The Effect of Managers on Public Service Provision: Evidence from Medicaid, SNAP, and TANF*

Link for Most Recent Version

Erik James†

November 12, 2024

Abstract

This paper studies how public sector managers impact both the quantity and quality of public service provision. I use novel administrative data containing case review decisions for three public benefit programs in the state of Texas: Medicaid, SNAP, and TANF. In this setting, managers oversee teams of caseworkers deciding whether to permit or deny household applications and can influence both the quantity of applications reviewed and the quality of decision-making. I document wide variation in quantity- and quality-based measures of performance across manager teams. I exploit variation in caseworker-manager assignments to show that managers explain 8-10% of the overall variation in caseworker performance. I find that higher manager quantity-based performance does not come at the cost of quality. Replacing managers in the lowest quartile of quantity-based performance with those at the 75th percentile would increase output by 5.6%, enough to eliminate the organization-wide case backlog that led to months-long delays in case review decisions.

*I am grateful for support from The RIDGE Program and the Boettner Center at the University of Pennsylvania. The content is solely the responsibility of the author and does not represent the official views of the USDA, University of Pennsylvania, or any other supporting entity.

†Erik James: The Wharton School, University of Pennsylvania (ecjames@wharton.upenn.edu)
I would like to thank my advisors, Fernando Ferreira, Jessie Handbury, Benjamin Lockwood, and Alex Rees-Jones for their advice on this project. I would also like to thank Michael Best and Corinne Low for valuable comments.

1 Introduction

Managers are key intermediaries within large public sector organizations that implement policy and guide day-to-day operations. Yet the extent that manager differences impact public service provision is not well understood. Public sector managers face unique challenges in influencing organizational performance. They are often constrained by rigid rules, limited budgets, and a lack of performance-based incentives. While these restrictions may limit managers' influence, managers may be important precisely because there are limited ways public sector organizations can motivate and direct their workers, thus relying more on the manager to influence and oversee teams. Furthermore, measuring how managers influence both the quantity and quality of public services is critical for understanding variation across managers. Managers may differ in their productivity or the way they prioritize competing objectives, which have different implications for the types of policies that are most likely to improve welfare.

This paper studies how differences in public sector managers impact both the quantity and quality of public service provision. I start by quantifying differences in quantity-based and quality-based measures of performance across managers within a public sector organization. I then measure the extent that differences in worker performance are driven by variation across managers. Next, I determine if managers achieve higher quantity at the cost of quality. Finally, I quantify the impact of manager differences on public service provision, overall and relative to workers. These goals have important implications for effective policy implementation, bureaucratic performance, and the standardization of public sector decision-making.

I study public sector managers that supervise teams of caseworkers determining eligibility for U.S. benefit programs. In 2022 the U.S. federal government distributed $930 billion in benefits to low-income households through Medicaid, SNAP, and TANF.[1][2] Information frictions exist in identifying who is eligible for these benefits, which has led to the creation of large public sector bureaucracies in each state dedicated to determining program eligibility. I study one of these bureaucracies, Texas Health and Human Services ("Texas HHS" hereafter), which allocated $68 billion in benefits in Texas in 2022, over 7% of the nationwide total. In this setting, caseworkers review household applications and determine whether to permit or deny requests for benefits. Managers supervise teams of caseworkers and provide on-

1. SNAP stands for Supplemental Nutrition Assistance Program and was formerly formerly known as the food stamp program. TANF stands for Temporary Assistance for Needy Families, and offers cash assistance. Medicaid offers health insurance.

2. Data for fiscal year 2021-2022 is from the Kaiser Family Foundation State Health Facts for Medicaid Spending, USDA SNAP Data Tables for SNAP spending, and the Office of Family Assistance for TANF spending.

the-job training, monitor performance, and implement program guidelines. These managers have limited tools and financial incentives to improve performance, but the number of cases their teams review (quantity) and the accuracy of their team's decision-making (quality) are consequential for the timeliness of services, household outcomes, program size, and the uniformity and fairness of public service provision. Many features of this setting are common across the public sector.[3]

An important contribution of this paper is measurement. I use novel data containing the universe of 30 million caseworker eligibility determination decisions from 2018 to 2023 to construct quantity- and quality-based measures of performance for each worker under every manager within a large public sector organization. I measure quantity using throughput: the number of cases reviewed per caseworker-month. This tracks closely with how productivity is defined in the private sector (Syverson 2011). Measurement of throughput is facilitated by the production of a homogenous product with one key non-manager input (caseworkers) with minimal multitasking, collaboration, or variation in non-labor inputs. I consider quality to be the accuracy of case decisions, for which I measure two important components: permissiveness and the miss rate. Permissiveness is the share of cases that are permitted, while the miss rate is a proxy for the share of cases that are incorrectly denied, i.e. false negative errors.[4] I do not observe false positive errors directly, but I return to this later in the paper.

I begin by documenting substantial variation in quantity- and quality-based outcomes across manager teams. This is facilitated by cases being assigned by the same automated statewide queues and then reviewed under uniform administrative guidelines. This means that local or regional variation in case review demand, program eligibility, or review guidelines will not drive differences in manager outcomes. I find that managers at the 90th percentile of throughput have 31% higher throughput relative to those at the 10th percentile. Repeating this separately for both quality-based measures, I find that managers at the 90th percentile have 9% (5 p.p.) higher permissiveness and 35% (2 p.p.) higher miss rate than managers at the 10th percentile, respectively. These differences in throughput ("productivity") and permissiveness *within* a public sector organization are similar to or slightly smaller than variation documented across workers, offices, and plants in the existing literature.

Next, I measure the impact of managers and find that managers explain an important

3. There are many public sector settings where supervised workers make repeated binary classification decisions, where both the decision-making accuracy and amount of decisions made are important for performance and public service provision. This includes bureaucrats granting permits, judges granting bail, radiologists diagnosing pneumonia, police officers searching vehicles, and caseworkers administering social programs more broadly.

4. The miss rate is the share of cases that are denied, reapply in the next 3 months, and are then permitted to receive benefits. Reapplying is how in practice applicants address incorrect denials.

part of the variation in caseworker throughput, permissiveness, and miss rate. Following Abowd, Kramarz, and Margolis (1999), I exploit variation in caseworker-manager assignments (i.e. "switchers") to decompose the causal effects of managers from the causal effects of caseworkers.[5] Caseworkers switch managers as a result of churn and reassignment of both managers and caseworkers to new teams. Using both an event study and variance decomposition, I find that managers explain 8-10% of the variation in overall caseworker throughput, permissiveness, and miss rate. This is a large effect considering that managers supervise teams of about 10 caseworkers.

I then investigate whether manager throughput (quantity) comes at the cost of accuracy (quality), and whether differences in productivity or differences in preferences seem to explain variation in these measures across managers. Differences in manager productivity would imply that managers are making caseworkers more or less productive and would create variation in throughput and accuracy that is positively correlated. On the other hand, differences in manager preferences would imply that managers are instead shifting caseworker decision-making rather than making them better or worse. This would create variation in throughput and accuracy that is negatively correlated and suggest that higher manager throughput is achieved at the cost of accuracy. To investigate this, I use the empirical joint distribution of manager impacts on throughput and accuracy. In my setting, managers maximize accuracy by minimizing false positive and false negative decision-making errors. I do not observe false positive errors, so I use the method described in Chan, Gentzkow, and Yu (2022) to measure differences in manager impacts on false positive errors and accuracy using differences in manager impacts on permissiveness and the miss rate.[6]

I document important variation in throughput and accuracy across managers, but find that throughput and accuracy are uncorrelated, suggesting that high throughput managers do not sacrifice accuracy. Furthermore, this suggests that there is variation in both manager productivity and preferences, yet neither factor dominates, leading to substantial heterogeneity across managers.[7] This finding is invariant to the assumptions made about the relative cost of false positive and false negative errors. I contrast these findings for managers with

5. I conduct standard checks for endogenous manager assignment and use the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023) to address limited mobility bias and sampling error.

6. This requires that (i) the miss rate is a relevant proxy for false negatives and (ii) caseworker-manager switches are uncorrelated with drift in the share of cases that should be permitted after conditioning on observed differences in case composition.

7. I explore what manager characteristics and behavior are correlated with manager throughput, accuracy, and decision-making. I find a mild correlation between manager tenure and throughput as well as variation in throughput and accuracy across regions and by urbanicity. For permissiveness, I find that Black managers have higher permissiveness than white managers and more permissive managers also award higher average SNAP benefit amounts, suggesting they are either also more generous on the intensive margin or have better benefit targeting.

caseworkers where throughput is correlated with differences in decision-making, which makes the relationship between caseworker throughput and accuracy depend on the relative cost of false positive and false negative errors.

An important implication of these findings is that a naive staffing policy that selects managers to increase throughput without considering accuracy or decision-making won't have unintended positive or negative impacts.[8] Enacting a one-time cut of the 10% lowest throughput managers and rehiring according to the existing manager distribution would increase total output for Texas HHS by 2.1% with no impact on accuracy, permissiveness, or the miss rate.

I conclude by illustrating that differences in manager impacts have important implications for public service provision, overall and relative to caseworkers. Shifting the lowest throughput quartile of managers to the 75th percentile of throughput would increase organization-wide output by 5.6%, or 1.5 million cases between 2018 and 2023. This is comparable in size to the backlog of cases Texas HHS faced in 2021 and 2022 that led to months-long delays in case review. Shifting the least accurate quartile of managers to the 75th percentile of accuracy would increase their accuracy by 1.5 p.p. and organization-wide accuracy by 0.4 p.p. from 2018 to 2023. For permissiveness, shifting the least permissive quartile of managers to the 75th percentile would increase the organization-wide share of cases permitted by 1.4%, which would increase program costs by at least $406 million over this period, or just under $1 million per manager-year. I also provide the first evidence of the relative importance of managers and workers in the public sector. I find the impact of shifting manager throughput, accuracy, permissiveness, and miss rate on a per worker basis are 3-4.5 times larger than the impacts of the same exercise for caseworkers.

This paper contributes to four literatures. First is the literature quantifying the impacts of differences across managers. This include a nascent literature on public sector managers that quantifies how differences in managers impact both quantity- and quality-based measures of public sector performance (Bloom et al. 2015; Rasul and Rogger 2018; Choudhury, Khanna, and Makridis 2020; Bertrand et al. 2020; Janke, Propper, and Sadun 2019; Limodio 2021; Fenizia 2022; Munoz and Otero 2022; Muñoz and Prem 2024) In this literature, causal evidence for public sector managers is relatively rare, reaches mixed conclusions regarding whether managers matter (e.g. Janke, Propper, and Sadun 2019; Munoz and Otero 2022), and does not estimate separate causal effects of managers on both quantity- and quality-based performance measures.[9] While evidence in the private sector for within-firm differences

8. This assumes that existing managers that are not fired do not respond to the policy.

9. Closest to this work, Fenizia 2022 shows that managers in a similar setting drive quantity-based productivity without impacting error rates, but does not explore whether or not managers impact either error rates or permissiveness. Janke, Propper, and Sadun 2019 investigates the impact of public sector CEOs on

in managers is more robust, it also does not quantify how managers separately impact quantity and quality (e.g., Lazear, Shaw, and Stanton 2015; Adhvaryu et al. 2019; Adhvaryu, Kala, and Nyshadham 2022; De Stefano, Bidwell, and Camuffo 2022; Giardili, Ramdas, and Williams 2023; Metcalfe, Sollaci, and Syverson 2023). This is important for measuring manager performance, especially in the public sector where revenue or profit measures often don't exist, organizational objectives are less clear, and the quality of public services is particularly important.

Second, I contribute to the literature that documents systematic differences in decision-making and quality of public sector agents. This includes literature focusing on differences in decision-making and permissiveness across agents (Anwar and Fang 2006; Maestas, Mullen, and Strand 2013; Dobbie, Goldin, and Yang 2018; Autor et al. 2019; Chan, Gentzkow, and Yu 2022; Arnold, Dobbie, and Hull 2022; Feigenberg and Miller 2022; Cook and East 2023) and differences in quality more broadly (Fredriksson, Öckert, and Oosterbeek 2013; Chetty, Friedman, and Rockoff 2014b; Khan, Khwaja, and Olken 2016; Bandiera et al. 2021; Best, Hjort, and Szakonyi 2023; Mulhern 2023). I connect this literature with the management literature by showing that the systematic differences in decision-making and quality (accuracy) are in part explained by differences in management. I also simultaneously quantify the impact of agents on quantity (throughput) in a setting where agents can reveal their preferences by choosing their productive quality and quantity. While I find no correlation for managers, I show that differences in caseworker quantity and decision-making are correlated, suggesting that differences in the amount of work done could explain differences in the way agents make decisions.

Third, this paper contributes to work on U.S. benefit program design, administration, and incomplete program take-up (Boadway, Marceau, and Sato 1999; Currie 2006; Prendergast 2007; Kleven and Kopczuk 2011; Bettinger et al. 2012; Rossin-Slater 2013; Bhargava and Manoli 2015; Ganong and Liebman 2018; Deshpande and Li 2019; Finkelstein and Notowidigdo 2019; Homonoff and Somerville 2021; Gianella et al. 2022; Wu and Meyer 2022; Cook and East 2023; Elzayn et al. 2024). This paper provides some of the first empirical evidence for how Medicaid and SNAP administrators – both caseworkers and managers – impact program outcomes, participation, and benefit generosity. This suggests that the agency and effectiveness of administrators should be an important consideration in the design of public benefit programs.

Finally, this paper contributes to the literature documenting differences in productivity across workplaces (Syverson 2004; Foster, Haltiwanger, and Syverson 2008; Syverson 2011;

both quantity- and quality-based measures of performance, but finds that CEOs have minimal impact on these outcomes.

Bartelsman, Haltiwanger, and Scarpetta 2013; Chandra et al. 2016; Fenizia 2022). I provide additional evidence that the differences in productivity (throughput) within a highly uniform public sector organization are similar to or slightly smaller than variation documented in other settings.

This paper proceeds as follows. In Sections 2 and 3, I provide institutional background for my public sector setting and describe my data. In Section 4, I measure differences in manager performance across manager teams. In Section 5, I quantify the causal impact of managers on caseworker throughput, permissiveness, and the miss rate. In Section 6, I determine if managers achieve higher quantity at the cost of quality, and explore whether differences in manager productivity or preferences explain differences in throughput and accuracy across managers. In Section 7, I quantify the extent manager differences impact public service provision. In Section 8, I conclude.

2 Institutional Background

2.1 Texas Health and Human Services

Texas Health and Human Services (Texas "HHS") administers multiple federal public benefit programs in the state of Texas, including the Supplemental Nutrition Assistance Program (SNAP), Medicaid, the Children's Health Insurance Program (CHIP), Medicaid for the Elderly and Disabled (MEPD), and Temporary Assistance for Needy Families (TANF). It is one of the largest agencies in the Texas state government, with about 37,000 employees.[10] There are a wide range of functions performed by Texas HHS to administer these public benefit programs, but one of the largest functions is determining program eligibility. This includes reviewing applications for benefits and managing existing cases to verify continued eligibility. There are over 7,000 full-time employees allocated to this function, the majority of which are caseworkers and supervisors (referred to hereafter as "managers"). The salaries of caseworkers and managers cost $215 million per year during 2018 to 2023. Caseworkers are the lowest-level employees that review cases and make decisions. Managers supervise a team of caseworkers and do not review cases. Caseworkers and managers are located in more than 300 offices across the state. On average, each office has about 5 managers, each which oversees a team of about 11 caseworkers.[11]

10. Government Salaries Explorer Database Published by the Texas Tribune, https://salaries.texastribune.org/departments/.

11. There are three different levels of caseworkers. Level 1 caseworkers are typically within their first year and are still learning and ramping up their case review. Level 2 caseworkers are the vast majority of caseworkers and cases reviewed. Level 3 caseworkers begin to transition into an assistant manager role, where they not only review cases but provide assistance to other caseworkers. In addition, there are other

2.2 Background on Managers

Managers are the second level of the bureaucratic hierarchy for eligibility determination operations at Texas HHS above caseworkers. They are responsible for guiding day-to-day operations of the case review process. This includes providing on-the-job training, monitoring production, assigning tasks, resolving issues, and motivating and evaluating workers. In addition, managers are key for implementing new policy changes on the ground by introducing caseworkers to new review protocols and computer functionality. They also do administrative work relating to coordinating special investigations and reviews, preparing reports, approving schedules, monitoring compliance, and managing the physical office location. Managers are described as having "considerable latitude for the use of initiative and independent judgement".[12]

Despite their wide set of responsibilities, managers have limited tools to improve their team's performance. Related to staffing, they are not involved in the hiring process, cannot choose who is staffed to their team, and cannot choose the size of their team. In terms of worker incentives, managers cannot set pay or use performance incentives, and they have limited ability to influence the promotion or firing of workers, which are uncommon events. Managers have limited control over some important parts of the case review process. They do not determine the scope of the case review work, have no say in the organization-wide state review guidelines for how cases are to be reviewed, and have minimal control over what productive inputs they can use. On top of these limited tools, managers also have very limited incentives to improve their team's performance. Managers do not have their own performance incentives and often stay in their role for many years without any increases in pay linked to their performance.[13] Managers are incentivized to do well to qualify for promotion and avoid termination, but these are rare events.

Even though managers are limited in certain ways and are not the ones reviewing cases, they are anecdotally quite important for a variety of reasons. First, the case review process is incredibly complex and many caseworkers have limited experience. Managers have many years of experience reviewing cases prior to becoming managers, so they can help answer questions, provide on-the-job training, and help resolve unusual issues for their workers. Second, managers are responsible for implementing new program guidelines and explaining to staff how new processes will work. Doing this well is important for avoiding delays and errors. Third, managers shape caseworker decision-making and trade-offs. Caseworkers are

specialist caseworkers that are either deployed at hospitals or specialize in the review of MEPD cases. This paper focuses only on the standard level 2 caseworkers.

12. Texas Works Supervisor I Job Description, https://hr.sao.texas.gov/Compensation/JobDescriptions/R5630.pdf.

13. For both managers and caseworkers, about 75% of variation in salary is explained by experience and a time fixed effect.

balancing the speed of their work with the number of errors they will make so they will meet their performance guidelines. In addition, caseworker decision-making trades off between the possibility of false positive and false negative errors. Managers evaluate caseworker performance and advise them about how to balance these trade-offs in situations where formal guidance for decision-making is not fully available or information is limited. Fourth, managers are often physically needed to check caseworkers. The computer system can request a manager override, which requires the manager to process and approve the request in order for the caseworker to continue working.

Table 1 provides descriptive information on managers and caseworkers in this setting from 2018 to 2023 available from public employee records. The average manager tenure at Texas HHS is about 11.7 years, with 3.6 years in the manager role. This highlights that managers often are caseworkers for many years prior to promotion to manager, making them very familiar with the work they are managing. The nominal average manager salary from 2018-2023 is about $50,000, which is 43% larger than the average caseworker salary. In terms of demographics, about 80% of managers are female, which reflects that women are a large share of the workforce across Texas HHS. In terms of race and ethnicity jointly categorized, 51% are Hispanic, 24% are Black, and 24% are white.[14] Education is not recorded for 43% of managers, but among managers who do report, 65% have a 4-year college degree. Lastly, manager promotion is rare and termination is exceedingly rare. Over a six-year period, only 18% of managers are promoted from their role within Texas HHS. This is the main performance incentive for managers and caseworkers, yet it is relatively uncommon for both.

14. The overall population of Texas is 40% Hispanic, 40% white, 12% Black, and 5% Asian.

Table 1: Manager and Caseworker Descriptives: 2018-2023

	Managers	Caseworkers
Tenure at Texas HHS (years)	11.7	6.26
Tenure as Manager (years)	3.6	
Gross Annual Salary (in dollars)	50,331	34,676
Female	.80	.85
Race/Ethnicity: Hispanic	.46	.49
Race/Ethnicity: Black	.30	.28
Race/Ethnicity: White	.21	.20
Education: 4-Year College or Higher	.43	.28
Education: Less than 4-Year College	.23	.23
Education: Missing	.34	.49
Promoted	.18	.02
Terminated	.01	.03
Number of Workers	959	7,143

Notes: Summarizes descriptive information for managers and caseworkers from the quarterly Texas public employee records dataset for 2018-2023.

As an organization, Texas HHS' objective is to provide benefits accurately and in a timely fashion. Manager and caseworker performance are evaluated based on these objectives. Managers and caseworkers have their performance reviewed on a routine basis and are rated as either below, meeting, or exceeding expectations, but there are limited financial implications of these reviews. Managers and caseworkers are evaluated on the number of cases reviewed and the error rate. Caseworkers need to spend at least 95% of their time "online" reviewing cases and have an average case processing time within service standards to meet expectations. The service standards for cases vary, but for most cases are between 35 and 45 minutes. These requirements imply that there are case review thresholds that caseworkers and manager teams need to meet for the different levels of performance. The second performance metric is referred to as the "error rate".[15] A sample of five cases per caseworker per month are reviewed by an external quality assurance team to check the accuracy of decision-making. In addition, managers can pick up errors during the day as they assist and review ongoing work. Caseworkers and teams are supposed to have an error rate below 5%. An error could either be an incorrect denial or an incorrect approval, but it can also be awarding an incorrect level of benefits for SNAP or TANF. It is not clear if managers and workers meet this threshold as an organization, and for at least several years after the

15. This paper does not use any formal data from the quality assurance process, nor was this process active for large parts of the study period. I proxy for decision-making errors from my data in a different way.

onset of COVID this formal quality assurance process was stopped altogether.

2.3 Case Assignment and Review Process

Case assignment is the process of assigning case actions to caseworkers for review. A case action is created when the electronic case management system determines that caseworker review is needed to determine the status of a particular case, i.e. whether the household is permitted to receive a particular benefit and if relevant how much their benefit amount is.[16] Most of the time this is when the households submits either an initial application to start receiving benefits or a recertification to be certified to continue receiving benefits. All other caseworker actions that do not correspond with the submission of an application by the client are called "incomplete reviews".[17] Case actions are allocated into electronic queues and assigned to caseworkers by the electronic computer system based on their relative priority.

Even though caseworkers and managers are located in offices dispersed across the state, the process of case assignment is largely done through a centralized statewide process. Cases from the entire state are electronically categorized into 5 different queues (also referred to as "tracks") based on the type of case and sorted by priority level.[18] Managers assign caseworkers to work on one of these queues at a given point in time. Not all caseworkers are qualified to work all types of cases, but almost all caseworkers are qualified to work SNAP and Medicaid cases, which represent a dominant share of overall case volume. Upon finishing a case, a caseworker receives the next case in the statewide queue in what is described as a "next up" assignment process.[19] Caseworkers only work on one case at a time, conducting an interview if required and then deciding whether or not to permit or deny the case based on standardized review guidelines and eligibility criteria.[20] Permitted cases then receive benefits, while denied cases must reapply or appeal their case decision. Despite households having multiple case actions associated with their case across months and years, caseworkers

16. Almost all changes to a case's status occur through a case action implemented by a caseworker, with the exception of terminating benefits for households that do not recertify.

17. For example, validating whether or not additional documentation was submitted for an expedited application, dealing with a reported change in income, or a request to change a household's address would create an incomplete review action.

18. One queue is for SNAP and Medicaid (including CHIP) applications, a second is for SNAP and Medicaid (including CHIP) recertifications and incomplete reviews, a third for TANF cases, a fourth is for Medicaid for the Elderly and People with Disabilities ("MEPD") cases, and a fifth is for "missing information" cases. Missing information cases have been previously reviewed by a caseworker and pended because they could not be disposed in their current form.

19. This process should be thought of as quasi-random conditional on queue, within queue variation in case type, and the time of case assignment.

20. In some situations when the application fails a major barrier like being largely incomplete or failing to complete the interview, the case is not reviewed and instead reassigned later to another caseworker via the missing track after giving the applicant time to address the issue.

are reassigned each time an additional action is taken.[21]

3 Data

3.1 Data Structure

My main dataset is Texas HHS' "Case Action" data from January 2018 through September 2023. The Case Action data contains one record for each caseworker decision (i.e. "action") to permit (i.e. approve) or deny a request for benefits.[22] The Case Action data includes records for multiple programs including Medicaid,[23] SNAP, and TANF cases. It also includes records for multiple types of applications, including initial applications, recertifications, and "incomplete reviews".[24] The data has about 50 million case actions. Often a household's application is reviewed for multiple programs or subcategories within a program, which are often reviewed by the same caseworker. In these situations, I randomly sample one of the benefit programs for the application-caseworker pair.[25] This results in about 30 million remaining caseworker actions.

The data contain information for each application including when and how the application was submitted, the specific type of benefits applied for, and some information about the household.[26] It then includes if the case was permitted or denied, the amount of benefits awarded (if relevant), when the decision was made, what caseworker took the action, the caseworker's office location, and who was the caseworker's manager at the time. I aggregate the data to construct a caseworker-month dataset where I identify the primary manager for each caseworker in every month.[27]

21. Only for joint applications are multiple actions reviewed by the same caseworker.

22. Not all requests are associated with an application. For example, if a client reports a change in income, this will trigger a review request to determine if benefits should continue to be permitted (sustained), reduced, or removed.

23. This includes the Children's Health Insurance Program ("CHIP") and Medicaid for the Elderly and People with Disabilities ("MEPD").

24. Initial applications are the initial request to be permitted benefits. Recertifications are applications that request continuing benefits prior to their expiration. Incomplete reviews are all other actions that are not associated with a formal application submitted by the household. These include income checks, requests for changed benefit amounts, changes in address, and follow-up on verification requirements that are delayed for expedited SNAP cases.

25. Calculations of the miss rate use the entirety of the data prior to this sampling protocol.

26. This includes whether the head of household is a senior, eligible household size, and eligible number of kids. I do not observe income or true household size or number of children for all applications, only those that are approved.

27. The primary manager for a caseworker-month must manage at least 70% of the caseworker's cases in that month, otherwise the primary manager is left missing.

3.2 Caseworker-Level Outcome Measures

For each caseworker under every manager I create three monthly measures of quantity- and quality-based performance: throughput, permissiveness, and the miss rate. I measure quantity using throughput: the number of cases reviewed per caseworker-month. Throughput reflects the amount of output per unit input and tracks closely with how productivity is often defined in existing research (Syverson 2011).[28] A feature of my setting is that the production process creates one tangible output (cases disposed) using one key input (caseworkers), making measuring throughput much less complex than in many other settings. Furthermore, there is no multi-tasking, variation in non-labor inputs, collaboration across workers or teams, or part-time concerns. In addition, my detailed caseworker data will allow me to measure and adjust for differences in input quality across managers. I discuss adjustments for different types of cases disposed in Section 4.

I consider quality in my setting to be the accuracy of case decisions.[29] Accuracy is maximized by minimizing the number of false positive and false negative errors. I measure two important components of accuracy: permissiveness and the miss rate. Caseworker permissiveness, also referred to as the "permit rate" or "approval rate", is the share of cases that are permitted to receive benefits. This is often thought of as a proxy for differences in worker leniency or preferences in existing work. The miss rate is the share of cases that are denied where the applicant then submits a new application within 3 months and is successfully approved. The miss rate is a relevant proxy for false negative decision-making errors in this setting because in practice incorrectly denied applicants address their situation by reapplying.[30] I do not observe data on false positives, which I return to in Section 6.

I also measure the average benefit amount for approved SNAP cases for each caseworker-month. The average benefit amount for SNAP cases is a function of household size and net income and is relevant for thinking about benefit generosity on the intensive margin and targeting of public benefits (e.g., Finkelstein and Notowidigdo 2019; Homonoff and Somerville 2021). While I observe income for approved cases, the strict nature of Medicaid requirements in Texas implies that very few non-SNAP households receiving benefits report having earned income. This makes it impossible with my data to measure the income or

28. In my setting, a manager is more productive if they have both higher throughput and higher accuracy.

29. In this type of setting, the time to dispose a case would often be an importance performance metric, which reflects the time between when the application is filed and a case determination is made. However, in my setting caseworkers review applications one at a time and spend about 35-45 minutes on each case. Given this and the statewide case assignment process, the time to dispose a case will largely reflect the time it took for the application to be assigned to the caseworker that then disposed it, and will not relate to caseworker or manager performance.

30. When applicants reapply they receive a different quasi-randomly assigned caseworker. While applicants that are incorrectly denied can dispute the decision, this is rare and very inefficient for applicants.

incurred medical costs for households receiving public health insurance benefits.

3.3 Supplemental Datasets

I supplement the Case Action data with two additional data sources. First, I use the program participation record data from Texas HHS that has benefit program-by-month information about program participation for each case. I use this to obtain previous and past participation for every case relative to the time an application decision was made. This data includes additional demographic and income information for participating households. Second, I use a public employee records dataset with the demographics, wages, and tenure of all employees at Texas HHS. A subset of about 40% of the managers in my Case Action data have been linked to this additional employee information.

4 Measuring Differences in Performance Across Manager Teams

In this section I provide descriptive evidence of differences across manager teams in throughput, permissiveness, and the miss rate. It is not clear that there will be persistent variation in these quantity- and quality-based measures of performance; managers have limited financial incentives, limited toolkits for driving outcomes, and standardized decision-making guidelines their teams need to follow in this public sector setting. Furthermore, variation in throughput across these public sector managers may not be as relevant compared to variation in within industry productivity documented in the private sector. Regarding permissiveness and the miss rate, existing work documents variation across public sector workers, but has yet to measure the same variation across public sector managers.

Often there are many challenges with comparing managers across a large organization because of variation in tasks, policy, resources, or clientele across locations. However, program administration at Texas HHS is very centralized and uniform, with limited administrative variation across counties or regions. This includes benefit program requirements, case review guidelines, training programs, worker objectives, worker incentives, computer systems, and the case assignment process.

The centralized statewide case assignment process is a particularly useful feature for comparing managers and caseworkers. Caseworkers across the state of Texas regardless of their location are reviewing the same cases from the same five statewide queues. Cases are automatically assigned to caseworkers in a "next-up" fashion by a centralized computer system. This means that differences in outcomes across caseworkers and managers are not driven by local or regional differences in case review demand or eligibility. It also means that caseworkers and managers are not able to manipulate the cases they are assigned, making case composition very comparable conditional on queue (i.e. the type of case being worked) and time of assignment.

One challenge this assignment process does not address are differences in case composition across caseworkers based on which queues they are assigned. Managers determine the queue that caseworkers are assigned to on a given day or even hour. This means that different caseworkers will work on different queues for different amounts of time and receive different types of cases (e.g. SNAP vs. TANF, initial application vs. recertification). These cases may have different levels of difficulty and chances of being permitted. To address this issue, I residualize my outcomes by removing differences in outcomes explained by differences in case type, other observed application-level case characteristics, and variation across time. I do this by regressing the outcomes of interest on observable case characteristics including case type by time fixed effects, then use the regression residual as my outcome of interest.[31] Residualizing outcomes is preferred in my setting over including controls due to computational feasibility.

Figure 1 shows the distribution of log throughput, permissiveness, and miss rate across manager teams. The gray histogram represents unadjusted outcomes, while the red bars represent outcomes after residualization has removed variation explained by differences in case composition, observed case characteristics, and across time. Starting with throughput, managers on average dispose 180 cases per caseworker per month. After residualization, managers at the 90th percentile have 31% higher throughput than managers at the 10th percentile, a 0.3 log point or 49 cases per caseworker difference. For permissiveness, managers are on average permitting 56% of cases. Residualizing permissiveness dramatically reduces variation across managers because different types of cases in different time periods can have very different likelihoods of being permitted to receive benefits.[32] Managers at the

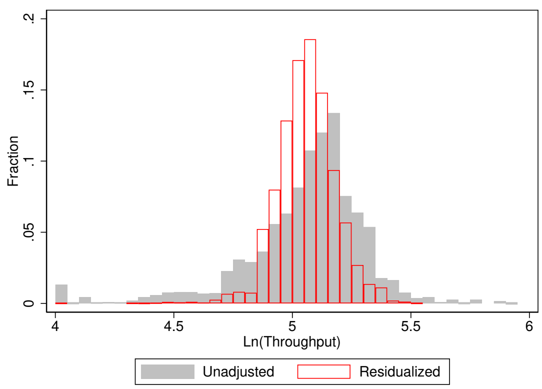
31. For permissiveness and the miss rate, I run an application-level regression of the indicator for whether or not a case was permitted or missed on program type (SNAP, Medicaid, MEPD, TANF) by application type (initial, recertification, incomplete review) interactions with the month the application was submitted, caseworker organization, what county the case came from (and whether or not that county is within the caseworker's county or region), how the application was submitted, an indicator for prior participation, an indicator for prior application submission, and whether or not the head of household is a senior. I also include program type by application type specific controls for the month of disposition and more granular program subcategories. I would like to control for more specific case-level characteristics like household size, the presence of kids, primary language, and household income, but these variables are either not available or possibly correlated with caseworker attributes due to possible differences in reporting in my data. I then average the regression residual across each caseworker-month. For throughput, I do the same process with case shares at the caseworker-month level, since throughput is not a case-specific outcome.

32. For example, the permit rate for SNAP recertifications is over 70%, while the permit rate for a TANF initial application is less than 30%.

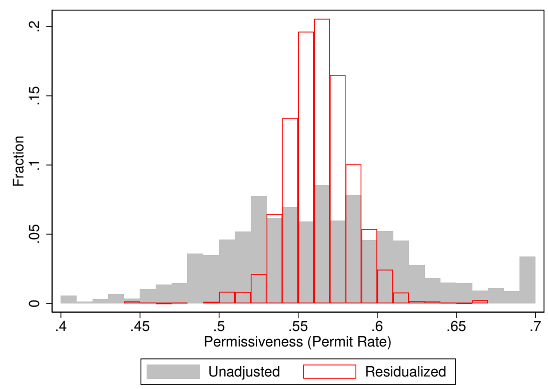
90th percentile of permissiveness are 9.3% (5 p.p.) more permissive than those at the 10th percentile. For the miss rate, the average miss rate is 6.5%. This means that 6.5% of all applications reviewed are denied but reapply and are permitted successfully in the next 3 months. Residualizing reduces the variation in the miss rate across managers less than for permissiveness. Managers at the 90th percentile have a 35% (2.0 p.p.) higher miss rate than those at the 10th percentile, where a higher miss rate is worse for manager performance.

Figure 1: Distribution of Log Throughput, Permissiveness, and Miss Rate Across Manager Teams

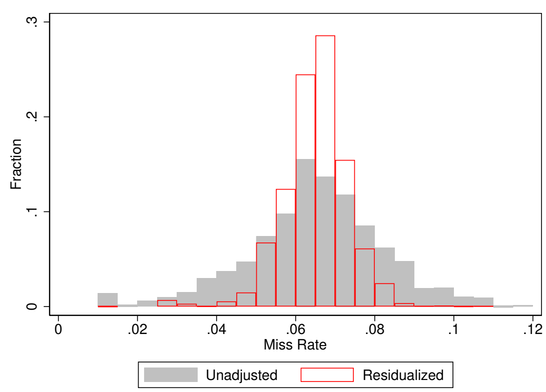
(a) Ln(Throughput)



(b) Permissiveness



(c) Miss Rate



Notes: plots distribution of log throughput, permissiveness, and miss rate across managers, which for each manager is the weighted average across all caseworker-month observations assigned to that manager. The unadjusted distribution are the outcomes "as-is". The residualized distribution has been adjusted for differences in case composition and across time, renormalized around the overall outcome mean. The unadjusted and residualized standard deviations for log throughput are 0.3 and 0.12, for permissiveness are 0.07 and 0.02, and for the miss rate are 0.02 and 0.01, respectively. The unadjusted and residualized difference between the 90th and 10th percentile for log throughput are 0.55 and 0.3, for permissiveness are 0.15 and 0.05, and for the miss rate are 0.04 and 0.02, respectively.

To facilitate comparison of variation in throughput to estimates of variation in productivity in other settings, I calculate the variation in within-office-year log throughput.[33] I find that the standard deviation of within-year unadjusted log throughput across offices within a year is 0.20 with an interquartile range of 0.23 and a difference between the 90th and 10th percentile of 0.45. After residualization, which controls for detailed differences in case composition and variation across time, the standard deviation of log throughput is 0.14 with an interquartile range of 0.16 and a difference between the 90th and 10th percentile of 0.33. This estimate of the standard deviation of productivity across offices within this statewide public sector organization is smaller than estimates across offices in a similar nationwide public sector organization of 0.37 (Fenizia 2022) and smaller than within industry variation in productivity for manufacturing between 0.2 and 0.4 (Syverson 2004; Foster, Haltiwanger, and Syverson 2008; Bartelsman, Haltiwanger, and Scarpetta 2013). This is more comparable to the standard deviation in productivity documented across hospitals of 0.17 (Chandra et al. 2016).

I compare residualized differences in manager and caseworker permissiveness to variation in permissiveness documented in other public sector settings for non-manager agents.[34] For managers, the standard deviation for residualized permissiveness is 2 p.p. with a difference between the 90th and 10th percentile of 5 p.p., while for caseworkers this is 5 p.p. and 12 p.p., respectively.[35] As expected, variation across managers is often a bit smaller than variation in permissiveness across non-manager agents in other settings, while caseworker variation is very comparable (Maestas, Mullen, and Strand 2013; Dobbie, Goldin, and Yang 2018; Autor et al. 2019; Chan, Gentzkow, and Yu 2022; Feigenberg and Miller 2022; Cook and East 2023).[36]

Together, this evidence suggests that there are important differences in throughput, permissiveness, and the miss rate across manager teams. However, these differences do not represent the causal impact of managers since these differences could be driven by a variety of factors including differences in the quality of caseworkers assigned to managers. This is addressed in the next section.

33. I identify 200 office locations with more than 3 standard level 2 caseworkers per year and I average unadjusted and residualized log throughput across caseworker-months within an office-year. I then remove differences across years for unadjusted log throughput.

34. As in Figure 1, this averages residualized permissiveness across all caseworker-month observations weighted by number of cases, either across managers or caseworkers.

35. The standard deviation and difference between the 90th and 10th percentile are similar for the causal effects of managers (2 p.p. and 5 p.p.) and caseworkers (4 p.p. and 10 p.p.) estimated in Section 5.

36. Maestas, Mullen, and Strand (2013) finds a s.d. of 6 p.p. for SSDI disability examiners (Figure 3), Dobbie, Goldin, and Yang (2018) finds a s.d. of 3 p.p. for bail judges (Figure 1), Autor et al. (2019) finds a s.d. of 5 p.p. for disability insurance judges (Figure 3), Chan, Gentzkow, and Yu (2022) finds a s.d. of 1.0-1.5 p.p. for radiologists diagnosing pneumonia (Online Appendix Table A.5), Feigenberg and Miller (2022) appears to find a s.d. of less than 2 p.p. for motor vehicle searches of highway patrol troopers, Cook and East (2023) finds a s.d. of 0.03 for caseworkers reviewing SNAP cases (Figure 2).

5 Quantifying the Effect of Managers on Caseworker Performance

In this section I estimate the causal impact of managers on quantity- and quality-based measures of performance. More specifically, I quantify what share of the overall variation in caseworker throughput, permissiveness, and the miss rate are explained by differences in managers. I exploit variation in caseworker-manager assignments ("switchers") to identify caseworker and manager effects using the method pioneered by Abowd, Kramarz, and Margolis (1999) (henceforth, "AKM"). I first estimate the simple AKM model and discuss identification challenges. Next, I use an event study to illustrate the impact of managers while validating the AKM identification assumptions before completing additional specification checks. Then I include a more detailed discussion of the AKM estimation before concluding with a variance decomposition that measures what share of the variation in caseworker performance is explained by differences in managers and caseworkers.

5.1 Identification

I model residualized caseworker-month outcomes as a linear and additively separable function of a caseworker component and manager component.[37] For caseworkers i, manager m, and time t

$$\widehat{y}_{it} = \alpha_i + \theta_{m(i,t)} + u_{it} \tag{1}$$

The residualized dependent variable $\widehat{y}_{it}$ is either log throughput, permissiveness, or the miss rate. α_i is the time-invariant portable part of caseworker throughput and $\theta_{m(i,t)}$ is the time-invariant portable component of manager throughput, which are also referred to as fixed effects. u_{it} is the error term. There are no additional time or case composition controls since all of this variation has been removed by the residualization process described in Section 4. I use the residualization process rather than include controls for computational reasons.[38] I estimate the model using ordinary least squares on the largest connected set in my data, which contains over 98% of caseworker-month observations. In addition, I drop the transition month when a caseworker's manager changes because the manager impact may take time to phase in.

37. Even though I observe a caseworker's office, I assume that offices have no impact separate from the caseworker and manager. This may be a reasonable assumption in my setting given that there are no clear differences in production technology and tasks across offices.

38. In Appendix Tables 15 to 17 I show my variance decomposition results are similar when instead of residualizing outcomes I use less granular case composition controls and time fixed effects.

Table 1 shows the adjusted R^2 for the estimation of Equation 1 for log throughput in Column (2), permissiveness in Column (4), and the miss rate in Column (6). The other columns show the amount of variation explained by caseworkers.[39] Comparing these show that the adjusted R^2 increases when manager fixed effects are included, even after controlling for differences across caseworkers.[40]

Table 2: Analysis of Variance: Log Throughput, Permissiveness, and the Miss Rate

	Ln(Throughput)		Permissiveness		Miss Rate	
	(1)	(2)	(3)	(4)	(5)	(6)
Observations	158,525	158,517	158,613	158,605	158,613	158,605
R-Squared	0.381	0.399	0.355	0.377	0.231	0.249
Adjusted R-Squared	0.357	0.371	0.329	0.348	0.201	0.214
Caseworker FE	X	X	X	X	X	X
Manager FE		X		X		X

Notes: Shows the adjusted R^2 of regressions of caseworker outcomes on fixed effects, with Columns (2), (4), and (6) corresponding to Equation 1, which when compared to Columns (1), (3), and (5) highlight the additional variation in caseworker outcomes explained by managers. Each column either uses caseworker residualized log throughput, permissiveness, or the miss rate as the dependent variable. The fixed effects included are noted. Given that the dependent variable has already been residualized, there are no time or case composition controls.

This model utilizes "switchers" to estimate the separate manager and caseworker components. Switchers are caseworkers with more than one manager and managers with more than one caseworker. In my setting, managers have multiple caseworkers in a given month t, effectively making all managers switchers. However, caseworkers do not work with multiple managers at the same time. Therefore, I rely on managers and caseworkers "moving" teams across time periods to connect teams. These moves can occur within or across physical office locations.

For this model to identify the causal effect of managers and caseworkers, the assignment of caseworkers to managers through this switching process must be as good as random conditional on the caseworker fixed effect.[41] More specifically, (1) drift in outcomes and caseworker-manager switches must be uncorrelated and (2) there cannot be sorting based on unmodeled match effects (Card, Heining, and Kline 2013). The first assumption implies that managers cannot sort strategically based on caseworker-specific shocks or trends in the

39. A manager fixed effect alone has an adjusted R^2 of 7-9%. See Appendix Tables 1 and 2.

40. In addition, managers are explaining this variation in caseworker outcomes across multiple caseworkers in a given period. This difference in adjusted R^2 is small relative to similar estimates for managers without worker-specific fixed effects (e.g. Bertrand and Schoar 2003; Fenizia 2022) but is larger than those in a setting including worker-specific effects but with larger team sizes (e.g. Adhvaryu et al. 2019).

41. It is worth nothing that in this setting both caseworkers and managers can switch teams, making it important to consider both types of moves when thinking about exogenous assignment to managers.

outcomes of interest. For example, if good managers systematically switched to caseworker teams on positive outcome trends or experiencing positive outcome shocks, the impact of managers would be biased upwards. The second assumption requires that managers and caseworkers do not sort based on their comparative advantage working with each other relative to other caseworkers and managers. What is not a violation of assumption (2) is sorting based on the caseworker fixed effect. For example, "better" managers sorting to or being strategically paired with teams of better or worse caseworkers is not a violation of these assumptions.

Qualitatively, caseworkers and managers at Texas HHS typically move to address staff shortages created by high staff turnover. Over half of managers active in 2018 at the beginning of my time period had either left Texas HHS or changed to another role by the end of 2023. Staff are required to change teams to fill in gaps created by turnover when requested. These decisions are usually based on proximity and the logistical feasibility rather than comparative advantage. One identification concern is that "better" managers tend to be sent to teams whose exiting manager was not just "worse" put whose team was on a worse negative trends prior to their exit. If this is occurring, I will be able to see this differential pattern in my data. Moves can also occur when caseworkers and managers request to move offices, which is usually done to be closer to home.[42] There are no direct financial incentives for moving offices and limited overall financial incentives for managers, which suggests that managers are unlikely to move strategically to improve their performance.

5.2 Event Study and Other Diagnostic Checks

I first use an event study to characterize the evolution of caseworker outcomes around a change in manager event. This event study demonstrates how caseworker-manager switches identify the causal impact of the manager and tests the first identification assumption that caseworker-manager switches are uncorrelated with drift in outcomes.

The event study employs a specific subset of the data: caseworkers who spend four consecutive months with one manager before shifting to another manager for the next four months, allowing for up to one transition month and requiring pseudo-balanced data.[43] I

42. For example, a large metro area may have over 10 offices. A manager may be originally promoted to a manager position in an office farther from their house but then switch when an opening is created at a closer office. When a manager position opens, either a switch can be mandated or an internal job posting is listed that caseworkers or other managers from other teams or offices can apply for. Internal promotions of caseworkers within the same office do occur, and are about half of caseworker promotions to manager. However, this does not imply the caseworker is overseeing their prior team given that there are multiple managers and teams per office.

43. For event time $k = -4$ to $k = -1$ the caseworker must have at least 85% of their cases overseen by the outgoing manager. Event time $k = 0$ is the first period where the share overseen by the outgoing manager

refer to these as caseworker manager change "events", and I find that 1,640 caseworkers have at least one event, which is about 27% of caseworkers.

For caseworkers i, months t, and event time k, I estimate the following event study with a discrete and continuous treatment associated with a caseworker's change in manager event

$$\widehat{y}_{it} = \sum_{k \neq 1} [\pi_0^k D_{it}^k + \pi_1^k D_{it}^k \widehat{\Delta M}_i] + \sigma_i + \epsilon_{it}$$
$$where\ \widehat{\Delta M_i} = \widehat{\theta}_{i,incoming} - \widehat{\theta}_{i,outgoing} \tag{2}$$

$\widehat{y}_{it}$ is either residualized caseworker log throughput, permissiveness, miss rate, or another outcome of interest like the participation rate. D_{it}^k is an indicator for event time k and $\widehat{\Delta M}_i$ is the change in estimated manager fixed effect the caseworker experiences for throughput, permissiveness, or the miss rate at time 0. Lastly, σ_i is a caseworker fixed effect. This event study is estimated on the balanced sample with binned endpoints at $k = -5, 5$. I bootstrap my standard errors.

The coefficient π_0^k is the effect at event time k of receiving a new manager. The coefficient of interest π_1^k is the additional effect at event time k of receiving a manager with a higher or lower estimated fixed effect relative to the prior manager ($\widehat{\Delta M_i}$). Using the example of throughput, this is the differential effect on caseworker throughput of receiving a new manager that is higher versus lower throughput relative to the previous manager. The continuous treatment $\widehat{\Delta M_i}$ is not directly observable and is instead estimated using the AKM model in Equation 1. Estimating the independent variable using Equation 1 can create an artificial correlation between the dependent and independent variable that could bias the results (Fenizia 2022).[44] To address this issue, I estimate Equation 1 using a caseworker leave-out.[45] This also allows me to summarize my results using a differences in differences estimator.[46]

drops below 85%. I then require the incoming manager to oversee at least 70% of the caseworker's cases in event time $k = 1$ and 85% for $k = 2$ through $k = 4$. The pseudo-balanced data assumption allows one month in $k \in \{-4, -3, -2\}$ and $k \in \{2, 3, 4\}$ to not exist or fail the following assumptions or be missing.

44. In small samples, a shock to the outcome variable will not be fully averaged out of the estimated manager component in Equation 1. This means that a positive shock to the dependent variable will also raise the estimated manager component, creating an artificial correlation in Equation 2. See Fenizia (2022) for a more detailed explanation.

45. When obtaining $\widehat{\Delta M_i}$ for each caseworker i, I estimate Equation 1 separately for each caseworker leaving out the observations for caseworker i for $k \in [-8, 8]$. The window $k \in [-8, 8]$ was chosen to make sure that the outcomes of interest for $k \in [-4, 4]$ are not autocorrelated with the caseworker data used to estimate Equation 1. This window was chosen based on the autocorrelation across Equation 1 residuals four months apart being 0.04 for log throughput, 0.06 for permissiveness, and 0.05 for the miss rate. In Appendix Figures 4 and 5 I show my results are similar using an event time leave-out and a broader caseworker leave-out that removes data for other caseworkers experiencing the same change in manager event.

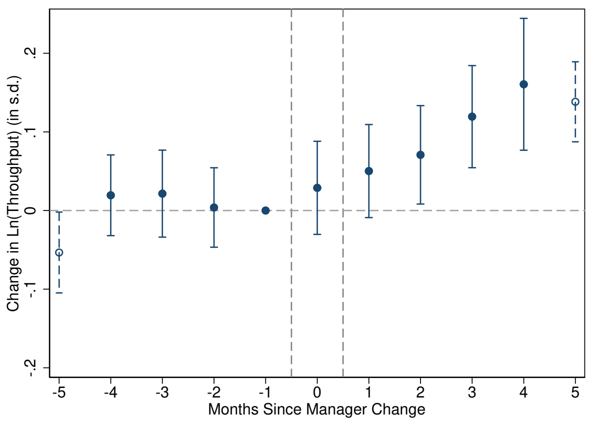
46. Difference in differences results compare the post period $k \in \{2, 3, 4\}$ relative to the pre period $k \in \{-4, -3, -2\}$, but are otherwise the same as the event study. This is done to show ancillary results more

Figure 2 shows the event study results, i.e. the estimated π_1^k from Equation 2. The x-axis plots event time months relative to time 0, which is when caseworkers experience the change in manager event. Changes in caseworker outcomes on the y-axis are standardized for comparability across outcomes. Estimates for the unbalanced binned points are denoted at event time $k = -5$ and $k = 5$. Prior to the change in manager, there is no evidence of pre-trends. This suggests that caseworkers that are about to receive managers with a higher or lower fixed effect are not on differential trends, speaking to the validity of the first identification assumption. If higher or lower throughput managers were selecting caseworkers based on outcome trends, caseworker throughput would be trending differentially prior to the manager switch.

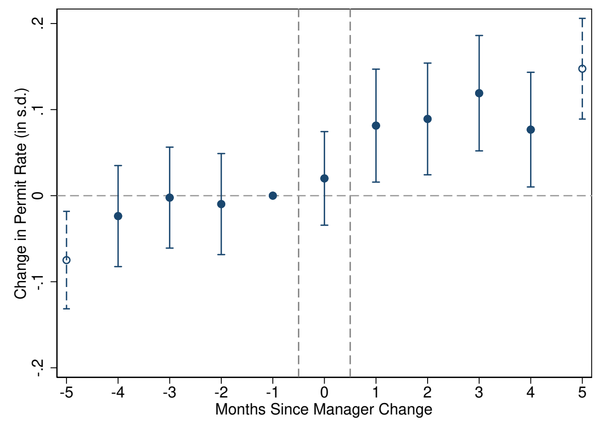
compactly. The event time leave-out approach from Fenizia 2022 estimates a separate regression for the effect in each event time period, and hence cannot be summarized using a simple pre and post.

Figure 2: Impact of a One Standard Deviation Change in Manager Fixed Effect on Caseworker Outcomes (in standard deviations)

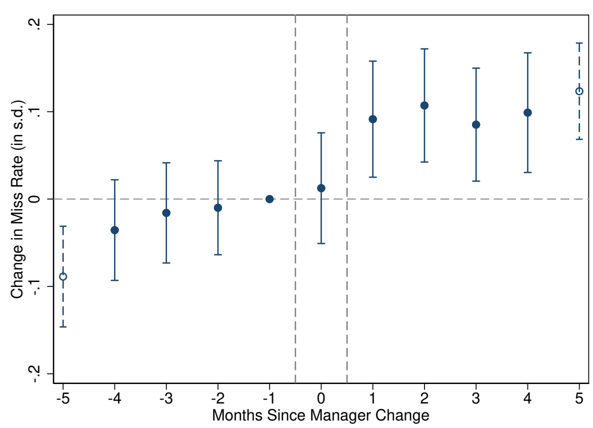
(a) Log Throughput



(b) Permissiveness



(c) Miss Rate



Notes: These plots show the impact on a caseworker outcome of receiving a manager at time 0 with a one standard deviation higher estimated manager fixed effect for that outcome, which is π_1^k in Equation 2. This includes caseworkers with pseudo-balanced data from 4 periods before to 4 periods after the month they experience a change in manager. The caseworker outcomes are standardized and the treatment (the change in manager fixed effect) is also standardized relative to the distribution of manager fixed effects. For example, subplot (a) shows the impact of a one standard deviation change in manager log throughput fixed effect on a caseworker's log throughput, in standard deviations.

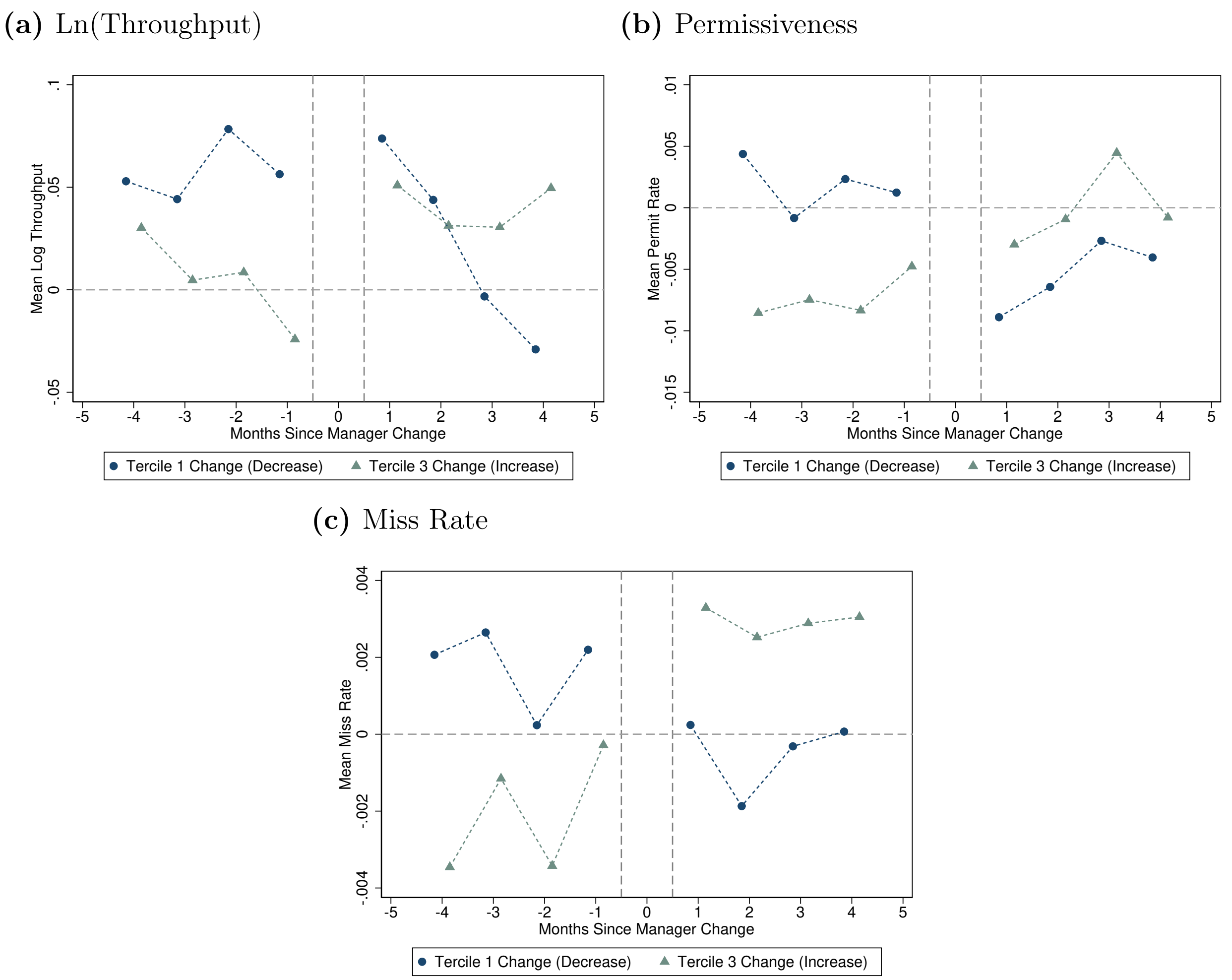
Interpreting for log throughput first, a one standard deviation increase in the manager's log throughput fixed effect increases a caseworker's log throughput by 0.1 standard deviations. The magnitude of the manager effect is also about 0.1 standard deviations for permissiveness and the miss rate as well. These effects represent a 4% increase in throughput, a 1.2% (0.7 p.p.) change in permissiveness, and an 11% (0.4 p.p.) change in miss rate. These results show that managers impact caseworker outcomes, and that these effects manifest within the first few months after a manager change.

However, this event study doesn't speak to the second identification assumption of sorting based on comparative advantage. To do this, I split the events in my event study into terciles based on the change in manager fixed effect at time 0 ($\widehat{\Delta M_i}$). I plot the mean caseworker outcomes for the bottom tercile (decrease in fixed effect) and top tercile (increase in fixed effect) events in Figure 3.[47] Given the limited number of events, the caseworker average outcomes in event time are noisy and standard errors are not included. As expected, bottom tercile events that experience decreases in the manager fixed effect at time 0 do experience a decrease in outcomes, and vice versa for the top tercile events. These changes also occur immediately with one exception; when a caseworker gets a manager with lower throughput. Examples in the literature often focus on wages related to job changes where an immediate shift is usually expected. However, with throughput the delayed onset patterns could be rationalized in multiple ways, including taking time to understand your new manager or gradually declining motivation.

Sorting based on caseworker-manager match-specific components would imply that decreases in throughput for caseworkers getting a lower throughput manager would be smaller than the increases in throughput for caseworkers getting a higher throughput manager. Ignoring the adjustment period after the manager switch, the magnitudes of the changes look comparable, providing no evidence of sorting on comparative advantage. Because a higher miss rate is worse, this logic would be reversed for the miss rate. I see if anything that caseworkers getting a higher miss rate ("worse") manager experience a larger increase in miss rate compared to those getting a lower miss rate ("better") manager, the opposite of what would be predicted by sorting on comparative advantage. Permissiveness is not an outcome where higher or lower permissiveness is better or worse and where it would not make sense to sort based on comparative advantage. Regardless, the change in magnitude is roughly similar.

47. These are the outcome means of the residualized outcomes $\widehat{y}_{it}$ after residualizing a second time to remove differences across caseworkers.

Figure 3: Average Caseworker Outcome Residuals Summarized By Event Time and Change in Estimated Manager Fixed Effect ($\widehat{\Delta M_i}$) Tercile



Notes: This plots the mean of caseworker residualized outcomes (after further removing caseworker fixed effects) in event time relative to the manager change at event time 0. Caseworkers are split based on their $\widehat{\Delta M_i}$ tercile, which is the change in manager fixed effect at time 0. Caseworkers in tercile 1 experience on average a decrease in $\widehat{\Delta M_i}$ while caseworkers in tercile 3 on average experience an increase in $\widehat{\Delta M_i}$. This includes caseworkers with pseudo-balanced data from 4 periods before to 4 periods after the month they experience a change in manager. The average caseworker outcome on the y-axis is not standardized.

Another way to investigate sorting on comparative advantage is to compare the estimation of Equation 1 to a saturated model. If caseworkers and managers were sorting based on comparative advantage, this would mean that caseworkers would have notably different outcomes with different managers. This would mean that caseworker and manager fixed effects alone would not do a good job of explaining outcomes, and that a saturated model with caseworker by manager fixed effects would better explain the model. Appendix Table 3 and Appendix Table 4 show the change in adjusted R^2 from the unsaturated model to a saturated model with caseworker by manager interactions is comparable to the change when a manager fixed effect is added. This suggests that there is limited scope for match-specific effects.

I also investigate if additive separability is satisfied. It could be that high-throughput managers may have different impacts on low-throughput and high-throughput caseworkers, which would violate additive separability. To check this, I divide observations into a grid based on their estimated caseworker and manager fixed effect quartiles, then look at the patterns of mean residuals from Equation 1. This is shown in Appendix Figures 6 to 8. The plots show no particular pattern of the residuals, and the range of the residuals is less than 2% as big as both the variation in the dependent variable and the sum of estimated caseworker and manager fixed effects. This means that additive separability is a reasonable assumption in this setting.

5.3 Estimation Details and Variance Decomposition

After estimating Equation 1, I use a variance decomposition to measure the relative importance of differences in caseworker and manager effects for throughput, permissiveness, and miss rate. The variance of the outcome $\widehat{y}_{it}$ can be decomposed into its various components as follows (Abowd, Kramarz, and Margolis 1999)

$$var(\widehat{y}_{it}) = var(\alpha_i) + var(\theta_{m(i,t)}) + 2cov(\alpha_i, \theta_{m(i,t)}) + var(u_{it}) \tag{3}$$

The ratio of the estimated variance of the manager effects $var(\widehat{\theta}_{m(i,t)})$ relative to the variance of the outcome $var(\widehat{y}_{it})$ is the share of the overall variation in the outcome explained by variation in managers.

There are three common issues that often arise when estimating these variance and covariance objects using Equation 1. The first issue is adjusting for multiple connected sets, but this is not an issue in my setting since the largest connected set contains over 98% of the data. The second issue is "limited mobility bias", which has been discussed at length and for which there are multiple proposed solutions (Andrews et al. 2008; Kline, Saggio, and Sølvsten 2020; Best, Hjort, and Szakonyi 2023).[48] In my data the amount of variation is relatively good compared to baselines established in the literature. Managers in my data have on average 15.5 different caseworkers, of which 10 are switchers. From the caseworker

48. Limited mobility bias can arise in situations where there is limited variation in caseworker-manager assignment where it will be challenging for the model to separately identify the caseworker from the manager contributions to the outcome of interest. This means that in cases where sampling error leads one of the fixed effects to be overestimated, the other will be underestimated, creating a spurious negative correlation between the caseworker and manager fixed effects. This would bias the variance terms upwards while biasing the covariance between the caseworker and manager components downwards.

perspective, I find that 55% of caseworkers have more than one manager. The third issue for estimation is consistency, which in this setting requires the number of observations for each caseworker-manager group to tend toward infinity.[49] Sampling error can be dealt with using either split sample or shrinkage approaches (Kane and Staiger 2008; Chetty, Friedman, and Rockoff 2014a; Finkelstein, Gentzkow, and Williams 2016; Best, Hjort, and Szakonyi 2023). I correct for both limited mobility bias and sampling error using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). This addresses sampling error similar to the standard shrinkage approach but also the possibility of covariance in the sampling errors across workers predicted by limited mobility bias.[50] In the main text I report only the results from the covariance shrinkage approach because I find that my results are relatively stable across specifications.[51]

The variance decomposition results are shown in Table 3. The first row is the standard deviation of the residualized outcome variable across caseworker-months, with each of the three main outcomes in each column.[52] The second set of rows report the standard deviation for the caseworker and manager effects, their correlation, and the standard deviation of the summed caseworker and manager fixed effect. The third set of rows report the share of the overall variance in the outcome explained by a particular component, which is calculated from the variance and covariance components, not the standard deviations and correlation in the set of rows above. Finally, the manager-caseworker ratio expresses how much outcome variation the manager explains relative to the caseworker.

49. The median number of months per caseworker-manager pairing in my data is 7, so the error in the estimates has likely not asymptoted to zero, especially for uncommon pairs.

50. I bootstrap Equation 1 to estimate the variance-covariance matrix of the sampling error across all workers (i.e. caseworkers and managers). I then use this to construct the optimal shrinkage matrix that minimizes the mean squared prediction error for the worker fixed effects. The intuition here is that limited mobility bias would predict that the sampling error across certain caseworker and manager pairings are negatively correlated, and the shrinkage matrix will take this into account.

51. For a comparison of the unadjusted estimates to the standard shrinkage and covariance shrinkage approaches for each outcome, see Appendix Tables 11 to 13.

52. This differs from the standard deviation of the histograms in Section 4 because this is the standard deviation across caseworker-month observations, not managers i.e. manager teams.

Table 3: Variance Decomposition: Share of Variation In Caseworker Outcomes Explained by Differences in Managers and Differences in Caseworkers

	Ln(Throughput)	Permit Rate	Miss Rate
SD of Outcome	0.446	0.070	0.032
SD of Case Worker Effects	0.282	0.043	0.016
SD of Manager Effects	0.128	0.023	0.009
Caseworker-Manager Effect Correlation	–0.225	–0.263	–0.300
SD of Caseworker + Manager	0.282	0.043	0.016
Share Caseworker	0.402	0.381	0.25
Share Manager	0.082	0.104	0.094
Share Covariance	–0.041	–0.052	–0.046
Share Caseworker + Manager	0.402	0.379	0.252
Manager Caseworker Ratio	0.206	0.273	0.377
Adjusted R-Squared	0.372	0.347	0.214
Number of Observations	152145	152229	152229
Number of Caseworkers	5817	5817	5817
Number of Managers	876	876	876
Number of Caseworker-Manager Pairs	12647	12652	12652
Number of Connected Sets	1	1	1

Notes: The variance decomposition shows what share of the overall variation in caseworker outcomes are explained by variation across managers and caseworkers. This is done by bootstrapping Equation 1 and adjusting using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). Comparisons to the standard estimates and standard shrinkage estimators are shown in Appendix Tables 11 to 13. Each column looks at a different outcome and a different set of estimated caseworker and manager fixed effects specifically for that outcome. "SD of Outcome" is the standard deviation of the outcome variable in each column. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" is obtained by dividing "Share Manager" by "Share Caseworker". Statistics for the estimation of Equation 1 are included at the bottom.

I find that managers explain 8-10% of the variation in throughput, permissiveness, and the miss rate and differences in managers explain between 20% to 38% as much of the variation in outcomes as differences in caseworkers. While this may make managers seem unimportant relative to caseworkers, managers impact outcomes for multiple caseworkers. I show that manager differences actually end up being more important than caseworker differences in Section 7. In Appendix Tables 18 and 19 I show that managers explain a similar share of the variation in permissiveness and the miss rate for different case types, suggesting that the role of managers is similar for Medicaid and SNAP as well as for initial applications and recertifications. Regarding the matching of caseworkers, the negative correlation between estimated caseworker and manager fixed effects suggests managers and caseworkers display negative assortative matching.[53]

The interpretation of the size of this effect depends on the size of manager teams, which in my context are about 10 caseworkers.[54] This reflects that managers in my setting have a "supervisory" role as opposed to being in a higher-level management position. While I am not aware of existing estimates for manager impacts on permissiveness or the miss rate, there are many existing comparisons for throughput, or "productivity". The share of variation in throughput explained by managers in my setting is very comparable to estimates for managers in other settings (Adhvaryu et al. 2019; Fenizia 2022; Giardili, Ramdas, and Williams 2023), but these managers manage more workers and hence have a bigger overall impact. In contrast, Metcalfe, Sollaci, and Syverson 2023 find that managers explain 25-35% of the variation in store productivity, but stores tend to have only about 4 employees. Another important factor to contextualize my results is that manager effects in my setting are estimated controlling for caseworker fixed effects as opposed to store, office, or plant fixed effects, which is uncommon. Therefore, my estimates do not include relevant mechanisms through which managers can drive outcomes (e.g. retaining experienced workers) and control for more confounding factors than plant- or office-specific controls.

Overall, these results demonstrate that managers in my setting have important impacts on multiple dimensions of quantity- and quality-based measures of performance.

5.4 Robustness

Manager vs. caseworker moves - my identification strategy uses variation in caseworker-manager assignments generated by both caseworkers and managers moving teams. One concern here is that including variation from caseworker moves may bias the estimation of manager effects because caseworkers that move also receive new peers. In Appendix Tables 7 to 9 I split my event study results for change in manager events that are very likely to be from a manager move and events that are more likely to be a caseworker move.[55] The impact

53. This negative correlation is after adjusting for limited mobility bias. This finding is relatively common (e.g. Andrews et al. 2008; Adhvaryu et al. 2019; Fenizia 2022; Metcalfe, Sollaci, and Syverson 2023), and could reflect either intentional negative assortative matching or "selection-based" negative assortative matching. Metcalfe, Sollaci, and Syverson (2023) explains that selection-based negative assortative matching results from selection rather than intentional sorting. For example, matches of low-throughput managers with low-throughput caseworkers may be unstable and lead to another switch occurring.

54. For example, CEOs explaining 8% of the variation in productivity of an entire company would be far more impactful than supervisors explaining 8% of the variation in productivity for a team of 10 workers.

55. I do not observe formal team designations, only a caseworker's manager assignment. Therefore, when more than one caseworker changes from the same old to the same new manager in the same period, this suggests that the manager moved. When only one caseworker changes from one manager to another in a

of managers is larger for events where the manager likely moved and smaller for events where the caseworker is more likely to have moved. This suggests my results are not driven by peer effects. Instead, this is consistent with what would be expected under negative assortative matching; caseworkers that receive a better manager on average receive worse caseworker peers.

Movers - variance decomposition results can be sensitive based on the sample included in the estimation (Andrews et al. 2008). There are many caseworkers that never change managers and whose effects are only identified via the movement of other caseworkers and managers. These non-moving caseworker effects may not be as well identified, or non-moving caseworkers may be systematically different than caseworkers that move, which is relevant for external validity. In Appendix Table 14, I show the variance decomposition estimated using only the subsample of caseworker movers. I find that under this specification my results for managers are largely unchanged.

Standard regression controls - I repeat the variance decomposition using unadjusted (non-residualized) outcomes and instead include much less granular controls in the regression. These controls are the share of cases per case type (SNAP vs. Medicaid vs. MEPD vs. TANF and initial vs. recertification vs. incomplete review) and month fixed effects. I show these results in Column (2) of Appendix Tables 16 to 17. I find that managers explain a similar amount if not more of the overall variation in outcomes.

Robustness for missing queue cases - in my setting there are cases assigned via the missing queue (also referred to as the missing track).[56] The missing track cases are not specifically denoted in my data, but can be proxied for because the majority of them are reviewed immediately prior to the application's due date. I can then control for this proxy for missing cases when residualizing outcomes, which may be important since their characteristics may differ from other standard cases. I show these results in Column (3) of Appendix Tables 16 to 17. I find that my results are largely unchanged.

given period, this suggests either the caseworker moved or that there were no other caseworkers on their team with suitable balanced data for the event study.

56. These are cases where no determination to permit or deny is made by the first caseworker because the application failed a major review barrier. For example, this occurs when an incomplete application is submitted or the applicant fails to complete the interview. These cases are later assigned to a caseworker after the applicant has been given a chance to rectify the issue.

6 Does Higher Manager Throughput Come at the Cost of Accuracy?[57]

In this section I explore whether higher manager throughput (quantity) is achieved at the cost of accuracy (quality). This depends on what drives differences in outcomes across managers: differences in manager productivity, differences in preferences, or a combination of both. This has important implications for the effectiveness of different policies aimed at improving or standardizing performance.

I start by discussing the variation in manager throughput and accuracy that would result from differences in manager productivity and preferences, and what implications these would have for the effectiveness of staffing policies. Second, I measure differences in manager accuracy using estimated differences in manager permissiveness and miss rate from Section 5 by adapting methods from Chan, Gentzkow, and Yu (2022) to my context. Third, I plot the empirical joint distribution of manager throughput and accuracy and evaluate the impact of a naive staffing policy that imposes a one-time cut of the bottom 10% of managers based on throughput without considering their accuracy or decision-making. Fourth, I repeat the same empirical exercise for caseworkers to compare and contrast across worker types.

6.1 Motivation: Implications of Differences in Manager Productivity and Preferences

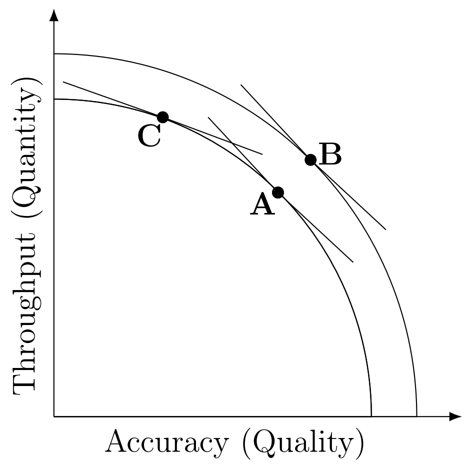
In this setting, managers engage in a production process where they contribute to caseworker production via throughput (quantity) and accuracy (quality). Managers want to maximize their impact on these two objectives. However, managers likely face a trade-off between these two objectives; managers can increase throughput by helping their caseworkers work faster but this may decrease the accuracy of the worker's decisions.[58] Each manager has a production possibilities frontier that represents the impact on caseworker throughput they can achieve for a given impact on caseworker accuracy. This is illustrated in Figure

57. Unless otherwise stated, reference in this section to manager or caseworker throughput, accuracy, permissiveness, miss rate, or false positive rate refer to the estimated causal impact of managers or caseworkers on those outcomes.

58. This trade-off is based in part on the assumption that caseworkers have the ability to "choose" throughput and accuracy in this setting and managers can influence this choice. For example, caseworkers can speed up decision-making by flipping a coin to decide the case, but this would come at the cost of accuracy. Similarly, caseworkers could increase accuracy by spending a long time on each case while reducing throughput. Therefore, even if managers can make productive improvements to throughput or accuracy without costing the other or have no ability to productively increase one of these measures, they can still achieve higher throughput or accuracy by influencing their caseworkers' decision-making in this way.

4. I consider two dimensions in which managers vary.[59] First, managers that vary in their productivity will be located on different production possibilities frontiers. In Figure 4, Manager B is more productive than Manager A; they can achieve higher throughput for any given impact on accuracy. Second, managers can vary in their productive preferences. In Figure 4, Manager A and Manager C have the same production possibilities frontier yet choose different locations on that frontier based on where their indifference curves create a tangency. Differences in preferences lead Managers A and C to have different throughput and accuracy. These differences in preferences could reflect that managers prefer different things or having different beliefs about the efficacy or costs of the actions they take.

Figure 4: Illustrative Manager Production Possibilities Frontier



Notes: illustrates differences in manager productivity and preferences using the manager production possibilities frontier for throughput and accuracy.

Differences in manager productivity and preferences could create similar variation in manager impacts for throughput and accuracy but for different reasons and with very different empirical patterns for the joint distribution of manager throughput and accuracy. In Figure 5, I plot three illustrative examples of the joint distribution of manager throughput and accuracy assuming that there is variation in both manger throughput and accuracy.[60] Subfigures 5(a) and 5(b) have similar variation in manager throughput and accuracy, but very different joint distributions. In Subfigure 5(a), variation across managers is driven by differences in productivity. Each manager is on a different production possibilities frontier and their productivity can be strictly ranked relative to one another. Situations where variation is driven

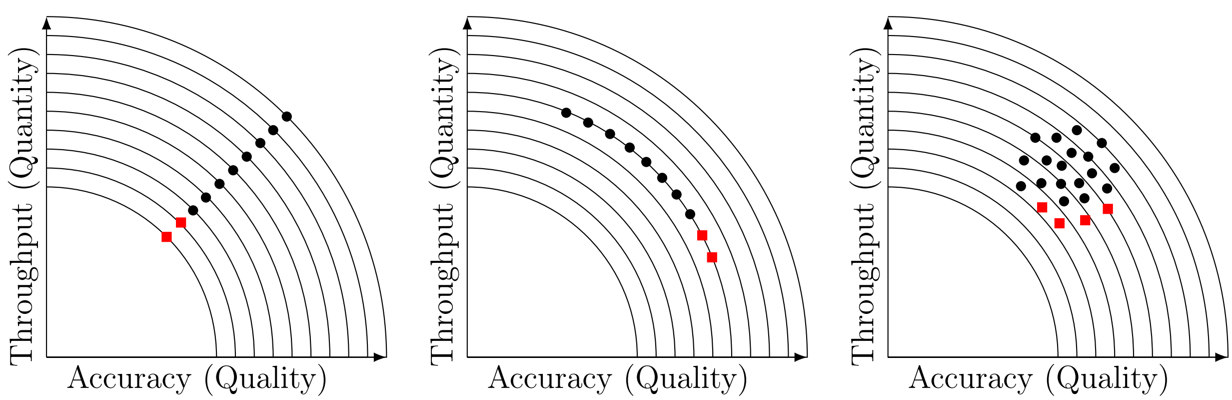
59. I assume that managers can be ordered by skill and there is no comparative advantage across managers. This means that managers are not allowed to be both more and less productive than another manager at two different levels of throughput. Otherwise, variation in the shape of the production function could create variation in manager impacts that in the cross-section cannot be distinguished from differences in productivity.

60. This represents cross-sectional variation across managers, where each dot is a different manager.

by productivity will result in a positive correlation between manager impacts on throughput and accuracy. Such an empirical relationship would be very hard to rationalize with other explanations like differences in manager preferences. In addition, variation in throughput and accuracy due to productivity is not necessarily inefficient.[61] In contrast, Subfigure 5(b) is an example where variation across managers is driven by differences in preferences. All of the managers have the same productivity but locate at different points on their common production possibilities frontier. This variation in preferences reflects a cross-sectional "quantity-quality" trade-off and results in a negative correlation between manager throughput and accuracy, suggesting that high throughput managers achieve higher throughput at the cost of accuracy. A negative correlation is highly suggestive of differences in preferences, but differences in productivity and comparative advantage cannot be ruled out. This is important because variation in preferences suggests that there is inefficiency relative to the social planner's preferred decision rule created by a principal agent problem. The final scenario in Subfigure 5(c) shows a "cloud" of managers. In this illustrative scenario there is variation in both manager productivity and preferences; there is variation both across and along the production possibilities frontiers. This suggests that variation in manager productivity and preferences are both relevant but neither factor dominates, creating a more nuanced story where there is a lot of heterogeneity across managers. This would be reflected by no correlation or a weak correlation between manager throughput and accuracy.

Figure 5: Illustrative Empirical Distributions of Manager Throughput and Accuracy

(a) Variation in Productivity **(b)** Variation in Preferences **(c)** Variation in Both



Notes: shows hypothetical cross-sectional joint distributions of manager impacts on throughput and accuracy under three different scenarios. Subfigure (a) shows variation resulting from differences in manager productivity. Subfigure (b) shows variation resulting from differences in manager preferences. Subfigure (c) shows variation resulting from a mix of both. This is used to implications for accuracy of a staffing policy that removes the lowest throughput managers denoted with red squares.

61. If managers were not assumed to be on their production frontier this pattern could be created by managers with the same production technology where some are not performing on their production frontier, which would be inefficient.

The different empirical joint distributions created by different drivers of manager variation for throughput and accuracy have important implications for staffing policies that change the composition of the manager workforce. I consider a candidate policy where Texas HHS wants to increase per-manager throughput by firing the lowest throughput managers. This would be similar to providing a lump-sum incentive to retain the best managers based on their past performance, which are both policies Texas HHS could implement.[62] This also reflects the type of policy that would be considered when quality is not fully observable or is costly to observe, yet is an important part of manager performance. The unintended impact of this policy on accuracy depends on the underlying driver of variation in manager throughput and accuracy. When variation is driven by differences in manager productivity as in Subfigure 5(a), this is a very effective policy because the lowest throughput managers (red squares) are the least productive managers with the lowest accuracy. Firing the lowest throughput managers improves throughput, accuracy, and manager productivity. In contrast, when variation is driven by differences in preferences as in Subfigure 5(b) this policy has negative unintended consequences. The lowest throughput managers are precisely the ones with the highest accuracy, leading to a worsening of accuracy while having no impact on average productivity. A more effective policy when variation is driven by differences in preferences would be implementing uniform decision-making guidelines that shift low-throughput managers to higher-throughput at the cost of accuracy, aligning decision-making with the preferences of the social planner. Finally, if differences in manager throughput and accuracy reflect a large amount of heterogeneity as in Subfigure 5(c), there will be minimal impacts of this policy on accuracy because manager throughput and accuracy are uncorrelated or weakly correlated. This provides smaller increases to average manager productivity than in Subfigure 5(a). However, assuming Texas HHS cares about throughput and accuracy, a more effective policy would instead target managers with low-productivity, i.e. both low-throughput and low-accuracy.

In the remainder of this section, I work towards plotting the empirical relationship between manager impacts on throughput and accuracy and evaluating the impacts of a naive staffing policy that replaces the lowest throughput managers. Next I describe how I measure differences in manager impacts on accuracy.

6.2 Measuring Manager Impacts on Accuracy

In my setting, accuracy reflects the quality of manager decision-making and is characterized by managers minimizing false positive and false negative decision-making errors. False

62. These are illustrative policy examples that are assumed to be one-off changes that would not induce responses from managers that were not fired.

positive errors (Type I errors) are situations where a case is permitted that should have been denied. False negative errors (Type II errors) are situations where a case is denied that should have been permitted. I define the causal impact of a manager j on accuracy (A_j) as the following

$$A_j = -(1-\beta)FN_j - \beta FP_j \tag{4}$$

FN_j is the causal impact of manager j on the share of cases that are false negatives and FP_j is the causal impact of manager j on the share of cases that are false positives. Here I assume that accuracy is linear in the share of cases that are false positives or false negatives, hence the level of decision-making errors is not relevant for considering the manager's impact on accuracy.[63] β reflects the relative cost (or "weight") of false positive errors relative to false negative errors for the social planner.[64] This depends on a variety of factors and is not known.[65] However in the context of manager decision-making, β reflects a manager's decision-making preferences, which may deviate from the social planner and create a principal agent problem.[66]

In the prior section, I measured the causal impact of managers on the miss rate, which is a proxy for false negative errors. However, I do not observe false positives and have not measured the impact of managers on false positive errors. This scenario mirrors a one-sided selection model where the true outcome (i.e. whether the case should have been permitted or denied) only reveals itself when cases are denied, but not when they are permitted. To navigate around this issue, I use methods from Chan, Gentzkow, and Yu (2022) to infer differences in manager impacts on false positives from differences in manger impacts on permissiveness and false negatives (i.e. false negatives), which is precisely what I quantify in Section 5.[67]

63. I also assume the cost of false positive and false negative errors are homogeneous, and the only variation in cost of errors is between false positive and false negative errors (β).

64. In general, the government incurs the cost of false positive errors by providing additional benefits to ineligible households. In contrast, the cost of false negative errors is borne by applicants who do not receive benefits they are eligible for or are forced to reapply in order to obtain those benefits. In both cases additional administrative costs of case review are imposed on the government to fix these errors, either to screen out incorrectly approved cases or to review the reapplications of incorrectly denied applicants.

65. I start by assuming $\beta = 0.5$, then show robustness for the full range of β from 0 to 1.

66. This would result in different permissiveness across managers. See Appendix Section C for more discussion.

67. Chan, Gentzkow, and Yu (2022) discuss how under certain conditions there is a one-to-one correspondence from "ROC space" based on false positives and false negatives to "reduced form space" defined by false negatives and permissiveness. This is used to create a test for differences in agent skill ("accuracy") in reduced form space without making assumptions about β or observing false positive errors. In my setting, I use this intuition to create a measure of accuracy according to the definition above and assuming different

I start by explaining the intuition while thinking about caseworkers reviewing cases, simplifying away from managers and causal impacts briefly. For caseworkers, the permissiveness is the share of cases permitted by the caseworker. By definition, permissiveness P_j for caseworker j is the share of cases permitted and is defined as[68]

$$P_j = S_j + FP_j - FN_j \tag{5}$$

where FP_j and FN_j are the share of cases that are false positives and false negatives. S_j is the share of cases received by the caseworker that should be permitted.[69] If a caseworker was not making any mistakes, they would have set $P_j = S_j$. Deviations from making no errors are reflected in FP_j, which increases the permit rate, and FN_j, which decreases the permit rate. P_j and FN_j are observed, but S_j and FP_j are not. Equation 5 can be interpreted the same way when thinking about manager causal impacts. A manager's causal impact on permissiveness is equal to their causal impact on false positives and false negatives plus the differences in the share of cases they allocate that should be permitted.[70]

I then assume that managers have a "common S_j" and take the difference in causal impacts between two managers j and j'. Section 5 measured these differences in causal impacts of mangers on P_j and on FN_j via the miss rate, which I can define relative to the "average manager" with a relative impact of 0. This transforms Equation 5 into[71]

$$FP_j = P_j + FN_j \tag{6}$$

This expression states that the relative difference in the impact on false positives for manager j relative to the average manager is equal to the relative difference in impact on permissiveness plus the relative difference in manager impacts on false negatives, both of which I have estimated. I then use Equation 6 to substitute for FP_j in Equation 4 to get an expression for differences in the causal impact of managers on accuracy in terms of permissiveness and

values for β.

68. The share of cases permitted is the share of cases correctly permitted (true positives) plus the share of cases that are false positives, $P_j = TP_j + FP_j$. The share of cases that should be permitted equals the share of cases correctly permitted (true positives) plus the share of cases that are false negatives, $S_j = TP_j + FN_j$. Subbing in for TP_j gives Equation 5.

69. This is usually referred to as the share of cases that are eligible, but in my setting the decision to permit focuses on making correct application review decisions that may not necessarily classify the cases' underlying eligibility.

70. For example, if there were no errors and managers received different shares of cases that should be permitted, shifting from one manager with a greater impact on P_j would imply that the manager's S_j increased the same amount.

71. First, comparing managers j and j' I get the expression $P_j - P_{j'} = (FP_j - FP_{j'}) - (FN_j - FN_{j'})$. However, I define j' as the "average manager" whose relative impact for standardized outcomes is 0, i.e. $P_{j'} = FN_{j'} = FP_{j'} = 0$.

the miss rate.

$$A_j = -(1+\beta)FN_j - \beta P_j \tag{7}$$

This shows how the joint distribution of differences in manager impacts for permissiveness and false negatives across managers now informs differences in accuracy. When managers increase permissiveness by ϵ, the miss rate must decrease by $\beta\epsilon$ to keep accuracy unchanged.

There are two key conditions for obtaining differences in manager impacts on false positive errors from impacts on permissiveness and the miss rate in this way. The first key condition is that the miss rate must be a relevant measure of false negatives. The miss rate is the share of cases that are denied that then reapply and are permitted within the following 3 months.[72] This is not a formal measure of false negative errors tied to ground truth by the evaluation from a third-party. However, the miss rate is likely to be highly correlated with the true false negative rate in my setting because the predominant way applicants address incorrect denials is by submitting a new application.[73] In addition to my miss rate being highly correlated with the false negative rate, I need to get the magnitude correct in order to correctly measure differences in false positive errors. In Section 6.5 I show my results are robust to scaling the miss rate up and down to reflect the possibility of mismeasurement.

The second key condition is the "common S_j" assumption, which is weaker than the assumption required under random assignment. When causal effects are identified using random assignment of cases across workers, $S_j = S$ would require that the share of cases that should be permitted is the same across the workers (e.g., Chan, Gentzkow, and Yu 2022). This would be sensible in the case of random assignment where the composition of cases is the same across workers, and validating random assignment would provide support for the assumption. In my setting the equivalent assumption for S_j goes back to the identification assumptions from the AKM model. The key condition is that drift in caseworker S_j (i.e. the share of cases that should be permitted) must be uncorrelated with caseworker-manager switches after controlling for differences in observed case composition. When a caseworker's manager changes, their permissiveness changes according to the manager's impact on permissiveness. The same happens for the miss rate (i.e. false negatives). These imply the impact of the manager on false positives unless the share of cases that the caseworker should be permitting is also changing systematically as the manager changes. The reason why S_j

72. In Chan, Gentzkow, and Yu 2022, the miss rate is defined as the share of patients diagnosed by a radiologist as not having pneumonia getting a new scan diagnosing pneumonia in the next 10 days.

73. An appeals process does exist, but applicants are often not aware of it and it takes much longer than simply reapplying and getting assigned to a new caseworker. In addition, only a small handful of denied cases if any are reviewed by the quality assurance team, meaning very few cases would be diverted from the miss rate due to the formal quality assurance process.

is likely not changing is that conditional on case type and time period controls, caseworkers receive similar cases from the same statewide queue through a "next up" assignment process with both the old and new manager.[74]

6.3 Manager Accuracy and the Empirical Joint Distribution of Manager Throughput and Accuracy

First I measure manager impacts on the share of false positive errors and accuracy from the joint distribution of manager impacts on the miss rate and permissiveness. For illustration purposes, I start by assuming the relative cost of false positive and false negative errors are the same ($\beta = 0.5$). Figure 6 plots the joint distribution of manager impacts on the miss rate and permissiveness normalized relative to the average manager.[75] The joint distribution has a negative correlation with a slope of -0.29.[76] In Subfigure 6(a), the heat map gradient represents the implied manager impact on false positives. When compared to the average manager, $FP_j = FN_j + P_j$. Hence managers along the dotted 45-degree line with a slope of -1 have equivalent impacts on false positives as the average manager. Then managers towards the upper right (green dots) have higher false positives, which is worse for accuracy. We can see that because the overall slope between the miss rate and permissiveness (-0.29) is less than -1, the implied false positive impacts are higher for managers with higher-permissiveness. This is expected in the sense that as permissiveness increases, there are more permitted cases that could become false positives and fewer that could become false negatives.

In Subfigure 6(b), the heat map gradient represents the impact on manager accuracy. Accuracy will be higher at a lower miss rate (shift straight down) and lower false positives (shift to bottom left), which when β is assumed to be 0.5 implies the following pattern for accuracy. Managers on the dotted line with slope -0.5 have the same accuracy as the average manager. We can see a few key insights about accuracy from this plot. First, when $\beta = 0.5$, the slope of the distribution between the miss rate and permissiveness (-0.29) is less than the slope of equivalent accuracy (-0.5, or $-\beta$). This implies under these assumptions that higher-permissiveness managers are less accurate. Second, the variation in accuracy

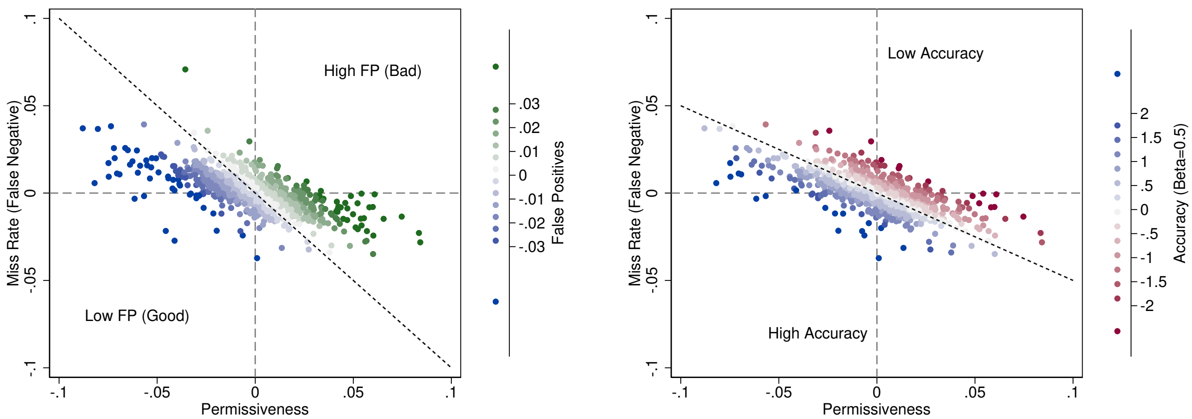
74. Managers assigning the caseworker different types of cases would not be a violation of this, because I have controlled for differences in observed case composition.

75. These are estimated fixed effects adjusted using the covariance shrinkage approach, so imprecisely estimated manager effects will be shrunk towards 0 and won't drive the variation between outcomes.

76. This negative correlation manifests for two reasons. First, as permissiveness increases, there are fewer cases that are denied that can become false negatives, which should reduce the miss rate. Second, as manager permissiveness increases, the conditional miss rate decreases because managers with higher permissiveness are more sure of the cases that they deny. See Appendix Figure 10.

Figure 6: Manager Impact Heat Map Plots for False Positives and Accuracy in Miss Rate and Permissiveness Space ($\beta = 0.5$)

(a) False Positives **(b)** Accuracy



Notes: These plots show how the joint distribution of manager impacts on permissiveness and the miss rate imply manager impacts on false positive rates and accuracy. The miss rate and permissiveness impacts are estimated from the AKM model in Equation 1 with the covariance shrinkage adjustment. False positives are determined based on Equation 6 and accuracy based on Equation 7 with the assumption that $\beta = 0.5$. Accuracy is standardized, but has a standard deviation of 0.8 p.p. relative to the standard deviation of permissiveness impacts of 2.2 p.p. and miss rate impacts of 1.0 p.p.

mainly goes across the width of the distribution, making the variation in manager impacts on accuracy (s.d. of 0.8 p.p.) less than the variation in permissiveness (s.d. of 2.2 p.p.) and the miss rate (s.d. of 1.0 p.p.). This is because the slope of the distribution between permissiveness and the miss rate (-0.29) is between -1 and 0 and because the assumption for $\beta = 0.5$ is not extreme.[77] However, the magnitude of the variation in accuracy is still important; caseworkers under managers at the 90th percentile of accuracy will make 1.9 p.p. fewer errors than those under a manager at the 10th percentile.[78]

Now that I have measured differences in manager impacts on accuracy, Figure 7 plots the cross-sectional empirical relationship between manager throughput and accuracy as discussed at the beginning of this section. Subfigure 7(a) shows the scatterplot of manager throughput and accuracy and the linear best fit line when $\beta = 0.5$. Subfigure 7(b) shows the same scatter for $\beta = 0.5$ but includes the linear best fit lines for various β between 0 to 1. This shows robustness from the extremes where only false positive errors matter ($\beta = 1$) and where only false negatives matter ($\beta = 0$). The scatterplot for manager impacts on throughput and accuracy is a big cloud of dots, and the slope of the best fit line in Subfigure 7(a) confirms that

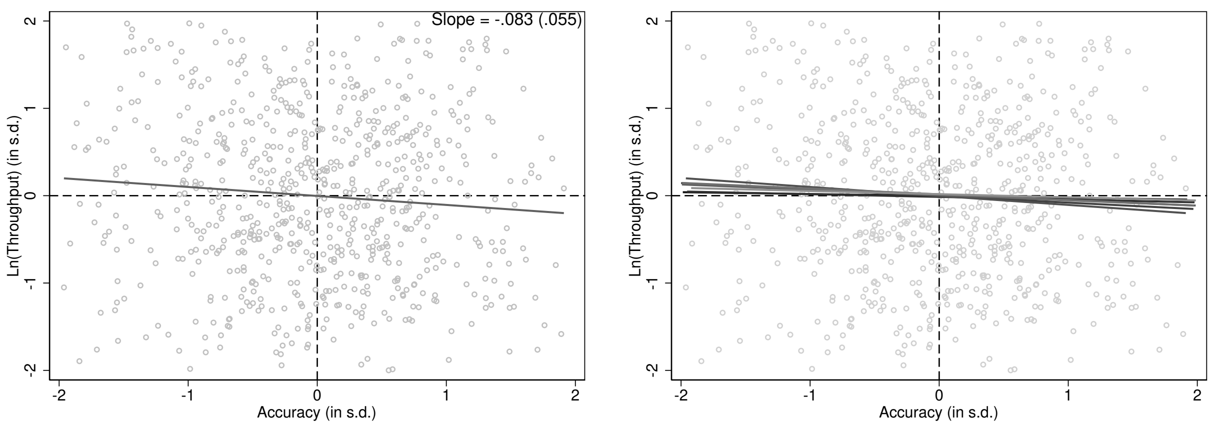
77. If the slope of -0.29 was instead less than -1 or greater than 0, this would result in greater variation in accuracy. See Appendix Section C for a more complete discussion of the relative importance of differences in manager accuracy and decision-making.

78. This varies from 1.6 p.p. to 3.6 p.p. depending on the value of β assumed. See Appendix Table 20.

Figure 7: Empirical Distribution of Manager Impacts on Ln(Throughput) and Accuracy

(a) Ln(Throughput) and Accuracy: $\beta = 0.5$

(b) Ln(Throughput) and Accuracy: β from 0 to 1



Notes: Shows the empirical distribution and best linear fit line between standardized manager log throughput and accuracy. Throughput impacts are the estimated manager impacts (fixed effects) for log throughput from the AKM model in Equation 1 with covariance shrunk estimates. Accuracy is obtained from the same impacts for permissiveness and the miss rate as defined in 7. The scatterplot in both subplots is the plot of standardized log throughput and accuracy impacts of managers assuming $\beta = 0.5$. The line in subplot (a) is the linear best fit between the two variables when $\beta = 0.5$. The lines in subplot (b) show the linear best fit under different values of $\beta \in \{0, 0.2, 0.4, 0.5, 0.6, 0.8, 1\}$, where $\beta = 0$ only weights false negative errors for accuracy and $\beta = 1$ only weights false positive errors for accuracy.

there is only a very weak negative correlation between manager throughput and accuracy.[79] The slopes plotted in Subfigure 7(b) for different β highlight that this finding is not sensitive to assumptions about the relative cost of false positives and false negatives. This shows that overall managers are not achieving higher throughput at the cost of accuracy. Instead, this is consistent with there being both variation in manager productivity and variation in manager preferences, but neither factor dominates. This is a more nuanced result where there is a lot of heterogeneity across managers in both productivity and likely preferences driving variation in throughput and accuracy. These findings contrast with cross-sectional evidence for public sector workers in Chan, Gentzkow, and Yu (2022) and Best, Hjort, and Szakonyi (2023) where differences in worker productivity or "type" was clearly the more important factor, leading to a positive correlation between two measures of worker quality. I find my situation is more nuanced, and has greater scope for differences in preferences possibly driving manager behavior.

To understand why manager throughput and accuracy are largely uncorrelated, Figure 8 shows the correlation between manager throughput and each decision-making component:

79. Appendix Figure 15 shows the conditional expectation function of throughput and accuracy to highlight that there is no evidence of non-linearities.

permissiveness, the miss rate, and the implied rate of false positive errors. Subfigure 8(a) shows that manager throughput and permissiveness is totally uncorrelated, a striking finding given the amount of variation across managers in both outcomes. In addition, Subfigure 8(b) shows a marginal positive relationship between throughput and the miss rate. which in part drives the marginal positive relationship between throughput and the implied false positive rate in Subfigure 8(c). Therefore, manager accuracy is not correlated with throughput because manager decision-making represented by false negative and false positive errors is uncorrelated with throughput. This is not because there is a lack of variation; managers are making different decisions and vary in their decision-making accuracy. In Appendix Section D I further validate this finding by looking at the reasons caseworker under a manager deny cases. Managers with different throughput don't deny cases for different reasons, providing evidence that managers are not systematically shifting the way they are evaluating cases to increase their throughput.

This lack of correlation between throughput and accuracy has important policy implications. It implies that naive policies that modify staffing focusing only on dimension of manager performance will have only small implications on the alternative dimension. To illustrate this, I consider a policy that would replace the bottom 10% of managers based on throughput.[80] I assume that these managers are re-hired from the distribution of existing managers. This is done to draw a new throughput, accuracy, permissiveness, and miss rate manager impact from the existing joint distribution.[81] Because manager throughput and accuracy are only marginally negatively correlated, this policy will remove managers with 0.1 s.d. higher accuracy (0.08 p.p. higher errors), which is not a statistically significant difference. The policy would result in an increase in total output of Texas HHS by 2.1% while having no impact on permissiveness, the miss rate, or accuracy.[82]

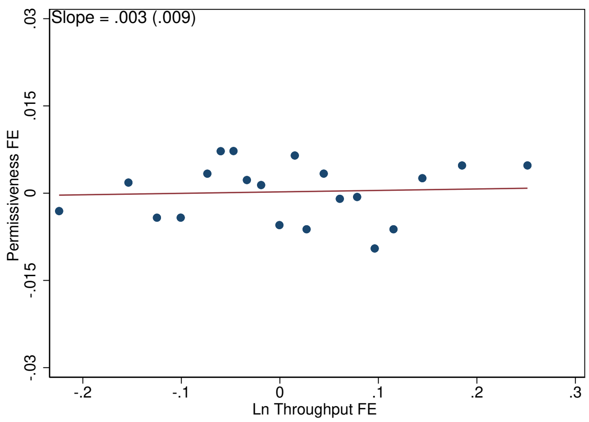
80. Given the notable levels of manager turnover, this would be similar to a one-time lump-sum incentive for higher throughput managers conditional on remaining at Texas HHS for a period of time.

81. Implementing this policy and measuring its benefits and costs would require additional information and empirical estimates. First, manager performance would need to be measured in a pre-period and the firing would need to be implemented at a particular point in time based on that period of measurement. Second, it is important to understand what types of staff Texas HHS can hire under existing incentives, how quickly, and at what cost. I find limited evidence that throughput and accuracy are correlated with experience in Appendix Section D, suggesting that new managers won't necessarily be far less productive. Third, I would need to quantify downstream costs of greater manager throughput, which for example include temporary or long-term increases in program participation or case review demands from recertification or reapplications. Fourth, there could be manager behavioral responses to increased chance of being fired for low throughput, which could also require a higher wage to compensate managers. To implement the same policy for high-throughput managers would require understanding which managers leave or get promoted, and how much increases in salary reduce manager attrition in this setting. For a similar policy analysis that discuss and address these issues in greater depth, see Chetty, Friedman, and Rockoff (2014b).

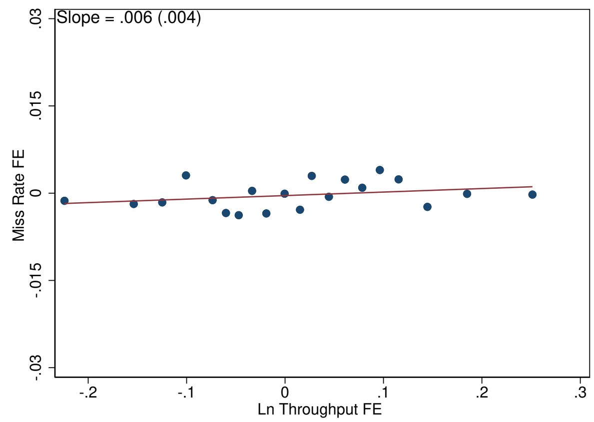
82. The implied change in accuracy would be less than a 0.01 p.p. decrease.

Figure 8: Conditional Expectation Functions for Estimated Manager Impacts

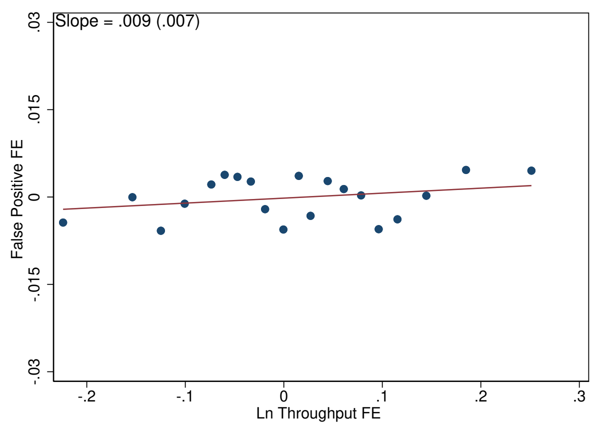
(a) Permissiveness and Ln(Throughput):



(b) Miss Rate and Ln(Throughput)



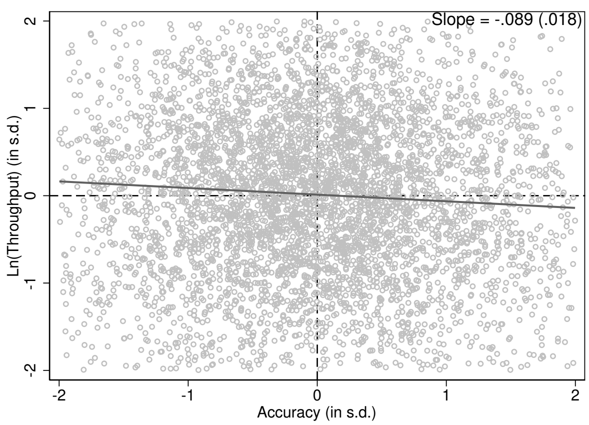
(c) Implied False Positive Rate and Ln(Throughput)



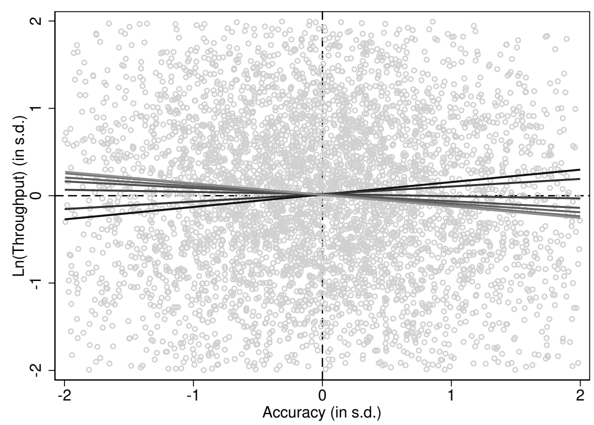
Notes: Plots the conditional expectation function between estimated manager impacts to show the relationship between manager log throughput and either permissiveness, the miss rate, or the implied false positive rate. All of these estimated impacts are covariance shrunk, and the manager impact on the false positive rate is obtained from Equation 6.

Figure 9: Empirical Distribution of Between Caseworker Impacts on Ln(Throughput) and Accuracy

(a) Ln(Throughput) and Accuracy: $\beta = 0.5$



(b) Ln(Throughput) and Accuracy: β from 0 to 1



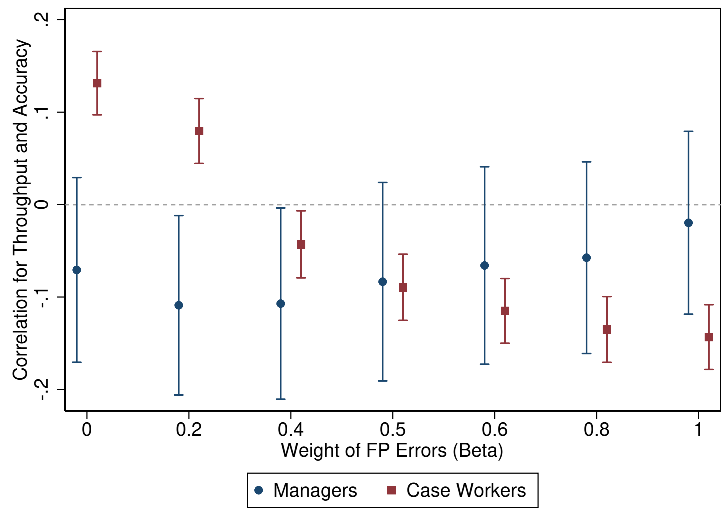
Notes: Shows the empirical distribution and best linear fit line between standardized caseworker log throughput and accuracy. Log throughput impacts are the estimated caseworker impacts for log throughput from the AKM model in Equation 1 with covariance shrunk estimates. Accuracy is obtained from the same impacts for permissiveness and the miss rate as defined in 7. The scatterplot in both subplots is the plot of standardized log throughput and accuracy impacts of caseworkers assuming $\beta = 0.5$. The line in subplot (a) is the linear best fit between the two variables when $\beta = 0.5$. The lines in subplot (b) show the linear best fit under different values of $\beta \in \{0, 0.2, 0.4, 0.5, 0.6, 0.8, 1\}$, where $\beta = 0$ only weights false negative errors for accuracy and $\beta = 1$ only weights false positive errors for accuracy.

6.4 The Empirical Joint Distribution of Caseworker Throughput and Accuracy

I follow the same process to measure differences in caseworker impacts on accuracy and investigate the empirical distribution of caseworker impacts on throughput and accuracy. In Appendix Figure 12 I find a similar joint distribution for caseworker impacts on the miss rate and permissiveness but with greater variance. Figure 9 shows the empirical relationship between caseworker impacts on throughput and accuracy. Subfigure 9(a) shows the caseworker scatterplot and best fit line when $\beta = 0.5$ while Subfigure 9(b) shows the best fit line for β between 0 and 1. Similar to managers, there is a large amount of heterogeneity across caseworkers driven by differences in both preferences and productivity. This leads to a weak negative correlation between caseworker throughput and accuracy when $\beta = 0.5$ that is very similar to that of managers. However, unlike for managers, Subfigure 9(b) shows that the relationship between caseworker throughput and accuracy differs depending on the value of β. When β is close to 0, there is a positive correlation between throughput and accuracy. However, when β is close to 1, there is a negative correlation.

This is explained further in Figure 10, which shows the estimated coefficient of throughput

Figure 10: Correlation Between Manager and Caseworker Impacts on Ln(Throughput) and Accuracy: β from 0 to 1



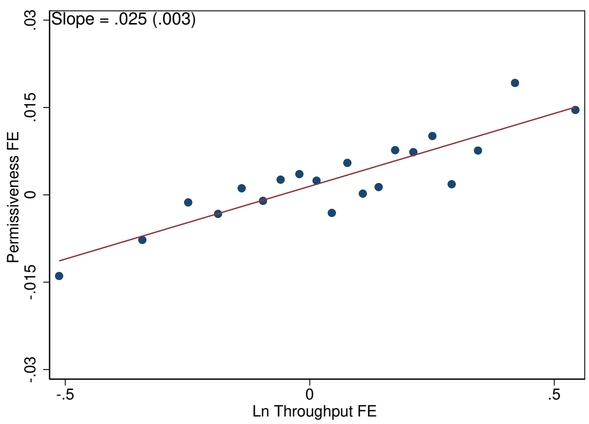
Notes: Shows the correlation coefficient between log throughput and accuracy for both managers and caseworkers under different assumptions of β, where $\beta = 0$ only weights false negative errors and $\beta = 1$ only weights false positive errors. The coefficient is from a regression of standardized worker log throughput impacts on standardized worker accuracy impacts, representing the correlation between the outcomes. These relationships are estimated on the full caseworker-month data clustering by either manager or caseworker.

regressed on accuracy, both for managers (blue circles) and caseworkers (red squares). The correlation for managers is quite consistent across the full range of β and if anything is slightly negative. In contrast, the relationship between caseworker throughput and accuracy is positive for low β and negative for high β. While this correlation is statistically significant, the correlation is not that strong in either direction.

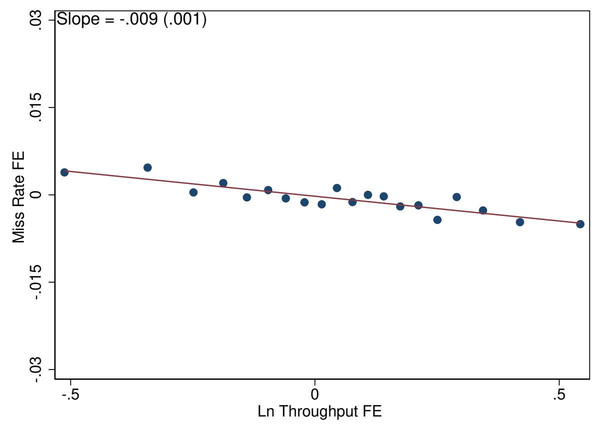
These results suggest that caseworker throughput is in fact correlated with differences in caseworker decision-making. In Figure 11, I plot the same relationships between caseworker impacts on throughput, permissiveness, the miss rate, and the implied false positive rate for caseworkers. In contrast to managers where there was largely no correlation between throughput and decision-making, caseworker throughput is positively correlated with permissiveness and negatively correlated with the miss rate. On average the highest throughput caseworkers have 3 p.p. higher permissiveness than the lowest throughput caseworkers, which is about 0.75 standard deviations. For the miss rate, the difference is about 1 p.p., or about 0.5 standard deviations. Given that permissiveness increases more than the miss rate decreases, this implies that false positives are also increasing as throughput increases. This creates an interesting scenario; caseworkers with higher throughput are more permissive, which means that they have a lower miss rate but a higher false positive rate. The implication for accuracy therefore depends on the relative importance of false positive and false negative errors, β.[83]

Figure 11: Conditional Expectation Functions for Estimated Caseworker Impacts

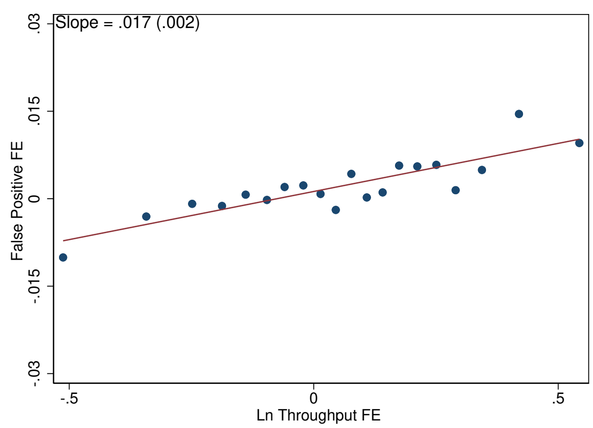
(a) Permissiveness and Ln(Throughput)



(b) Miss Rate and Ln(Throughput)



(c) Implied False Positive Rate and Ln(Throughput)



Notes: Plots the conditional expectation function between estimated caseworker impacts (i.e. fixed effects) on log throughput and either permissiveness, the miss rate, or the implied false positive rate according to Equation 6. All of these estimated impacts (fixed effects) are covariance shrunk. The standard deviation of estimated caseworker permissiveness and miss rate fixed effects are 4 p.p. and 2 p.p., respectively.

6.5 Other Analyses and Robustness

Differences in decision-making - In Appendix Section C I provide a more detailed discussion of what explains differences in manager decision-making. This ignores throughput

83. The intuition for this finding is highlighted in the heat plots in Appendix Figure 17, which shows the caseworker accuracy gradient over false positives and false negatives when $\beta = 0.1$ and when $\beta = 0.9$. Going from low- to high-throughput caseworkers means that on average the caseworkers are shifting from the upper left quadrant to the lower right quadrant of these plots, which could either increase or decrease accuracy depending on which plot you consider.

and focuses on the extent differences in manager accuracy and decision-making preferences describe variation in decision-making errors and permissiveness. I find that manager impacts on the false positive rate and false negative rate are negatively correlated, but this is sensitive to assumptions about the magnitude of the miss rate. I find that caseworker decision-making errors are even more negatively correlated, suggesting that differences in preferences likely play an important role in differences in caseworker decision-making. In both cases, there is also substantial variation in accuracy across workers.

Who are the different managers and what do they do differently? - In Appendix Section D I explore what managers with different performance and decision-making do differently and who are the different managers based on their experience, demographics, location, team composition, and other factors. I find a mild positive relationship between manager tenure and throughput as well as variation in throughput and accuracy within region and by urbanicity. For permissiveness, I find that Black managers have higher permissiveness than white managers. More permissive managers also award higher average SNAP benefit amounts, suggesting either that they are also more generous on the intensive margin or have better benefit targeting.

Mismeasuring the magnitude of false negatives - even if the the miss rate is a relevant measure of true false negatives, the magnitude of the differences in false negatives across managers and caseworkers could be mismeasured.[84] The magnitude of differences in manager impacts on the miss rate are important because they are used to infer differences in manager impacts on false positives and then accuracy. To address this, in Appendix Figure 20 I repeat my analyses scaling down the miss rate by 50% (0.5x) and then scaling up the miss rate by 100% (2.0x). This does not change the main finding that manager and caseworker throughput and accuracy are uncorrelated or relatively weakly correlated.[85]

84. It could be that when cases are improperly denied, only a randomly selected half of the improperly denied cases actually reapply. In this scenario, the miss rate will still be highly correlated with the true false negative rate, but the magnitude would be understated a factor of 2. On the other hand, it could be that half of denied applications that successfully reapply were correctly denied by the original caseworker, but these applicants improved their application prior to reapplying. In this case the miss rate would be overstated by a factor of 2.

85. The distribution of manager permissiveness and miss rate under these different assumptions are shown in Appendix Figures 13 and 14.

7 Implications of Manager Differences for Public Service Provision

In this section I illustrate the importance of manager differences for public service provision. Managers impact the quantity of public services provided via throughput, which has implications for the timeliness and administrative cost of public services. Managers impact the quality of public services via accuracy, which has important implications for the effectiveness of public benefit programs in directing benefits to those who are eligible. Finally, managers influence differences in decision-making regarding if public services are provided or not via permissiveness, which influence overall program size or cost. The size and cost of public benefit programs are relevant for taxpayers and also have important effects on households, businesses, and overall economic activity.

To do this, I measure how much outcomes would change if the bottom quartile of managers for a given outcome were shifted to the 75th percentile of performance.[86] Using throughput as an example, this would represents a hypothetical where the 205 managers with the lowest throughput contributions (fixed effects) were replaced with managers at the 75th percentile of throughput. I consider both how those manager's per manager-year performance would change as well as organization-wide performance. I compare the relative importance of managers and caseworkers by undertaking the same exercise for the bottom quartile of caseworkers, which represents 1,392 caseworkers. It is unclear whether shifting caseworkers or managers will have a bigger per-worker impact. Table 3 shows that caseworkers explain 2.5 to 5 times more of the variation in their own outcomes, but a manager's performance is relevant to a team of on average 6 caseworkers per manager in my data.[87] [88]

In this illustrative analysis I isolate one measure of manager performance at a time, whether that be throughput, accuracy, permissiveness, or the miss rate. When managers are replaced to improve one dimension of performance, this does not consider the consequences of how other dimensions might change. My findings from Section 6 show that higher manager throughput does not seem to be achieved at the cost of accuracy or by changing decision-making in other ways, suggesting that these concerns related to throughput may be of limited importance. However, manager impacts on the components of accuracy are or could be highly correlated.

86. I omit the first percentile to avoid any possible outliers, but this does not notably change the results.

87. This is smaller than the 10-11 caseworkers per team overall because I only use level 2 caseworkers in my analysis.

88. For overall outcomes for the 2018-2023 period, the amount of time each worker works as Texas HHS is also a factor. Caseworkers tend to have higher turnover than managers. The hiring patterns of managers and caseworkers during the period is relatively similar given that Texas HHS targets a constant number of caseworkers per manager.

Table 4 shows the impact on per-worker and organization-wide throughput, permit rate, miss rate, and accuracy from shifting the bottom quartile of managers or caseworkers for a given outcome to the 75th percentile.[89] Starting with throughput, shifting the lowest throughput managers to the 75th percentile of performance would increase the output of these managers by 3,428 cases per manager-year, a 30% increase. In aggregate, shifting the lowest throughput managers increases organization-wide output by 1.54 million cases reviewed, a 5.6% increase. This impact is about half as large as the 2.81 million impact from shifting the lowest throughput caseworkers but is achieved by shifting only 15% as many workers, which means that per worker managers have a 3.75 times greater impact than caseworkers.[90] Compared to the estimated importance of similar supervisory managers overseeing a comparable number of workers in a private sector setting from Lazear, Shaw, and Stanton (2015), my findings suggest that the value of these public sector managers relative to workers is similar if not higher.[91] To benchmark the relevance of increasing output by 1.54 million cases during this period, Texas HHS had at least a 1-2 month case backlog for large periods of 2021 and 2022 of between 0.8 and 1.2 million cases.[92] This led to significant delays for low-income households in receiving crucial benefits during this period.[93] While it may seem like case review demand is fixed and a large increase in output of this magnitude could not be realized, in practice Texas HHS during this period implemented several policies to reduce case review demand, mainly by waiving SNAP interviews and extending SNAP recertification periods. Hence the overall benefit of increased output should in part be interpreted through the value of policies that increase case review demand but provide greater screening or provide other services to applicants.

For accuracy assuming equal weight for false positive and false negative errors ($\beta = 0.5$), shifting the least accurate quartile of managers to the 75th percentile of accuracy would increase accuracy by 1.49 p.p. and would change overall accuracy by 0.36 p.p. from

89. More complete information is included in Appendix Tables 21 and 22.

90. Shifting approximately the bottom 760 caseworkers (13%) would achieve the same gain in output as shifting the bottom quartile of 205 managers.

91. Lazear, Shaw, and Stanton (2015) finds that replacing a manager from the 90th percentile with one at the 10th percentile would increase team output by about the same as adding one worker to a nine-member team. In my setting, the difference in the estimated log throughput fixed effect for managers at the 90th and 10th percentile is 0.302, which would increase an average caseworker's throughput from 180 to 243. For a team with at least 6 "standard" level 2 caseworkers, the better manager is about the same as adding two workers to a 6-member team.

92. Appendix Figure 9 shows the evolution of an estimate of the Texas HHS case backlog over time.

93. For example, in data reported on Texas HHS' website for SNAP applications for 2022 (https://www.hhs.texas.gov/about/records-statistics/data-statistics), less than 75% of SNAP initial applications were reviewed on time in 9 out of 12 months in 2022, with less than two-thirds reviewed on time in 6 out of 12 months. For recertifications, less than half were reviewed on time for several months in 2022 prior to recertifications being waived for a period of time.

Table 4: Impact of Shifting the Bottom Quartile of Managers or Caseworkers for a Given Outcome to the 75th Percentile

	Throughput (cases) (1)	Permit Rate (p.p.) (2)	Miss Rate (p.p.) (3)	Accuracy (p.p.) (4)
Per Worker-Year Impact				
Managers	3,428	3.94	1.89	1.49
Case Workers	1,107	8.65	3.02	2.48
Overall Impact				
Managers	$1.55*10^6$	0.77	0.43	0.36
Caseworkers	$2.81*10^6$	1.65	0.73	0.56
Manager-Caseworker Impact Ratio	3.75	3.17	4.00	4.37
Overall Baseline	$27.18*10^6$	56.46	6.47	

Notes: For a given outcome, reports the impact on that outcome of shifting the bottom quartile (omitting the first percentile) of workers (either managers or caseworkers) for that outcome to the 75th percentile of that outcome. This results in shifting either 205 managers or 1,392 caseworkers. Per worker-year impact on throughput is the increase in throughput per worker-year for the shifted workers. For the other outcomes this reflects the change in the rate for the shifted workers. The overall impact reports the total increase organization-wide for each measure from 2018 to 2023. The manager-caseworker impact ratio is the ratio of the overall impact per worker shifted for managers versus caseworkers. Overall baseline is the organization-wide outcome for the 2018 to 2023 period. This is not available for accuracy since assumptions were made to infer differences in impacts on accuracy but not levels of cases accurately reviewed.

2018 to 2023. This would represent 10,000 fewer errors over the time period, which have important costs on households that either need to reapply or do not receive benefits, on taxpayers via program costs, and on administrative costs due to the need to review additional reapplications and recertifications. While this measurement of accuracy relies on a variety of assumptions, it is informative to compare the impact relative to caseworkers making those same assumptions. The change in accuracy for managers is almost two-thirds of the gain from shifting the least accurate quartile of caseworkers, implying that the impact from shifting the lowest accuracy managers is 4.4 times greater than for caseworkers. This shows that differences in managers have an important role in impacting the quality of public service provision relative to workers.

For permissiveness, shifting the least permissive managers to the 75th percentile of permissiveness would increase those manager's permit rates by 3.94 p.p. (7.0%), which is on average an additional 503 permitted applications per manager year, and would increase organization-wide permit rates by 0.77 p.p. (1.4%). This is almost half as much as the change in organization wide permit rates achieved by shifting the least permissive quartile of caseworkers, implying that managers have a 3.2 times larger per worker impact than caseworkers. Using a back of the envelope calculation, I estimate that shifting the permissiveness of these 205 managers would generate at least \$406 million in additional program costs for tax payers from 2018-2023, or \$0.94 million per manager-year. This calculation is discussed in depth in Appendix Section E where I discuss heterogeneity in effects by program and case type, estimate the causal impact of managers on program participation, and quantify monthly per-applicant program costs for marginal applicants.[94] This illustrates that the differences in permissiveness between managers in the data translate into millions of dollars in variation in program cost.

Lastly, shifting the quartile of managers with the lowest miss rate to the 75th percentile of the miss rate would increase (worsen) those managers' miss rate by 1.89 p.p. and increase organization-wide permit rates by 0.43 p.p. (6.5%). This is on average a 4 times greater effect per worker than doing the same shift for the lowest miss rate caseworkers.

Together, these results suggest that differences in throughput across managers are important for the amount, timeliness, and administrative cost of public services, while differences in accuracy, the miss rate, and permissiveness translates into large differences in quality, program cost, and program size. This is true both in absolute terms and relative to the caseworkers that they manage. This suggests that policies targeting managers could be more effective than those targeting caseworkers at improving and standardizing bureaucratic performance, but this depends on a variety of unknown factors regarding how costly and feasible it would be to induce these changes for managers and caseworkers.

8 Conclusion

In this paper I explore the effect of mangers on the quantity and quality of public service provision. I show that managers explain 8-10% of the variation in quantity- and quality-based measures of caseworker performance. Next, I find that manager throughput (quantity) is not achieved at the cost of accuracy (quality). Instead I find that manager throughput and accuracy are uncorrelated, suggesting that there are important differences in both manager productivity and manager preferences. One implication of this is that a naive staffing policy that retains or removes workers based on throughput will not have unintended impacts on accuracy, or vice versa. Finally, I illustrate that manager differences have important

94. In short, I find that (i) manager impacts on permissiveness are relatively similar across the major case types, (ii) differences in manager permissiveness lead to persistent differences in long-term program participation at 6 months, and (iii) more permissive managers award if anything higher average SNAP benefit amounts. I cannot observe differences in cost for Medicaid participants, and therefore assume marginal applicants from the perspective of different manager decision-making incur average medical costs.

consequences for public service provision. For throughput and permissiveness separately, shifting the lowest quartile of managers to the 75th percentile of manager performance would increase organization-wide output by 5.6% and program costs by at least $406 million, or $0.94 million per manager-year.

These findings suggest that an important component of policy design should be focused on program administration, especially management. While differences in managers and workers are both important, I show that impacts of differences in management are 3-4.5 times larger than impacts of differences across workers and have important consequences for the efficacy of public policy. On the one hand, manager differences represent differences in manager productivity. Understanding how managers achieve different productivity or predicting which managers will be more productive is important. On the other hand, differences in decision-making across managers regarding production (quantity vs. quality) and case review decisions (false positive vs. false negative errors) suggest that managers are likely not functioning under a single socially optimal objective function, which is inefficient. This means that improving the administration of public benefit program policy by making middle-level managers more efficient and more standardized may have first-order benefits for public sector productivity and social welfare.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Adhvaryu, Achyuta, Vittorio Bassi, Anant Nyshadham, and Jorge Tamayo. 2019. "No Line Left Behind: Assortative Matching Inside the Firm" (September 23, 2019).

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2022. "Management and Shocks to Worker Productivity." *Journal of Political Economy* 130, no. 1 (January): 1–47.

Andrews, M. J., L. Gill, T. Schank, and R. Upward. 2008. "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society Series A: Statistics in Society* 171, no. 3 (June 1, 2008): 673–697.

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96, no. 1 (March): 127–151.

Arnold, David, Will Dobbie, and Peter Hull. 2022. "Measuring Racial Discrimination in Bail Decisions." *American Economic Review* 112, no. 9 (September): 2992–3038.

Autor, David, Andreas Kostøl, Magne Mogstad, and Bradley Setzler. 2019. "Disability Benefits, Consumption Insurance, and Household Labor Supply." *American Economic Review* 109, no. 7 (July): 2613–2654.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. 2021. "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats*." *The Quarterly Journal of Economics* 136, no. 4 (November 1, 2021): 2195–2242.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. "Cross-Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review* 103, no. 1 (February): 305–334.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. 2020. "The Glittering Prizes: Career Incentives and Bureaucrat Performance." *The Review of Economic Studies* 87, no. 2 (March 1, 2020): 626–655.

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *The Quarterly Journal of Economics* 118 (4): 1169–1208. Accessed April 17, 2024. JSTOR: 25053937.

Best, Michael Carlos, Jonas Hjort, and David Szakonyi. 2023. "Individuals and Organizations as Sources of State Effectiveness." *American Economic Review* 113, no. 8 (August): 2121–2167.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. “The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment*.” *The Quarterly Journal of Economics* 127, no. 3 (August 1, 2012): 1205–1242.

Bhargava, Saurabh, and Dayanand Manoli. 2015. “Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment.” *The American Economic Review* 105 (11): 3489–3529. JSTOR: 43821382.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. “Does Management Matter in Schools?” *The Economic Journal* 125, no. 584 (May 1, 2015): 647–674.

Boadway, Robin, Nicolas Marceau, and Motohiro Sato. 1999. “Agency and the Design of Welfare Systems.” *Journal of Public Economics* 73, no. 1 (July 1, 1999): 1–30.

Card, David, Jörg Heining, and Patrick Kline. 2013. “Workplace Heterogeneity and the Rise of West German Wage Inequality*.” *The Quarterly Journal of Economics* 128, no. 3 (August 1, 2013): 967–1015.

Chan, David C, Matthew Gentzkow, and Chuan Yu. 2022. “Selection with Variation in Diagnostic Skill: Evidence from Radiologists*.” *The Quarterly Journal of Economics* 137, no. 2 (May 1, 2022): 729–783.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. “Productivity Dispersion in Medicine and Manufacturing.” *American Economic Review* 106, no. 5 (May): 99–103.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.” *American Economic Review* 104, no. 9 (September): 2593–2632.

———. 2014b. “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood.” *American Economic Review* 104, no. 9 (September): 2633–2679.

Choudhury, Prithwiraj, Tarun Khanna, and Christos A Makridis. 2020. “Do Managers Matter? A Natural Experiment from 42 R&D Labs in India.” *The Journal of Law, Economics, and Organization* 36, no. 1 (March 1, 2020): 47–83.

Cook, Jason B., and Chloe N. East. 2023. “The Effect of Means-Tested Transfers on Work: Evidence from Quasi-Randomly Assigned SNAP Caseworkers.” Pre-published, June. Working Paper. National Bureau of Economic Research: 31307.

Currie, Janet. 2006. “The take-up of social benefits.” In *Public Policy and the Distribution of Income,* 80–148. Russell Sage Foundation.

De Stefano, Federica, Matthew Bidwell, and Arnaldo Camuffo. 2022. “Do Managers Capture the Value They Create? Drivers of Managers’ Value Capture in a Large Retail Chain.” *STRATEGIC MANAGEMENT JOURNAL* (Hoboken) 43, no. 10 (October): 1983–2011.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy* 11, no. 4 (November): 213–248.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108, no. 2 (February): 201–240.

Elzayn, Hadi, Evelyn Smith, Thomas Hertz, Cameron Guage, Arun Ramesh, Robin Fisher, Daniel E Ho, and Jacob Goldin. 2024. "Measuring and Mitigating Racial Disparities in Tax Audits*." *The Quarterly Journal of Economics* (September 26, 2024): qjae027.

Feigenberg, Benjamin, and Conrad Miller. 2022. "Would Eliminating Racial Disparities in Motor Vehicle Searches Have Efficiency Costs?*." *The Quarterly Journal of Economics* 137, no. 1 (February 1, 2022): 49–113.

Fenizia, Alessandra. 2022. "Managers and Productivity in the Public Sector." *Econometrica* 90 (3): 1063–1084.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence From Patient Migration*." *The Quarterly Journal of Economics* 131, no. 4 (November 1, 2016): 1681–1726.

Finkelstein, Amy, and Matthew J Notowidigdo. 2019. "Take-Up and Targeting: Experimental Evidence from SNAP*." *The Quarterly Journal of Economics* 134, no. 3 (August 1, 2019): 1505–1556.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98, no. 1 (March): 394–425.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2013. "Long-Term Effects of Class Size *." *The Quarterly Journal of Economics* 128, no. 1 (February 1, 2013): 249–285.

Ganong, Peter, and Jeffrey B. Liebman. 2018. "The Decline, Rebound, and Further Rise in SNAP Enrollment: Disentangling Business Cycle Fluctuations and Policy Changes." *American Economic Journal: Economic Policy* 10, no. 4 (November): 153–176.

Gianella, Eric, Tatiana Homonoff, Gwen Rino, and Jason Somerville. 2022. "Administrative Burden and Procedural Denials: Experimental Evidence from SNAP." *Working Paper* (July).

Giardili, Soledad, Kamalini Ramdas, and Jonathan W. Williams. 2023. "Leadership and Productivity: A Study of U.S. Automobile Assembly Plants." *Management Science* 69, no. 3 (March): 1500–1517.

Homonoff, Tatiana, and Jason Somerville. 2021. "Program Recertification Costs: Evidence from SNAP." *American Economic Journal: Economic Policy* 13, no. 4 (November): 271–298.

Janke, Katharina, Carol Propper, and Raffaella Sadun. 2019. "The Impact of CEOs in the Public Sector: Evidence from the English NHS." Pre-published, May. Working Paper. National Bureau of Economic Research: 25853.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." Pre-published, December. Working Paper. National Bureau of Economic Research: 14607.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors *." *The Quarterly Journal of Economics* 131, no. 1 (February 1, 2016): 219–271.

Kleven, Henrik Jacobsen, and Wojciech Kopczuk. 2011. "Transfer Program Complexity and the Take-Up of Social Benefits." *American Economic Journal. Economic Policy* (Nashville, United States) 3, no. 1 (February): 54–90.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-Out Estimation of Variance Components." *Econometrica* 88 (5): 1859–1898.

Lazear, Edward P., Kathryn L. Shaw, and Christopher T. Stanton. 2015. "The Value of Bosses." *Journal of Labor Economics* 33, no. 4 (October): 823–861.

Limodio, Nicola. 2021. "Bureaucrat Allocation in the Public Sector: Evidence from the World Bank." *The Economic Journal* 131, no. 639 (October 1, 2021): 3012–3040.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103, no. 5 (August): 1797–1829.

Metcalfe, Robert D., Alexandre B. Sollaci, and Chad Syverson. 2023. "Managers and Productivity in Retail." Pre-published, April. Working Paper. National Bureau of Economic Research: 31192.

Mulhern, Christine. 2023. "Beyond Teachers: Estimating Individual School Counselors' Effects on Educational Attainment." *American Economic Review* 113, no. 11 (November): 2846–2893.

Munoz, Pablo, and Cristobal Otero. 2022. "Managers and Public Hospital Performance." *Working Paper* (December).

Muñoz, Pablo, and Mounu Prem. 2024. "Managers' Productivity and Recruitment in the Public Sector." *American Economic Journal: Economic Policy.*

Prendergast, Canice. 2007. "The Motivation and Bias of Bureaucrats." *American Economic Review* 97, no. 1 (March): 180–196.

Rasul, Imran, and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *The Economic Journal* 128, no. 608 (February 1, 2018): 413–446.

Rossin-Slater, Maya. 2013. "WIC in Your Neighborhood: New Evidence on the Impacts of Geographic Access to Clinics." *Journal of Public Economics* 102 (June 1, 2013): 51–69. Accessed July 29, 2022.

Syverson, Chad. 2004. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy* 112, no. 6 (December): 1181–1222.

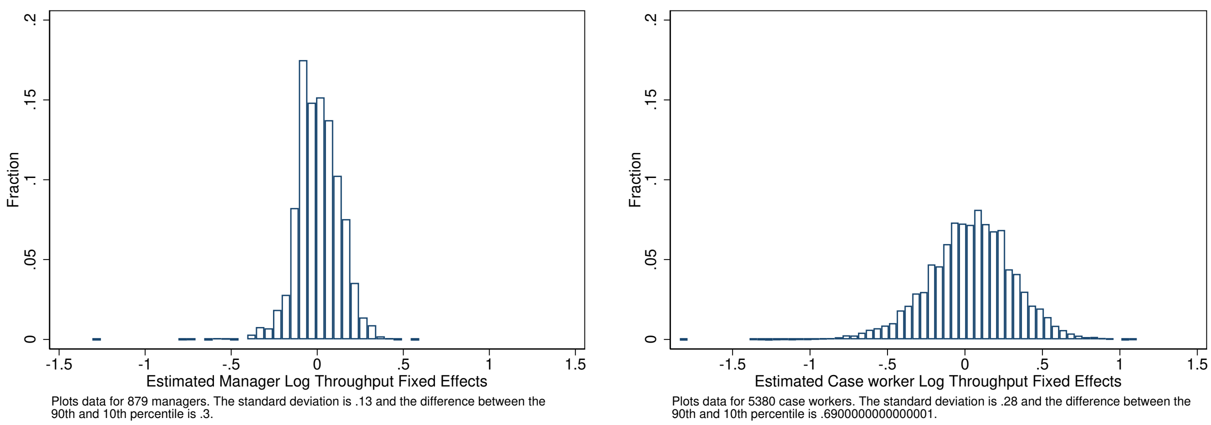
———. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49, no. 2 (June): 326–365.

Wu, Derek, and Bruce Meyer. 2022. "Certification and Recertification in Welfare Programs: What Happens When Automation Goes Wrong?" *Working Paper* (August 27, 2022).

A Appendix Figures

Figure 1: Distribution of Log Throughput Estimated AKM Model Fixed Effects

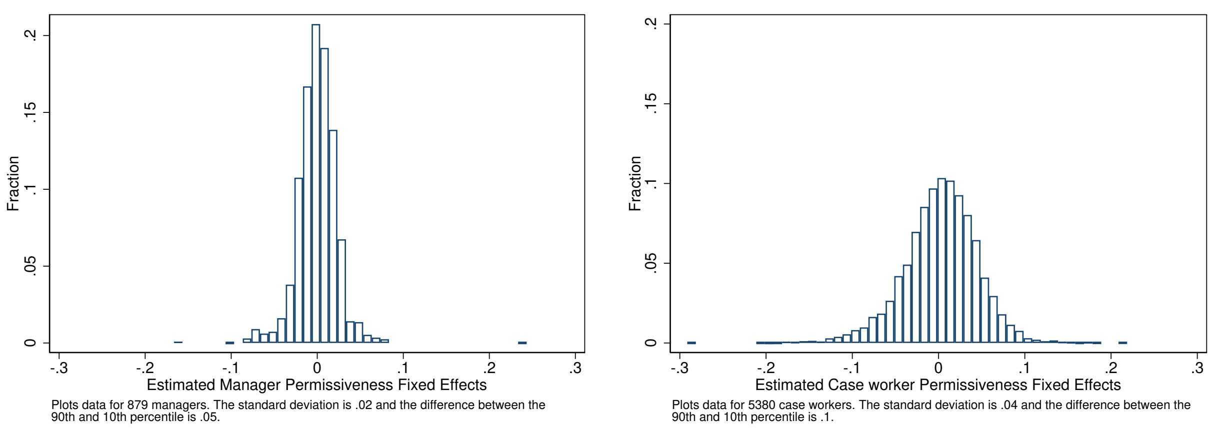
(a) Manager Fixed Effect **(b)** Caseworker Fixed Effect



Notes: Shows the distribution of estimated log throughput fixed effects from Equation 1 with the covariance shrinkage approach.

Figure 2: Distribution of Permissiveness Estimated AKM Model Fixed Effects

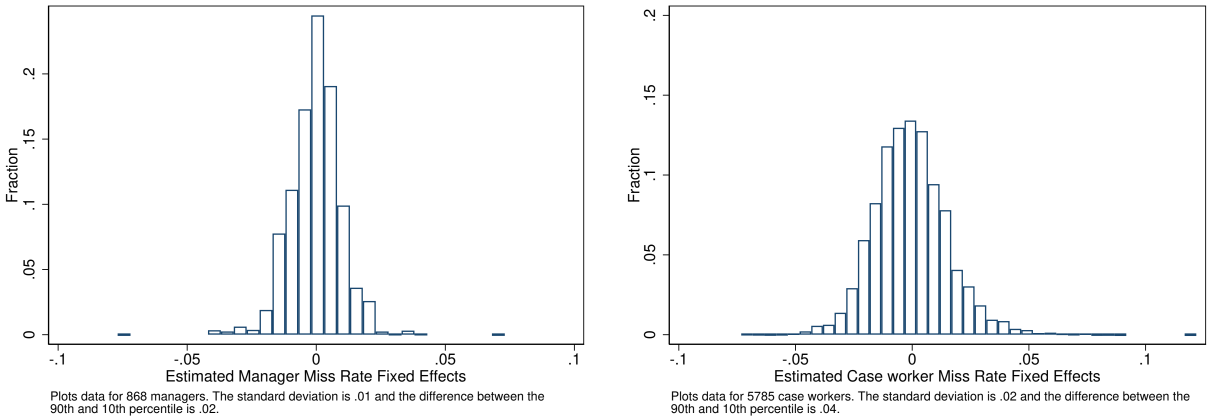
(a) Manager Fixed Effect **(b)** Caseworker Fixed Effect



Notes: Shows the distribution of estimated permissiveness fixed effects from Equation 1 with the covariance shrinkage approach.

Figure 3: Distribution of the Miss Rate Estimated AKM Model Fixed Effects

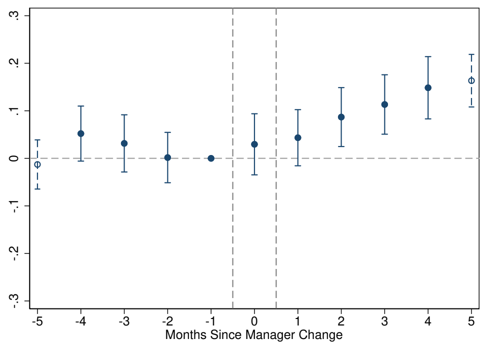
(a) Manager Fixed Effect **(b)** Caseworker Fixed Effect



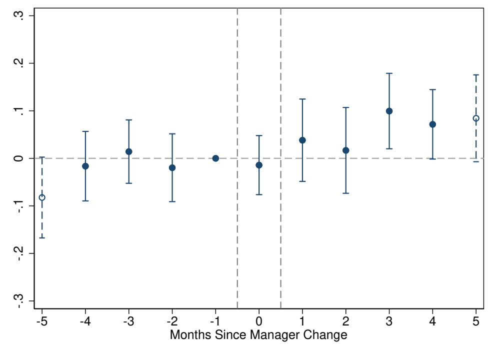
Notes: Shows the distribution of estimated miss rate fixed effects from Equation 1 with the covariance shrinkage approach.

Figure 4: Event Study Using "Within Event" Event Time Leave-Out

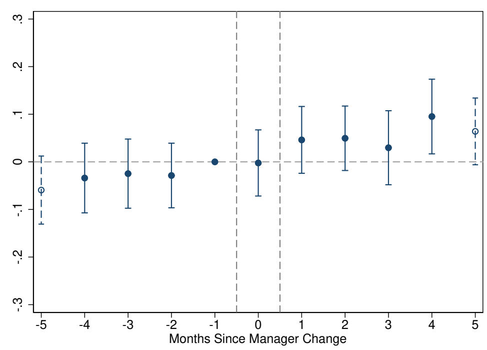
(a) Ln(Throughput) (in s.d.)



(b) Permissiveness (in s.d.)



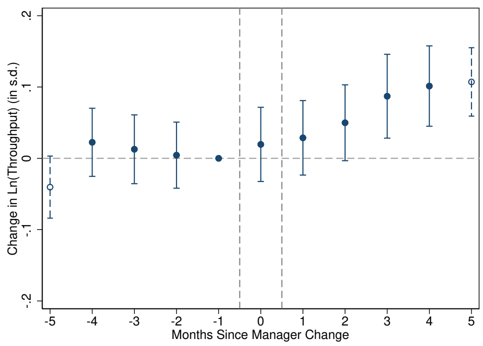
(c) Miss Rate (in s.d.)



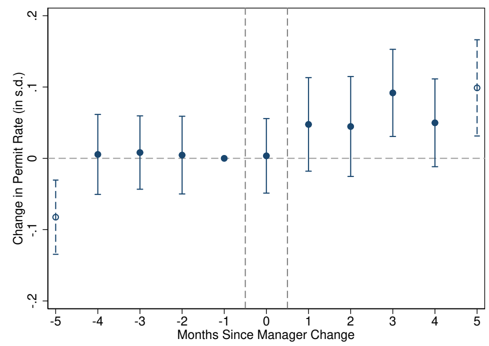
Notes: Estimates a separate regression for each event study point k using a caseworker specific i leave out that omits event time data within 5 periods of k and -1 for any caseworker experiencing the same change in manager event as caseworker i. The 5 period window is based on the level of autocorrelation observed across months. The regression is a long difference change in outcome between event time k and -1 denoted $\Delta y_i^k = \pi_0^k + \pi_1^k \widehat{\Delta M}_i^{L,k} + \Delta\epsilon_{it}$, where "L" denotes "leave-out" and $\widehat{\Delta M}_i^{L,k} = \widehat{\theta}_{i,incoming}^{L,k} - \widehat{\theta}_{i,outgoing}^{L,k}$ is the leave-out change in manager quality between event time k and event time -1. Standard errors are bootstrapped.

Figure 5: Event Study Using Event Group Leave-Out

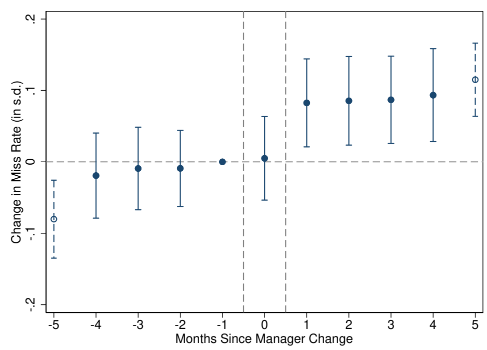
(a) Ln(Throughput) (in s.d.)



(b) Permissiveness (in s.d.)



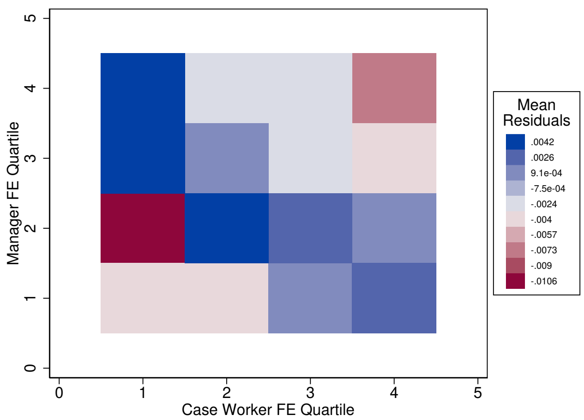
(c) Miss Rate (in s.d.)



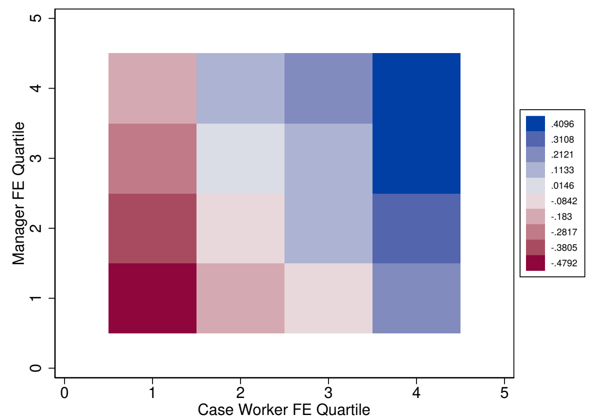
Notes: Event study that instead of a caseworker leave-out has a caseworker-specific event group leave-out. For each caseworker i, this omits caseworker i and any caseworkers experiencing the same change in manager event as caseworker i from periods $k = -8$ to $k = 8$ from the estimation of Equation 1 when obtaining the independent variable $\Delta\widehat{M_i}$ for the event study. Standard errors are bootstrapped.

Figure 6: Caseworker and Manager Match Robustness Plots: Log Throughput

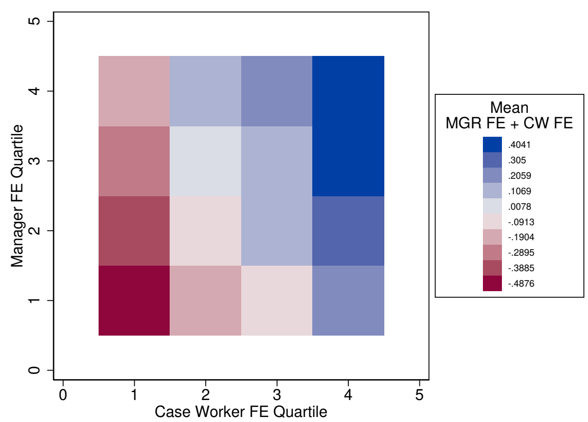
(a) Mean AKM residual



(b) Outcome Mean



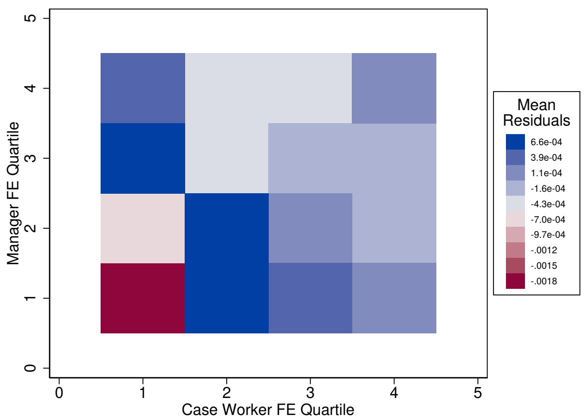
(c) Mean Caseworker + Manager FE



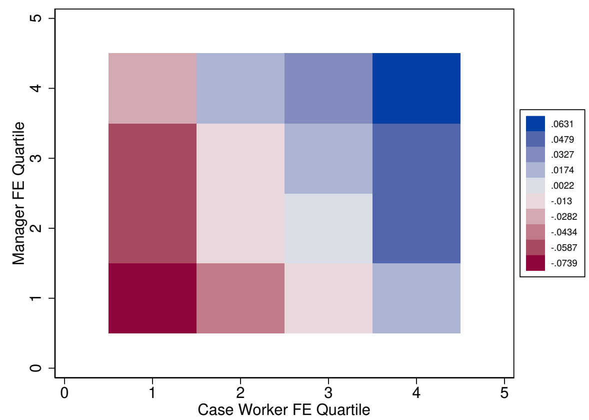
Notes: Each subplot divides caseworker month observations in the data into a grid using estimated caseworker and manager fixed effect quartiles for the outcome variable. This uses standard fixed effects without shrinkage corrections. Subplot (a) plots the mean residual from Equation 1 in each cell. Subplot (b) plots the mean of the outcome variable for outcomes in each cell. Subplot (c) plots the average value of the caseworker FE added to the manager FE in each cell.

Figure 7: Caseworker and Manager Match Robustness Plots: Permissiveness

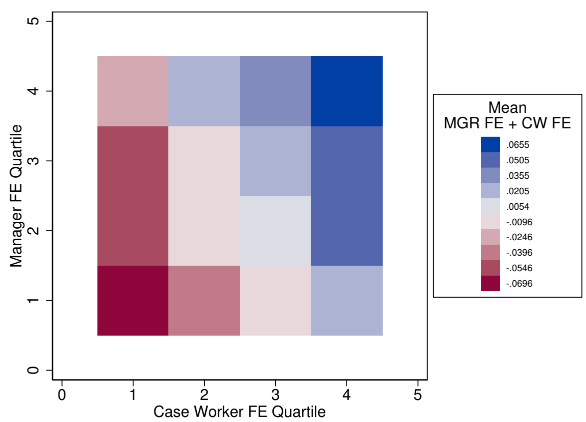
(a) Mean AKM residual



(b) Outcome Mean



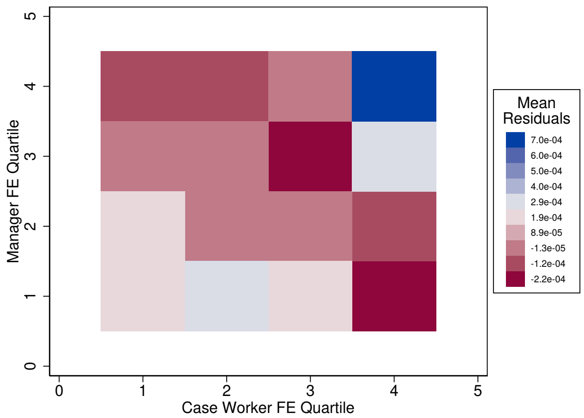
(c) Mean Caseworker + Manager FE



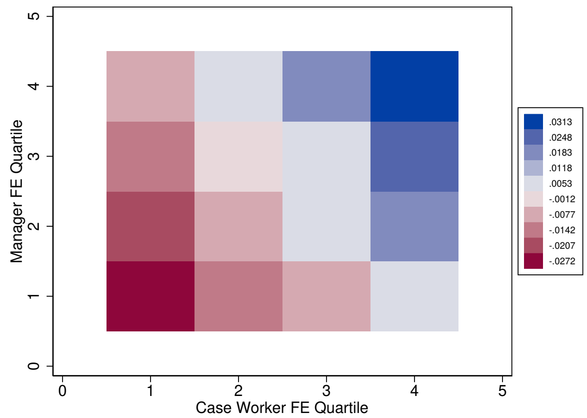
Notes: Each subplot divides caseworker month observations in the data into a grid using estimated caseworker and manager fixed effect quartiles for the outcome variable. This uses standard fixed effects without shrinkage corrections. Subplot (a) plots the mean residual from Equation 1 in each cell. Subplot (b) plots the mean of the outcome variable for outcomes in each cell. Subplot (c) plots the average value of the caseworker FE added to the manager FE in each cell.

Figure 8: Caseworker and Manager Match Robustness Plots: Miss Rate

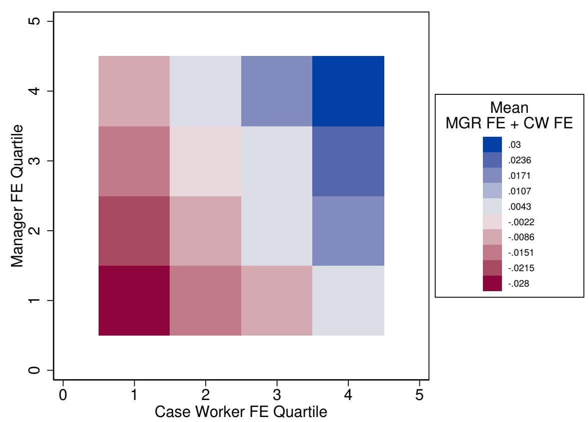
(a) Mean AKM residual



(b) Outcome Mean

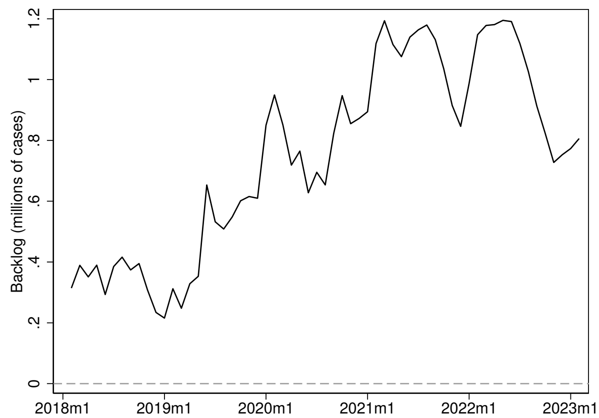


(c) Mean Caseworker + Manager FE



Notes: Each subplot divides caseworker month observations in the data into a grid using estimated caseworker and manager fixed effect quartiles for the outcome variable. This uses standard fixed effects without shrinkage corrections. Subplot (a) plots the mean residual from Equation 1 in each cell. Subplot (b) plots the mean of the outcome variable for outcomes in each cell. Subplot (c) plots the average value of the caseworker FE added to the manager FE in each cell.

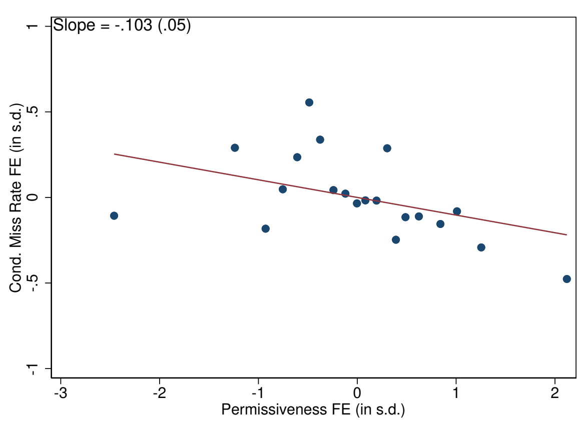
Figure 9: Texas HHS Case Backlog (in millions of cases)



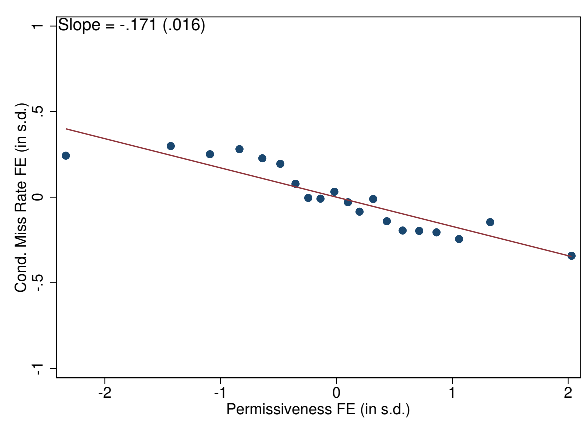
Notes: Shows the backlog of initial applications and recertifications for Texas HHS using data available for January 2018 to September 2023. A case is part of the backlog in month m if it was not disposed in the month that it was filed by the applicant. This only includes cases filed January 2018 or later and assumes that the majority of cases filed in February 2023 have been disposed by September 2023 and therefore appear in the data. The backlog starts at about 400,000 cases in 2018 and os measured using initial application and recertifications for Medicaid, MEPD, SNAP, TANF. This is relative to workers reviewing about 0.5 million cases per month.

Figure 10: Relationship Between Impacts on the Conditional Miss Rate and Permissiveness: Managers and Caseworkers

(a) Managers



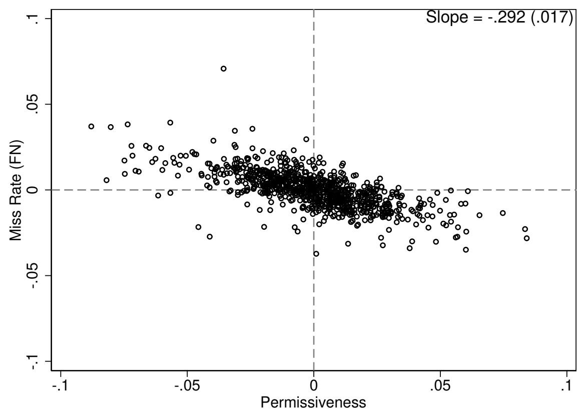
(b) Caseworkers



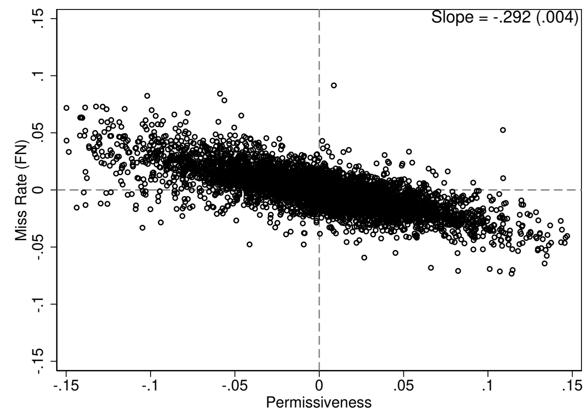
Notes: These plots show the conditional expectation function between estimated conditional miss rate fixed effects and permissiveness fixed effects, for managers and caseworkers. The conditional miss rate and permissiveness impacts are estimated from the AKM model in Equation 1 with covariance shrinkage adjustment. Both outcomes are standardized

Figure 11: Joint Distribution of Estimated Impacts for Miss Rate and Permissiveness: Managers and Caseworkers

(a) Managers



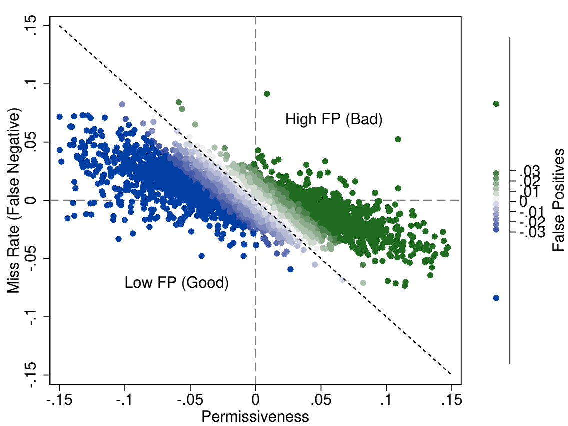
(b) Caseworkers



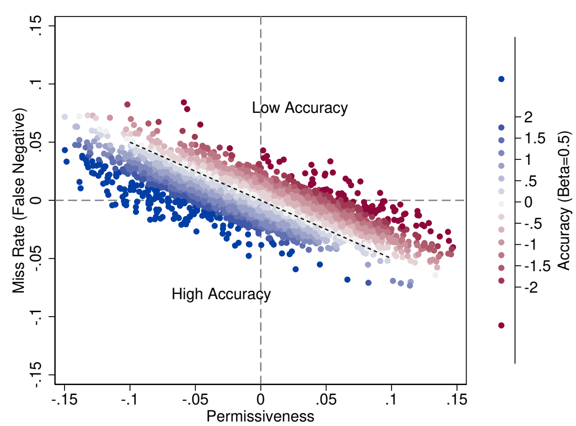
Notes: These plots show both the joint distribution of manager and caseworker impacts on permissiveness and the miss rate. The miss rate and permissiveness impacts are estimated from the AKM model in Equation 1 with covariance shrinkage adjustment.

Figure 12: Caseworker Impact Heat Map Plots in Miss Rate and Permissiveness Space: False Positives an Accuracy ($\beta = 0.5$)

(a) False Positives



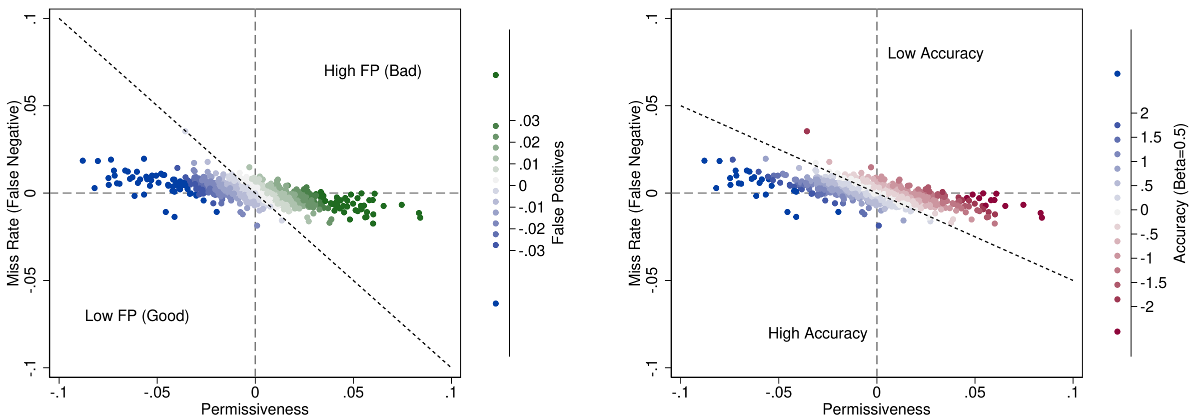
(b) Accuracy



Notes: These plots show both the joint distribution of caseworker impacts on permissiveness and the miss rate and uses the heat map gradient to show how these translate into false positives and accuracy. The miss rate and permissiveness impacts are estimated from the AKM model in Equation 1 with covariance shrinkage adjustment. False positives are determined based on Equation 6 and accuracy based on 7 with the assumption that $\beta = 0.5$.

Figure 13: Manager Impact Heat Map Plots in Miss Rate and Permissiveness Space With Scaled Down Miss Rate (0.5x): False Positives an Accuracy ($\beta = 0.5$)

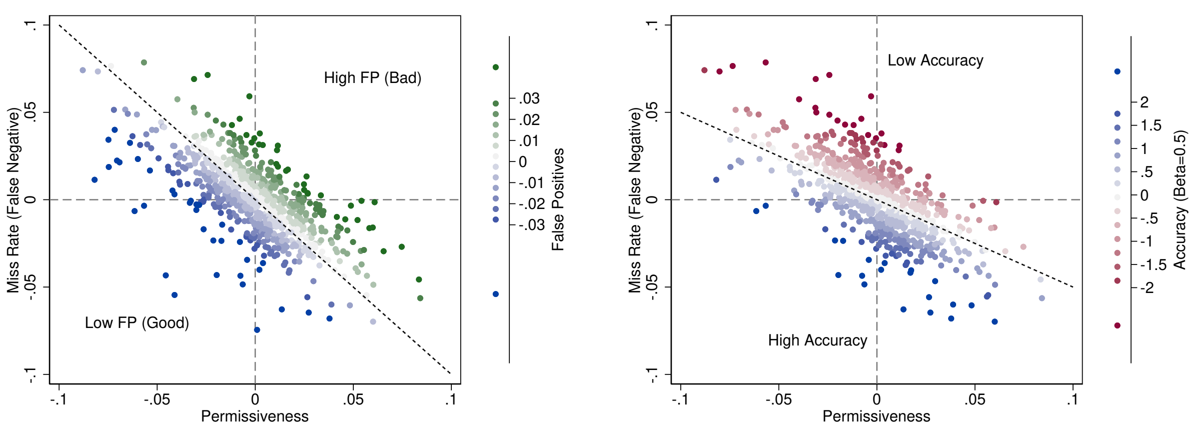
(a) False Positives **(b)** Accuracy



Notes: These plots show both the joint distribution of manager impacts on permissiveness and the miss rate after scaling the miss rate down by 50%. It uses the heat map gradient to show how these translate into false positives and accuracy. The miss rate and permissiveness impacts are estimated from the AKM model in Equation 1 with covariance shrinkage adjustment. False positives are determined based on Equation 6 and accuracy based on 7 with the assumption that $\beta = 0.5$.

Figure 14: Manager Impact Heat Map Plots in Miss Rate and Permissiveness Space With Scaled Up Miss Rate (2x): False Positives an Accuracy ($\beta = 0.5$)

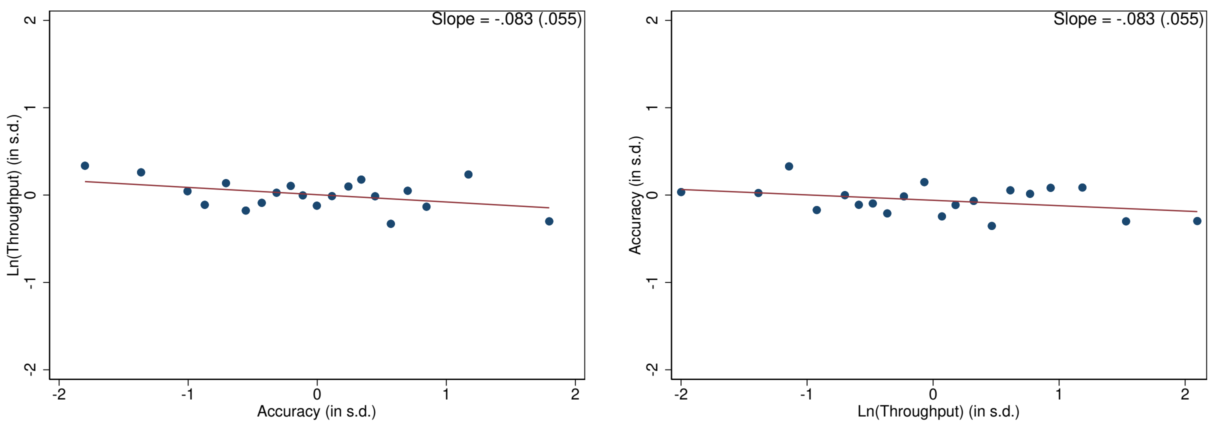
(a) False Positives **(b)** Accuracy



Notes: These plots show both the joint distribution of manager impacts on permissiveness and the miss rate after scaling up the miss rate by a factor of 2. It uses the heat map gradient to show how these translate into false positives and accuracy. The miss rate and permissiveness impacts are estimated from the AKM model in Equation 1 with the covariance shrinkage adjustment. False positives are determined based on Equation 6 and accuracy based on 7 with the assumption that $\beta = 0.5$.

Figure 15: Relationship Between Manager Impacts on Ln(Throughput) and Accuracy: Conditional Expectation Functions ($\beta = 0.5$)

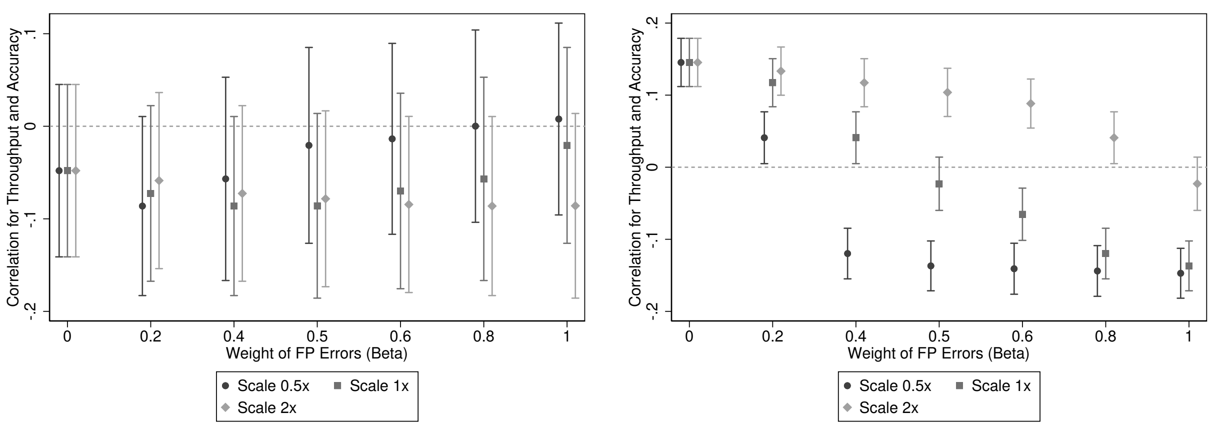
(a) Ln(Throughput) and Accuracy **(b)** Accuracy and Ln(Throughput)



Notes: Shows the conditional expectation function for manager log throughput and accuracy, both standardized. This confirms the relationship between manager log throughput and accuracy impacts is not nonlinear.

Figure 16: Relationship Between Impacts on Ln(Throughput) and Accuracy Using the Reapplication Rate: Reapplication Rate Magnitude Robustness

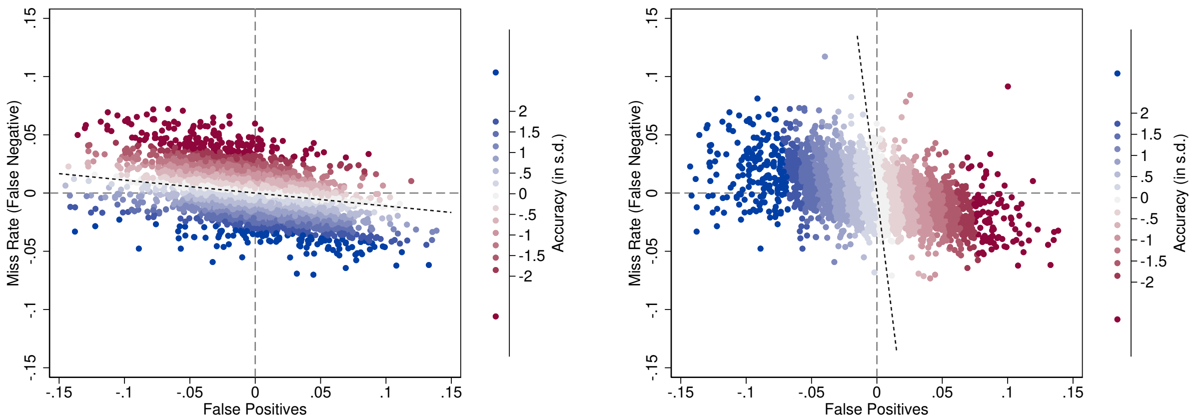
(a) Managers **(b)** Caseworkers



Notes: Shows the estimated correlation between manager and caseworker throughput and accuracy using the reapplication rate instead of the miss rate. Measures impacts for different relative cost of errors β and after scaling up and down the reapplication rate.

Figure 17: Caseworker Accuracy Gradient in False Positive and False Negative Space

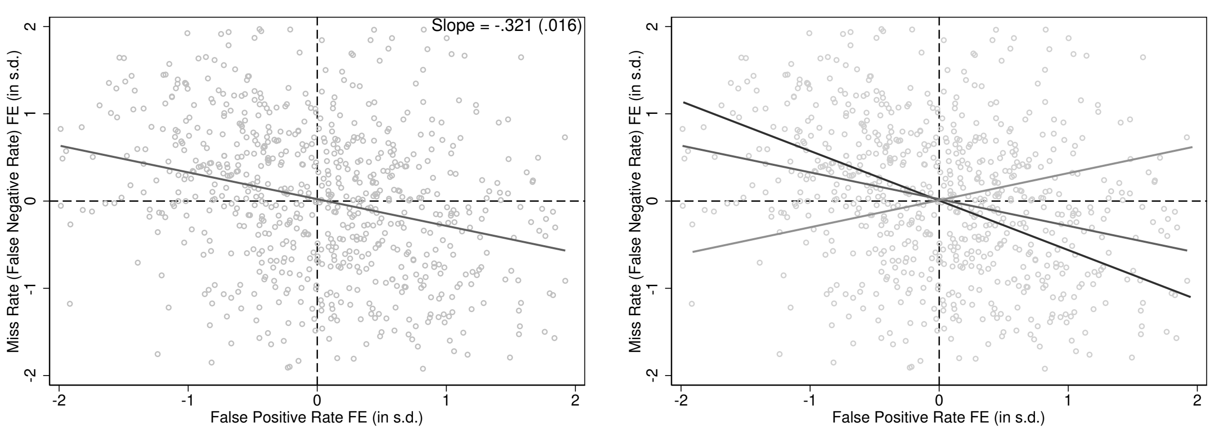
(a) Accuracy Gradient: $\beta = 0.1$ **(b)** Accuracy Gradient: $\beta = 0.9$



Notes: These plots show both the joint distribution of caseworker impacts on the miss rate from the estimation of the AKM model in Equation 1 and the implied impact on false positives determined based on Equation 6. The gradient on both subplots is the impact of caseworkers on accuracy (standardized) calculated based on 7 with the assumption that $\beta = .1$ or $\beta = .9$.

Figure 18: Empirical Distribution of Manager Impacts on False Negative Rate and False Positive Rate

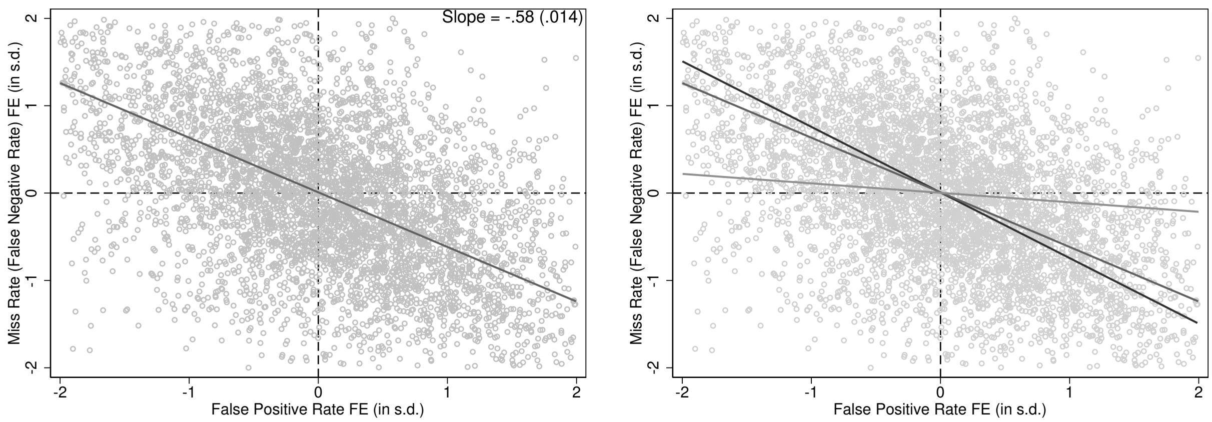
(a) Miss Rate Scale = 1.0x **(b)** Miss Rate Scaling from 0.5x to 2x



Notes: The scatterplot in both subplots is the plot of standardized miss rate (false negative rate) and false positive rate impacts of managers. Miss rate impacts are the estimated manager impacts (fixed effects) for miss rate from the AKM model in Equation 1 with covariance shrunk estimates. False positives is obtained from the same impacts for permissiveness and the miss rate as defined in 6. The line in subplot (a) is the linear best fit between the two variables. The lines in subplot (b) show the linear best fit under different assumptions for the scaling of the miss rate. The darkest line with the most negative slope reflects the scenario when the miss rate need to be scaled up by a factor of 2 to reflect false negatives and the lightest line with the most positive slope reflects the scenario when miss rate need to be scaled down by a factor of 0.5 to reflect false negatives.

Figure 19: Empirical Distribution of Caseworker Impacts on False Negative Rate and False Positive Rate

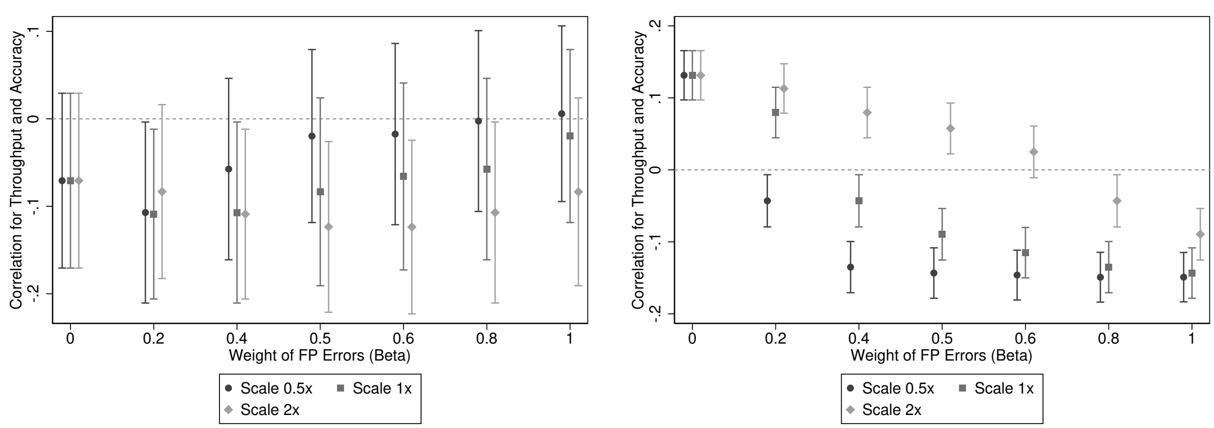
(a) Miss Rate Scale = 1.0x **(b)** Miss Rate Scaling from 0.5x to 2x



Notes: The scatterplot in both subplots is the plot of standardized miss rate (false negative rate) and false positive rate impacts of caseworkers. Miss rate impacts are the estimated caseworker impacts (fixed effects) for miss rate from the AKM model in Equation 1 with covariance shrunk estimates. False positives is obtained from the same impacts for permissiveness and the miss rate as defined in 6. The line in subplot (a) is the linear best fit between the two variables. The lines in subplot (b) show the linear best fit under different assumptions for the scaling of the miss rate. The darkest line with the most negative slope reflects the scenario when the miss rate need to be scaled up by a factor of 2 to reflect false negatives and the lightest line with the most positive slope reflects the scenario when miss rate need to be scaled down by a factor of 0.5 to reflect false negatives.

Figure 20: Relationship Between Impacts on Ln(Throughput) and Accuracy: Miss Rate Magnitude Robustness

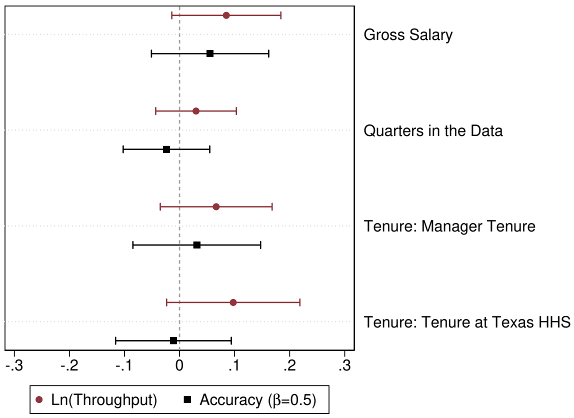
(a) Managers **(b)** Caseworkers



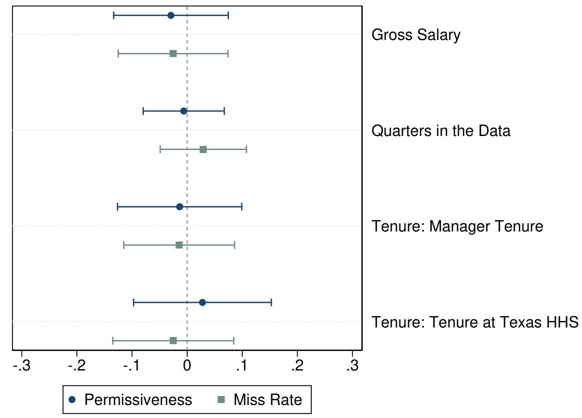
Notes: Plots the coefficient of a regression of standardized worker log throughput impacts on standardized worker accuracy impacts, representing the correlation between the outcomes. This considers different assumptions for β and scaling miss rate magnitudes up or down by a factor of 2. The impact on throughput is the estimated fixed effect from the AKM Model estimated using Equation 1, while the impact on accuracy is based on the estimated fixed effects for permissiveness and the miss rate according to Equation 7. These relationships are estimated on the full caseworker-month data clustering by either caseworker or manager.

Figure 21: Correlation Between Manager Experience and Manager Impacts

(a) Throughput and Accuracy



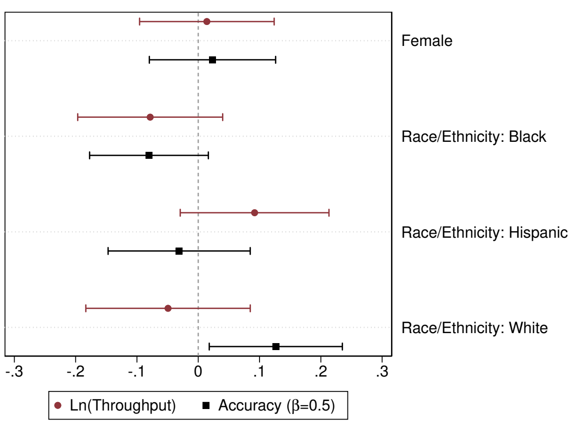
(b) Permissiveness and Miss Rate



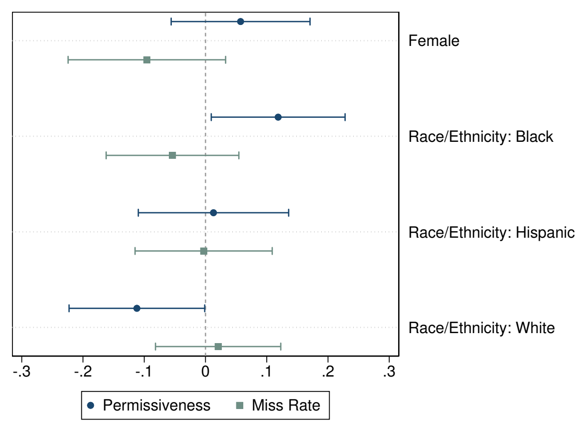
Notes: Plots the correlation between manager impacts and explanatory factors of interest. The manager impacts and explanatory factors are both standardized and should be interpreted in standard deviations. These relationships are estimated on the full caseworker-month data clustering by either caseworker or manager.

Figure 22: Correlation Between Manager Demographics and Manager Impacts

(a) Throughput and Accuracy



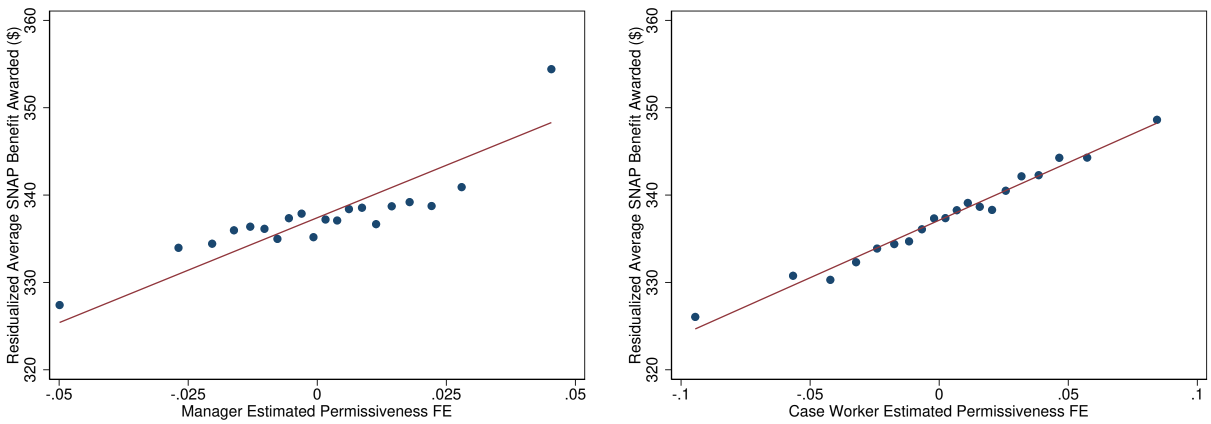
(b) Permissiveness and Miss Rate



Notes: Plots the correlation between manager impacts and explanatory factors of interest. The manager impacts and explanatory factors are both standardized and should be interpreted in standard deviations. These relationships are estimated on the full caseworker-month data clustering by either caseworker or manager.

Figure 23: Correlation Between Average SNAP Benefit Awarded and Manager Permissiveness Impact

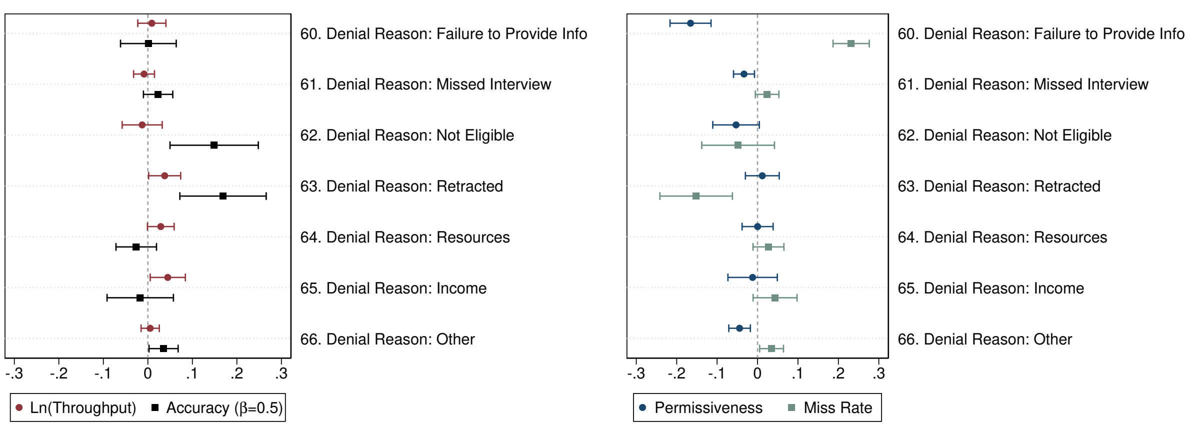
(a) Managers **(b)** Caseworkers



Notes: Shows the conditional expectation function between residualized average SNAP benefit amounts renormalized using the sample mean relative to differences in estimated manager or caseworker impacts for permissiveness.

Figure 24: Correlation Between Denial Reasons and Manager Impacts

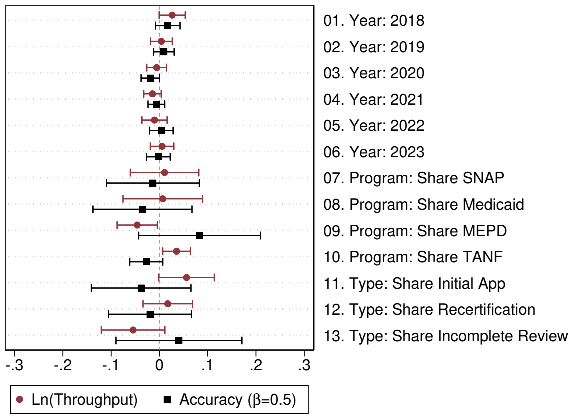
(a) Throughput and Accuracy **(b)** Permissiveness and Miss Rate



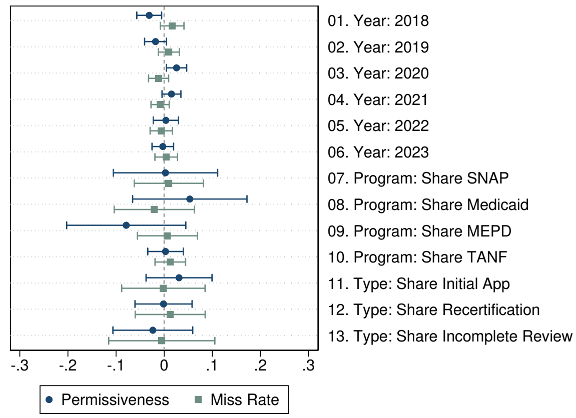
Notes: Plots the correlation between manager impacts and explanatory factors of interest. The manager impacts and explanatory factors are both standardized and should be interpreted in standard deviations. These relationships are estimated on the full caseworker-month data clustering by either caseworker or manager Failure to provide information means that a case was not complete enough to make a determination on its status, meaning it was missing relevant information. SNAP and TANF cases often have interviews, which if missed is a reason for denial. A case being denied for "not eligible" means that the composition of the household members made them ineligible for the benefit. A retracted case is retracted by the applicant, but this is very rare. Denials based on having too many resources (assets) or too much income are as expected..

Figure 25: Correlation Between Case Characteristics and Manager Impacts

(a) Throughput and Accuracy



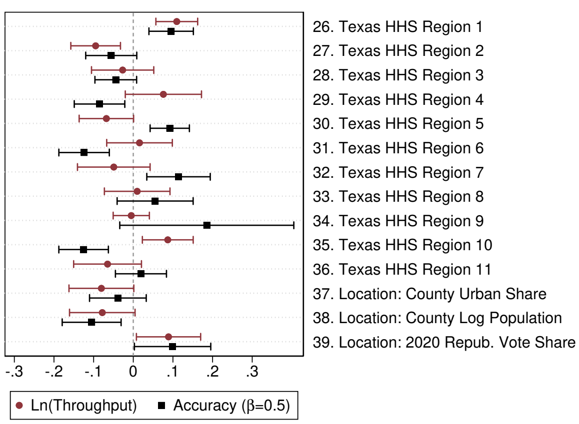
(b) Permissiveness and Miss Rate



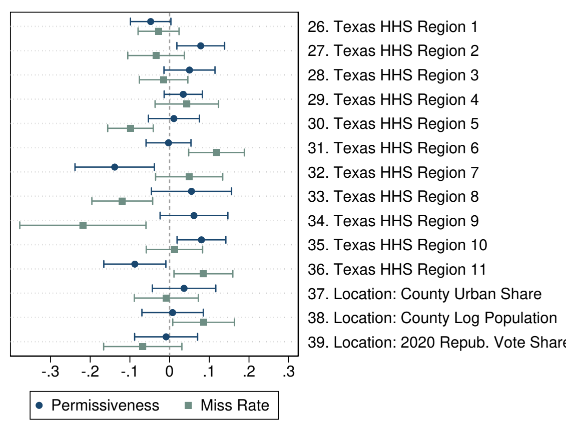
Notes: Plots the correlation between manager impacts and explanatory factors of interest. The manager impacts and explanatory factors are both standardized and should be interpreted in standard deviations. These relationships are estimated on the full caseworker-month data clustering by either caseworker or manager. The year factors are indicators for observations within a given year.

Figure 26: Correlation Between Administration Region/Location and Manager Impacts

(a) Throughput and Accuracy

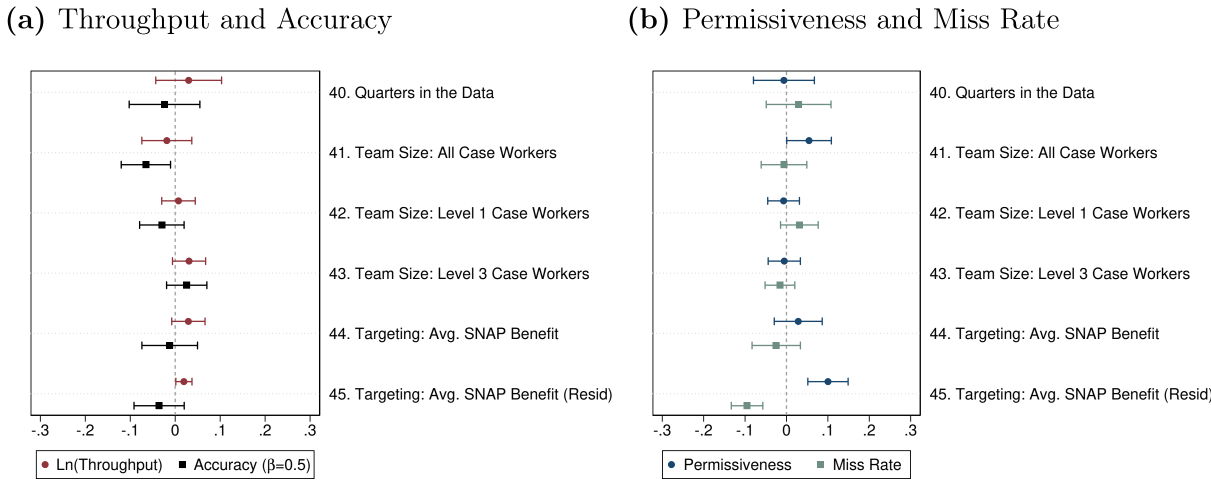


(b) Permissiveness and Miss Rate



Notes: Plots the correlation between manager impacts and explanatory factors of interest. The manager impacts and explanatory factors are both standardized and should be interpreted in standard deviations. These relationships are estimated on the full caseworker-month data clustering by either caseworker or manager.

Figure 27: Correlation Between Other Factors and Manager Impacts



Notes: Plots the correlation between manager impacts and explanatory factors of interest. The manager impacts and explanatory factors are both standardized and should be interpreted in standard deviations. These relationships are estimated on the full caseworker-month data clustering by either caseworker or manager.

B Appendix Tables

Table 1: Analysis of Variance: Log Throughput and Permissiveness

	Ln(Throughput)			Permissiveness		
	(1)	(2)	(3)	(4)	(5)	(6)
Observations	158,734	158,525	158,517	158,821	158,613	158,605
R-Squared	0.076	0.381	0.399	0.095	0.355	0.377
Adjusted R-Squared	0.071	0.357	0.371	0.089	0.329	0.348
Caseworker FE		X	X		X	X
Manager FE	X		X	X		X

Notes: Shows the adjusted R^2 of regressions of caseworker outcomes on fixed effects, with Columns (3) and (6) corresponding to Equation 1. Each column either uses caseworker residualized log throughput or permissiveness as the dependent variable. The fixed effects included are noted below. Given that the dependent variable has already been residualized, there are no time or case composition controls. The same table for the miss rate is included in Appendix Table 2.

Table 2: Analysis of Variance: Miss Rate and Reapplication Rate

	Miss Rate			Reapplication Rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Observations	158,821	158,613	158,605	158,821	158,613	158,605
R-Squared	0.079	0.254	0.272	0.07	0.231	0.249
Adjusted R-Squared	0.074	0.224	0.239	0.065	0.201	0.214
Case Worker FE		X	X		X	X
Manager FE	X		X	X		X

Notes: Each column is a regression with caseworker residualized miss rate or the reapplication rate. Columns (3) and (6) correspond with Equation 1. Given that the dependent variable has already been residualized, there are no time or case composition controls.

Table 3: Analysis of Variance with Saturated Model: Log Throughput and Permissiveness:

	Ln(Throughput)			Permissiveness		
	(1)	(2)	(3)	(4)	(5)	(6)
Observations	158,525	158,517	156,842	158,613	158,605	156,926
R-Squared	0.381	0.399	0.447	0.355	0.377	0.425
Adjusted R-Squared	0.357	0.371	0.4	0.329	0.348	0.377
Case Worker FE	X	X		X	X	
Manager FE		X			X	
Case Worker x Manager FE			X			X
Office FE						

Notes: Each column is a regression with caseworker residualized log throughput or permissiveness on the left hand side and different fixed effects on the right hand side. Columns (2) and (5) correspond with Equation 1 while Columns (3) and (6) are a saturated model with caseworker by manager fixed effects. Given that the dependent variable has already been residualized, there are no time or case composition controls.

Table 4: Analysis of Variance with Saturated Model: Miss Rate and Reapplication Rate

	Miss Rate			Reapplication Rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Observations	158,613	158,605	156,926	158,613	158,605	156,926
R-Squared	0.254	0.272	0.318	0.231	0.249	0.292
Adjusted R-Squared	0.224	0.239	0.26	0.201	0.214	0.233
Case Worker FE	X	X		X	X	
Manager FE		X			X	
Case Worker x Manager FE			X			X
Office FE						

Notes: Each column is a regression with caseworker residualized miss rate or reapplication rate on the left hand side and different fixed effects on the right hand side. Columns (2) and (5) correspond with Equation 1 while Columns (3) and (6) are a saturated model with caseworker by manager fixed effects. Given that the dependent variable has already been residualized, there are no time or case composition controls.

Table 5: Difference in Differences Impact of Manager Fixed Effect Changes on Caseworker Outcomes

	(1) Ln(Throughput)	(2) Permissiveness	(3) Miss Rate
1 s.d. Change in Manager Ln(Throughput)	0.088***		
	(0.022)		
1 s.d. Change in Manager Permissiveness		0.105***	
		(0.023)	
1 s.d. Change in Manager Miss Rate			0.113***
			(0.025)
Dependent variable mean	0	0	0
Dependent variable standard deviation	1	1	1
R-squared	0.343	0.312	0.201
Observations	70,627	70,655	70,655
Case Workers	1638	1638	1638
Distinct Events	917	917	917

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Plots the difference in differences version of the event study in Equation 2 that compares estimates in event time $k \in \{2, 3, 4\}$ to $k \in \{-4, -3, -2\}$, otherwise leaving everything else the same. Outcomes in each column are standardized.

Table 6: Difference in Differences Impact of Manager Fixed Effect Changes on Caseworker Outcomes: Not Standardized

	(1) Ln(Throughput)	(2) Permissiveness	(3) Miss Rate
1 s.d. Change in Manager Ln(Throughput)	0.038***		
	(0.010)		
1 s.d. Change in Manager Permissiveness		0.007***	
		(0.002)	
1 s.d. Change in Manager Miss Rate			0.004***
			(0.001)
Dependent variable mean	.02	0	0
Dependent variable standard deviation	.43	.07	.03
R-squared	0.343	0.312	0.201
Observations	70,627	70,655	70,655
Case Workers	1638	1638	1638
Distinct Events	917	917	917

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Plots the difference in differences version of the event study in Equation 2 that compares estimates in event time $k \in \{2, 3, 4\}$ to $k \in \{-4, -3, -2\}$, otherwise leaving everything else the same. Outcomes in each column are not standardized, and are interpreted as percentage change in log throughput and percentage point changes in permissiveness and the miss rate.

Table 7: Difference in Differences Impact of Manager Fixed Effect Changes on Caseworker Outcomes Split for Multi-Worker vs. Single-Worker Events: Ln(Throughput)

	(1) All	(2) Multi-Worker Events	(3) Single Worker Events
1 s.d. Change in Manager Ln(Throughput)	0.088***	0.095**	0.079**
	(0.022)	(0.033)	(0.029)
Dependent variable mean	0	0	0
Dependent variable standard deviation	1	1	1
R-squared	0.343	0.355	0.321
Observations	70,627	46,118	24,509
Case Workers	1638	1038	600
Distinct Events	917	317	600

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Plots the difference in differences version of the event study in Equation 2 for log throughput that compares estimates in event time $k \in \{2, 3, 4\}$ to $k \in \{-4, -3, -2\}$, otherwise leaving everything else the same. The sample of events in Column (1) include all caseworker events. Column (2) includes caseworker events where more than one caseworker shifts from a given old to new manager at the same time, indicating an event where the manager is likely to have moved. Column (3) include events where only one caseworker is impacted, which is more likely to be a caseworker move. This could be either a caseworker moves or that there wasn't clean pseudo-balanced data for the other workers impacted by the manager move. Outcomes in each column are standardized.

Table 8: Difference in Differences Impact of Manager Fixed Effect Changes on Caseworker Outcomes Split for Multi-Worker vs. Single-Worker Events: Permissiveness

	(1) All	(2) Multi-Worker Events	(3) Single Worker Events
1 s.d. Change in Manager Permissiveness	0.105^{***}	0.151^{***}	0.058
	(0.023)	(0.033)	(0.032)
Dependent variable mean	0	0	0
Dependent variable standard deviation	1	1	1
R-squared	0.312	0.309	0.319
Observations	70,655	46,133	24,522
Case Workers	1638	1038	600
Distinct Events	917	317	600

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Plots the difference in differences version of the event study in Equation 2 for permissiveness that compares estimates in event time $k \in \{2, 3, 4\}$ to $k \in \{-4, -3, -2\}$, otherwise leaving everything else the same. The sample of events in Column (1) include all caseworker events. Column (2) includes caseworker events where more than one caseworker shifts from a given old to new manager at the same time, indicating an event where the manager is likely to have moved. Column (3) include events where only one caseworker is impacted, which is more likely to be a caseworker move. This could be either a caseworker moves or that there wasn't clean pseudo-balanced data for the other workers impacted by the manager move. Outcomes in each column are standardized.

Table 9: Difference in Differences Impact of Manager Fixed Effect Changes on Caseworker Outcomes Split for Multi-Worker vs. Single-Worker Events: Miss Rate

	(1) All	(2) Multi-Worker Events	(3) Single Worker Events
1 s.d. Change in Manager Miss Rate	0.113***	0.158***	0.047
	(0.025)	(0.036)	(0.032)
Dependent variable mean	0	0	0
Dependent variable standard deviation	1	1	1
R-squared	0.201	0.206	0.192
Observations	70,655	46,133	24,522
Case Workers	1638	1038	600
Distinct Events	917	317	600

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Plots the difference in differences version of the event study in Equation 2 for the miss rate that compares estimates in event time $k \in \{2, 3, 4\}$ to $k \in \{-4, -3, -2\}$, otherwise leaving everything else the same. The sample of events in Column (1) include all caseworker events. Column (2) includes caseworker events where more than one caseworker shifts from a given old to new manager at the same time, indicating an event where the manager is likely to have moved. Column (3) include events where only one caseworker is impacted, which is more likely to be a caseworker move. This could be either a caseworker moves or that there wasn't clean pseudo-balanced data for the other workers impacted by the manager move. Outcomes in each column are standardized.

Table 10: Difference in Differences Impact of Manager Fixed Effect Changes on Caseworker Outcomes: Permissiveness and Program Participation

	(1) Permissiveness	(2) 3M Participation	(3) 6M Participation
1 s.d. Change in Manager Permissiveness	0.007***	0.005***	0.003**
	(0.002)	(0.001)	(0.001)
Dependent variable mean	0	0	0
Dependent variable standard deviation	.07	.05	.05
R-squared	0.312	0.216	0.14
Observations	70,655	70,655	68,063
Case Workers	1638	1638	1638
Distinct Events	917	917	917

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Plots the difference in differences version of the event study in Equation 2 for the miss rate that compares estimates in event time $k \in \{2, 3, 4\}$ to $k \in \{-4, -3, -2\}$, otherwise leaving everything else the same. Column (1) uses residualized permissiveness - the share of applications that were approved by the case worker - as the outcome of interest. Columns (2) and (3) use the share of cases that are participating in the program 3 months and 6 months after the caseworker decision to permit or deny the case, again residualized to adjust for differences in case composition and across time. All outcomes are not standardized and should be interpreted as percentage point changes.

Table 11: Variance Decomposition Method Comparison: Log Throughput

	Fixed Effects	(SE)	Shrinkage	Cov. Shrinkage
SD of Log Output	0.446		0.446	0.446
SD of Case Worker Effects	0.326	(0.003)	0.268	0.282
SD of Manager Effects	0.164	(0.004)	0.125	0.128
CaseWorker-Manager Effect Correlation	–0.280	(0.017)	–0.303	–0.225
SD of CaseWorker + Manager	0.321	(0.002)	0.259	0.282
Share Case Worker	0.537		0.363	0.402
Share Manager	0.135		0.079	0.082
Share Covariance	–0.075		–0.052	–0.041
Share Case Worker + Manager	0.519		0.340	0.402
Manager Case Worker Ratio	0.25		0.216	0.206
Adjusted R-Squared	0.372		0.372	0.372
Number of Observations	152115		152115	152115
Number of Case Workers	5817		5817	5817
Number of Managers	876		876	876
Number of Case Worker-Manager Pairs	12647		12647	12647
Number of Connected Sets	1		1	1

Notes: Shows the variance decomposition for log throughput using three different methods. Column (1) shows the standard variance decomposition with standard erorrs for the bootstrap in column (2). Column (2) shows the results from the standard shrinkage procedure that addresses only sampling error. Column (3) shows the results from the "covariance shrinkage" approach from Best, Hjort, and Szakonyi 2023 that also takes into account correlation in the sampling error across employees predicted by limited mobility bias. "SD of Outcome" is the standard deviation of the outcome variable. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker".

Table 12: Variance Decomposition Method Comparison: Permissiveness

	Fixed Effects	(SE)	Shrinkage	Cov. Shrinkage
SD of Approval Rate	0.070		0.070	0.070
SD of Case Worker Effects	0.048	(00)	0.041	0.043
SD of Manager Effects	0.027	(0.001)	0.021	0.023
CaseWorker-Manager Effect Correlation	–0.303	(0.017)	–0.338	–0.263
SD of CaseWorker + Manager	0.048	(00)	0.039	0.043
Share Case Worker	0.500		0.351	0.381
Share Manager	0.146		0.097	0.104
Share Covariance	–0.082		–0.063	–0.052
Share Case Worker + Manager	0.483		0.324	0.379
Manager Case Worker Ratio	0.291		0.279	0.273
Adjusted R-Squared	0.347		0.347	0.347
Number of Observations	152199		152199	152199
Number of Case Workers	5817		5817	5817
Number of Managers	876		876	876
Number of Case Worker-Manager Pairs	12652		12652	12652
Number of Connected Sets	1		1	1

Notes: Shows the variance decomposition for the approval rate using three different methods. Column (1) shows the standard variance decomposition with standard erorrs for the bootstrap in column (2). Column (2) shows the results from the standard shrinkage procedure that addresses only sampling error. Column (3) shows the results from the "covariance shrinkage" approach from Best, Hjort, and Szakonyi 2023 that also takes into account correlation in the sampling error across employees predicted by limited mobility bias. "SD of Outcome" is the standard deviation of the outcome variable. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker".

Table 13: Variance Decomposition Method Comparison: Miss Rate

	Fixed Effects	(SE)	Shrinkage	Cov. Shrinkage
	0.032		0.032	0.032
SD of Case Worker Effects	0.018	(00)	0.014	0.016
SD of Manager Effects	0.012	(00)	0.008	0.009
CaseWorker-Manager Effect Correlation	–0.340	(0.023)	–0.375	–0.300
SD of CaseWorker + Manager	0.018	(00)	0.014	0.016
Share Case Worker	0.370		0.209	0.250
Share Manager	0.134		0.079	0.094
Share Covariance	–0.075		–0.048	–0.046
Share Case Worker + Manager	0.352		0.192	0.252
Manager Case Worker Ratio	0.361		0.381	0.377
Adjusted R-Squared	0.214		0.214	0.214
Number of Observations	152229		152229	152229
Number of Case Workers	5817		5817	5817
Number of Managers	876		876	876
Number of Case Worker-Manager Pairs	12652		12652	12652
Number of Connected Sets	1		1	1

Notes: Shows the variance decomposition for the approval rate using three different methods. Column (1) shows the standard variance decomposition with standard erorrs for the bootstrap in column (2). Column (2) shows the results from the standard shrinkage procedure that addresses only sampling error. Column (3) shows the results from the "covariance shrinkage" approach from Best, Hjort, and Szakonyi 2023 that also takes into account correlation in the sampling error across employees predicted by limited mobility bias. "SD of Outcome" is the standard deviation of the outcome variable. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker".

Table 14: Variance Decomposition Using Only Moving Caseworkers

	Ln(Throughput)	Permit Rate	Miss Rate
SD of Outcome	0.430	0.068	0.030
SD of Case Worker Effects	0.259	0.039	0.014
SD of Manager Effects	0.123	0.021	0.008
CaseWorker-Manager Effect Correlation	–0.202	–0.189	–0.188
SD of CaseWorker + Manager	0.263	0.041	0.014
Share Case Worker	0.360	0.316	0.199
Share Manager	0.082	0.100	0.085
Share Covariance	–0.035	–0.034	–0.025
Share Case Worker + Manager	0.372	0.349	0.234
Manager Case Worker Ratio	0.229	0.317	0.428
Adjusted R-Squared	0.349	0.324	0.201
Number of Observations	93577	93625	93625
Number of Case Workers	2416	2416	2416
Number of Managers	844	844	844
Number of Case Worker-Manager Pairs	8030	8031	8031
Number of Connected Sets	1	1	1

Notes: This shows the variance decomposition of the estimated caseworker and manager effects from Equation 1 estimated using only caseworkers that have more than one manager during the study period and adjusted using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). Each column looks at a different outcome and a different set of estimated caseworker and manager fixed effects specifically for that outcome. "SD of Outcome" is the standard deviation of the outcome variable in each column. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker". Statistics for the estimation of Equation 1 are included at the bottom.

Table 15: Variance Decomposition Additional Robustness Specifications: Ln(Throughput)

	Main	Controls	Missing Track
SD of Outcome	0.445	0.594	0.250
SD of Case Worker Effects	0.282	0.343	0.158
SD of Manager Effects	0.128	0.244	0.073
CaseWorker-Manager Effect Correlation	–0.227	–0.173	–0.275
SD of CaseWorker + Manager	0.282	0.385	0.155
Share Case Worker	0.401	0.334	0.401
Share Manager	0.082	0.168	0.086
Share Covariance	–0.041	–0.041	–0.051
Share Case Worker + Manager	0.400	0.420	0.385
Manager Case Worker Ratio	0.206	0.504	0.216
Adjusted R-Squared	0.372	0.622	0.350
Number of Observations	152145	152145	152145
Number of Case Workers	5817	5817	5817
Number of Managers	876	876	876
Number of Case Worker-Manager Pairs	12647	12647	12647
Number of Connected Sets	1	1	1

Notes: This shows the variance decomposition of the estimated caseworker and manager effects from Equation 1 for log throughput for two different robustness specifications and adjusted using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). Column (1) is the standard variance decomposition result. Column (2) does not residualize the dependent variable and instead includes less granular case composition controls and month fixed effects in the regression. Column (3) adds an indicator that proxies for a case coming from the missing track (queue) to the residualization process to better control for differences in assignment to that queue. "SD of Outcome" is the standard deviation of the outcome variable in each column. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker". Statistics for the estimation of Equation 1 are included at the bottom.

Table 16: Variance Decomposition Additional Robustness Specifications: Permissiveness

	Main	Controls	Missing Track
SD of Outcome	0.069	0.137	0.060
SD of Case Worker Effects	0.042	0.063	0.036
SD of Manager Effects	0.022	0.043	0.019
CaseWorker-Manager Effect Correlation	–0.264	–0.169	–0.277
SD of CaseWorker + Manager	0.042	0.070	0.036
Share Case Worker	0.380	0.213	0.366
Share Manager	0.104	0.098	0.100
Share Covariance	–0.052	–0.024	–0.053
Share Case Worker + Manager	0.379	0.262	0.360
Manager Case Worker Ratio	0.273	0.460	0.275
Adjusted R-Squared	0.347	0.656	0.330
Number of Observations	152229	152229	152229
Number of Case Workers	5817	5817	5817
Number of Managers	876	876	876
Number of Case Worker-Manager Pairs	12652	12652	12652
Number of Connected Sets	1	1	1

Notes: This shows the variance decomposition of the estimated caseworker and manager effects from Equation 1 for permissiveness for two different robustness specifications and adjusted using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). Column (1) is the standard variance decomposition result. Column (2) does not residualize the dependent variable and instead includes less granular case composition controls and month fixed effects in the regression. Column (3) adds an indicator that proxies for a case coming from the missing track (queue) to the residualization process to better control for differences in assignment to that queue. "SD of Outcome" is the standard deviation of the outcome variable in each column. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker". Statistics for the estimation of Equation 1 are included at the bottom.

Table 17: Variance Decomposition Additional Robustness Specifications: Miss Rate

	Main	Controls	Missing Track
SD of Outcome	0.031	0.043	0.029
SD of Case Worker Effects	0.015	0.018	0.013
SD of Manager Effects	0.009	0.015	0.008
CaseWorker-Manager Effect Correlation	–0.300	–0.241	–0.298
SD of CaseWorker + Manager	0.015	0.021	0.013
Share Case Worker	0.250	0.186	0.223
Share Manager	0.094	0.127	0.082
Share Covariance	–0.046	–0.037	–0.040
Share Case Worker + Manager	0.252	0.239	0.225
Manager Case Worker Ratio	0.377	0.684	0.370
Adjusted R-Squared	0.214	0.462	0.186
Number of Observations	152229	152229	152229
Number of Case Workers	5817	5817	5817
Number of Managers	876	876	876
Number of Case Worker-Manager Pairs	12652	12652	12652
Number of Connected Sets	1	1	1

Notes: This shows the variance decomposition of the estimated caseworker and manager effects from Equation 1 for the miss rate for two different robustness specifications and adjusted using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). Column (1) is the standard variance decomposition result. Column (2) does not residualize the dependent variable and instead includes less granular case composition controls and month fixed effects in the regression. Column (3) adds an indicator that proxies for a case coming from the missing track (queue) to the residualization process to better control for differences in assignment to that queue. "SD of Outcome" is the standard deviation of the outcome variable in each column. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker". Statistics for the estimation of Equation 1 are included at the bottom.

Table 18: Variance Decomposition Split By Case Type: Permissiveness

	All	SNAP IN	SNAP RC	MA IN	MA RC
SD of Outcome	0.069	0.162	0.199	0.117	0.136
SD of Case Worker Effects	0.043	0.077	0.090	0.049	0.049
SD of Manager Effects	0.022	0.051	0.061	0.033	0.039
CaseWorker-Manager Effect Correlation	–0.264	–0.177	–0.394	–0.340	–0.322
SD of CaseWorker + Manager	0.043	0.085	0.086	0.049	0.052
Share Case Worker	0.381	0.228	0.202	0.174	0.129
Share Manager	0.104	0.100	0.095	0.081	0.086
Share Covariance	–0.052	–0.026	–0.055	–0.040	–0.034
Share Case Worker + Manager	0.380	0.275	0.188	0.174	0.147
Manager Case Worker Ratio	0.272	0.441	0.472	0.468	0.664
Adjusted R-Squared	0.347	0.199	0.115	0.118	0.073
Number of Observations	152229	141849	131547	143223	139889
Number of Case Workers	5817	5755	5721	5738	5705
Number of Managers	876	874	870	875	870
Number of Case Worker-Manager Pairs	12652	12419	12211	12387	12265
Number of Connected Sets	1	1	1	1	1

Notes: This shows the variance decomposition of the estimated caseworker and manager effects from Equation 1 for permissiveness split for different types of cases and adjusted using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). Column (1) is the standard variance decomposition result. Columns (2) to (5) focus on outcomes specifically for SNAP initial applications, SNAP recertifications, Medicaid initial application, and Medicaid recertifications, which are the largest groups of cases. "SD of Outcome" is the standard deviation of the outcome variable in each column. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker". Statistics for the estimation of Equation 1 are included at the bottom.

Table 19: Variance Decomposition Split By Case Type: Miss Rate

	All	SNAP IN	SNAP RC	MA IN	MA RC
SD of Outcome	0.031	0.075	0.142	0.049	0.060
SD of Case Worker Effects	0.015	0.023	0.046	0.013	0.017
SD of Manager Effects	0.009	0.015	0.032	0.008	0.013
CaseWorker-Manager Effect Correlation	–0.300	–0.364	–0.380	–0.434	–0.379
SD of CaseWorker + Manager	0.015	0.023	0.045	0.012	0.017
Share Case Worker	0.250	0.098	0.107	0.072	0.086
Share Manager	0.094	0.042	0.051	0.031	0.049
Share Covariance	–0.046	–0.023	–0.028	–0.020	–0.024
Share Case Worker + Manager	0.252	0.094	0.102	0.062	0.086
Manager Case Worker Ratio	0.377	0.432	0.480	0.433	0.571
Adjusted R-Squared	0.214	0.054	0.052	0.024	0.058
Number of Observations	152229	141849	131547	143223	139890
Number of Case Workers	5817	5755	5721	5738	5705
Number of Managers	876	874	870	875	870
Number of Case Worker-Manager Pairs	12652	12419	12211	12387	12265
Number of Connected Sets	1	1	1	1	1

Notes: This shows the variance decomposition of the estimated caseworker and manager effects from Equation 1 for the miss rate split for different types of cases and adjusted using the "covariance shrinkage" approach from Best, Hjort, and Szakonyi (2023). Column (1) is the standard variance decomposition result. Columns (2) to (5) focus on outcomes specifically for SNAP initial applications, SNAP recertifications, Medicaid initial application, and Medicaid recertifications, which are the largest groups of cases. "SD of Outcome" is the standard deviation of the outcome variable in each column. The second panel reports the standard deviation of the caseworker, manager, and caseworker + manager effects. It also reports the correlation between manager and caseworker effects. The third panel reports the share of the variation in the outcome variable explained by each of these components. These are calculated by taking the estimated variance (or covariance) of each component and dividing by the variance of the outcome variable. The final row "Manager Caseworker Ratio" expresses how much of the outcome variation managers explain relative to caseworkers, and is obtained by dividing "Share Manager" by "Share Caseworker". Statistics for the estimation of Equation 1 are included at the bottom.

Table 20: Difference in Manager Impact On Accuracy Between the 90th and 10th Percentile Manager: Full Set of Miss Rate Scaling and Relative Cost of Error β Assumptions

	Scale 0.5x	Scale 1.0x	Scale 2.0x
Beta=0.0	.012	.023	.047
Beta=0.2	.008	.016	.039
Beta=0.4	.013	.017	.032
Beta=0.5	.018	.019	.03
Beta=0.6	.022	.022	.03
Beta=0.8	.032	.027	.033
Beta=1.0	.041	.036	.037

Notes: Summarizes variation between the 90th and 10th percentile of manager impacts on accuracy weighted by caseworker-month. The Manager impact on accuracy calculated using the estimated manager impacts on permissiveness and the miss rate. The scaling refers to assumptions about how the number of cases missed in the miss rate is scaled up or down by a factor of two to reflect possible mismeasurement of the magnitude of false negative errors. The different values of β reflect different weights for the relative cost of false positive and false negative errors when measuring accuracy. This is in comparison to a difference in estimated manager impacts between the 90th and 10th percentile for permissiveness and the miss rate (scale 1.0x) of 4.7 p.p. and 2.3 p.p., respectively.

Table 21: Impact of Shifting the Bottom Quartile of Managers for Each Outcome to the 75th Percentile

	Throughput (cases) (1)	Permit Rate (p.p.) (2)	Miss Rate (p.p.) (3)	Accuracy (p.p.) (4)
Per Manager-Year				
Impact (cases/p.p.)	3,428	3.94	1.89	1.49
Baseline (cases/p.p.)	11,603	56.03	5.62	
Baseline Throughput (cases)	11,603	12,773	14,276	13,843
Overall				
2018-2023 Impact (million cases/p.p.)	$1.55*10^6$	0.77	0.43	0.36
2018-2023 Baseline (million cases/p.p.)	$27.18*10^6$	56.46	6.47	
Managers in Bottom Quartile	205	205	205	205
Manager-Years in Bottom Quartile	452	434	456	492

Notes: For a given outcome, reports the impact on that outcome of shifting the bottom quartile of managers for that outcome to the 75th percentile of that outcome. Omits the first percentile to remove any outliers. This involves shifting 205 managers, which have 400-500 manager-years worth of data between 2018 and 2023. Per manager-year impact is the amount throughput per manager-year has increased for the shifted managers, or for the other outcomes the change in the rate that cases are permitted, missed, or accurately decided has changed for the shifted managers. The baseline is the level of the outcome at baseline for the shifted managers, while baseline throughput is the average throughput per manager-year of the managers so that the number of additional cases permitted, missed, or accurately reviewed can be calculated. The overall impact for throughput reports the total increase in organization-wide output, while the other outcomes report the change in the organization-wide rate that cases are permitted, missed, or accurately decided. Overall impacts are for the entirety of the 2018 to 2023 period. Overall baseline is the organization-wide output (cases reviewed), permit rate, miss rate, and accuracy rate for the 2018 to 2023 period. Baseline numbers are not available for accuracy since differences in manager impacts on accuracy are inferred from differences in manager permissiveness and the miss rate, which does not make assumptions about the level of cases accurately reviewed.

Table 22: Impact of Shifting the Bottom Quartile of Caseworkers for Each Outcome to the 75th Percentile

	Throughput (cases) (1)	Permit Rate (p.p.) (2)	Miss Rate (p.p.) (3)	Accuracy (p.p.) (4)
Per Caseworker-Year				
Impact (cases/p.p.)	1,107	8.65	3.02	2.48
Baseline (cases/p.p.)	1,445	49.53	5.19	
Baseline Throughput (cases)	1,445	1,927	2,224	2,166
Overall				
2018-2023 Impact (million cases/p.p.)	$2.81*10^6$	1.65	0.73	0.56
2018-2023 Baseline (million cases/p.p.)	$27.18*10^6$	56.46	6.47	
Caseworkers in Bottom Quartile	1,392	1,392	1,392	1,392
Caseworker-Years in Bottom Quartile	2,534	2,799	3,073	2,967

Notes: For a given outcome, reports the impact on that outcome of shifting the bottom quartile of caseworkers for that outcome to the 75th percentile of that outcome. Omits the first percentile to remove any outliers. This involves shifting 205 caseworkers, which have 2,500-3,100 caseworker-years worth of data between 2018 and 2023. Per caseworker-year impact is the amount throughput per caseworker-year has increased for the shifted caseworkers, or for the other outcomes the change in the rate that cases are permitted, missed, or accurately decided has changed for the shifted caseworkers. The baseline is the level of the outcome at baseline for the shifted caseworkers, while baseline throughput is the average throughput per caseworker-year of the caseworkers so that the number of additional cases permitted, missed, or accurately reviewed can be calculated. The overall impact for caseworkers reports the total increase in organization-wide output, while the other outcomes report the change in the organization-wide rate that cases are permitted, missed, or accurately decided. Overall impacts are for the entirety of the 2018 to 2023 period. Overall baseline is the organization-wide output (cases reviewed), permit rate, miss rate, and accuracy rate for the 2018 to 2023 period. Baseline numbers are not available for accuracy since differences in caseworker impacts on accuracy are inferred from differences in caseworker permissiveness and the miss rate, which does not make assumptions about the level of cases accurately reviewed.

C What Explains Differences in Manager Decision-Making: Differences in Accuracy or Decision-Making Preferences?

In Section 6 of the main text, I explored whether differences in manager throughput and accuracy are driven by differences in productivity or preferences. More specifically, preferences in this context reflect "production preferences" between throughput and accuracy, i.e. quantity and quality. However, accuracy as defined in Section 6 can be achieved in different ways when optimizing over both false positive and false negative errors, reflecting differences in decision-making. In this section, I abstract from manager throughput and more formally explore what drives differences in manager decision-making: differences in manager accuracy or differences in manager decision-making preferences. I also determine the amount of variation in accuracy created under different assumptions.

In my setting, there are two outcomes from the decision making process: the rate of false positive and false negative errors. Differences in manager accuracy would imply that managers are making caseworkers better or worse decision-makers and would create variation in false positives and false negative errors that is positively correlated. On the other hand, differences in manager decision-making preferences would imply that managers are instead shifting caseworker decision-making rather than making them better or worse, which would create variation in false positive and false negative errors that is negatively correlated. For a given manager, permissiveness reflects how they are trading-off between the costs of these false-positive and false-negative errors.

To investigate this, I look at the empirical cross-sectional joint distribution of manager impacts on false negatives and false positives. However, I do not observe false positives in my setting. As discussed in Section 6, I follow Chan, Gentzkow, and Yu (2022) to infer manager impacts on false positives and accuracy using manager impacts on permissiveness and the miss rate. I measure relative differences in manager impacts and do not infer the level of false positive errors. Therefore, I do not plot the standard receiver operating characteristic ("ROC") curve, but rather compare differences between managers.

The two main conditions this transformation requires are that the miss rate is a relevant proxy for false negatives and that drift in the share of cases that should be permitted is uncorrelated with caseworker-manager switches. These are discussed in detail in Section 6. An additional requirement is that the magnitude of the miss rate is not mismeasured, which is important for inferring false positives. Therefore, I consider the relationship between false negatives and false positives using different scenarios where the magnitude of the miss rate is off by a factor of two in either direction to consider the reasonable range of scenarios. The miss rate could be understated if only a share of falsely denied cases reapply, while the miss rate could be overstated if some applicants are correctly denied, make improvements to their application, and are then successfully approved when they reapply.

Appendix Figure 11 plots the joint distribution of impacts on the miss rate and permissiveness for managers and caseworkers separately, normalized relative to the average worker. The joint distribution has a negative correlation with a slope of -0.29 for both managers and caseworkers. This negative correlation manifests for two reasons. First, as permissiveness increases, there are fewer cases that are denied that can become false negatives, which should reduce the miss rate.[95] Second, as manager permissiveness increases, the conditional miss rate decreases, which is shown in Appendix Figure 10. This could reflect that managers with higher permissiveness are more sure of the cases that they deny. Manager j in this plot will have different accuracy relative to the average manager at the origin if $\frac{FN_j}{P_j} \notin [-1, 0]$, regardless of the assumptions about the relative cost of errors β. Because the correlation between the miss rate and permissiveness is between -1 and 0, this implies that a large share of managers and caseworkers (about two-thirds) do not have strictly different accuracy relative to the average worker. This is consistent with having the same accuracy, but the actual difference in accuracy depends on what value is assumed for β.

In Appendix Figure 18, I plot the cross-sectional relationship between manager impacts on false negative and false positive errors, where false negatives are proxied by the miss rate and false positives are inferred under the conditions discussed in Section 6. Both outcomes are standardized. In Appendix Subfigure 18(a) I show the relationship assuming the magnitude of the miss rate is measured correctly, plotting the distribution and the best linear fit line representing the correlation between the two outcomes. Overall, there is a negative correlation between manager impacts on false positive and false negative errors of -0.32. This could be consistent with differences in preferences, but I cannot rule out differences in accuracy or comparative advantage also playing a role. In addition, there is still a lot of heterogeneity across managers, including differences in accuracy. This is reflected by many managers having strictly different accuracy than others based on having both lower false negative and false positive impacts (more accuracte), or both higher false negative and false positive impacts (less accurate). However, these result change under extreme assumptions about the miss rate being mismeasured. In Appendix Subfigure 18(b) I add the best linear fit lines to show robustness scenarios that reflect the range of reasonable scenarios for different assumptions about the magnitude of the miss rate. Scaling up the magnitude of the miss rate

95. Therefore, the overall relationship between permissiveness and the miss rate will be negative unless the conditional miss rate is positive enough to overcome the "mechanical" effect.

creates a more negative relationship (darkest color line), while scaling down the magnitude of the miss rate (lightest color line) can lead to a positive overall relationship between false positives and false negatives. Overall this suggests there is variation in manager accuracy and preferences, but the relative importance is sensitive to assumptions and hard to gauge.

Appendix Figure 19 plots the relationship between caseworker impacts on false positive and false negative errors. Overall, there is a stronger negative correlation for caseworkers of -0.58, and even under extreme assumptions about the miss rate being off by a factor of 0.5 the relationship remains negative. This suggests that differences in decision-making preferences seem to have an important role in explaining variation across caseworkers, though again other explanations cannot be ruled out. This also suggests that if anything decision-making preferences are more important for caseworkers than they are for managers.

These empirical distributions between manager impacts on false positives and false negative errors and the assumption for the relative cost of differences in errors are what determine the amount of variation in manager accuracy. For example, when variation in false positives and false negatives is negatively correlated and equal to the negative of the relative value of errors $-\beta$, the variation in accuracy goes across the width of the distribution (parallel), leading to less variation in accuracy. In contrast, when variation in false positives and false negatives is positively correlated and equal to the value of β, variation in accuracy goes along the length of the distribution (perpendicular), leading to more variation in accuracy. The overall negative relationship between false positive and false negatives decision-making errors in my setting mean will lead to less overall variation in accuracy, but will be impacted by the assumption for β.

In Appendix Table 20 I show the amount of variation in manager impacts on accuracy under different assumptions. I look at the difference in accuracy between a manager at the 90th percentile and 10th percentile of the accuracy impact distribution. This is in comparison to a difference in estimated manager impacts between the 90th and 10th percentile for permissiveness and the miss rate of 4.7 p.p. and 2.3 p.p., respectively. Under the standard miss rate scaling, variation across managers is between 1.6 p.p. and 3.6 p.p., which can be thought of as differences in the number of overall decision-making errors across managers. The variation in accuracy is lowest when the miss rate is scaled down and the miss rate is the only error that matters due to a low β. Variation in accuracy is always large for high β because of the variation in manager permissiveness used to infer false positives. This suggests that there is important variation in manager permissiveness under most assumptions, but it is often less than the variation in manager permissiveness and less than or comparable to the variation in manager miss rate.

D Who Are the Different Managers and What do They Do Differently?

In this section, I consider what other factors might explain variation across managers. I focus on who are the different managers and to a lesser extent what are they doing differently based on what I observe in my data. To look at what factors correlate with manager impacts, I estimate pairwise regressions of possible explanatory factors on the manager impacts of interest to get correlations. For caseworker i and month t,

$$\widehat{\theta}_{m(i,t)} = \gamma X_{it} + \widehat{\alpha}_i + \epsilon_{it} \tag{8}$$

where $\widehat{\theta}_{m(i,t)}$ is the estimated manager fixed effect, X_{it} is the explanatory factor of interest, and $\widehat{\alpha}_i$ is estimated caseworker fixed effect. For ease of interpretation, I standardize both $\widehat{\theta}_{m(i,t)}$ and X_{it}. I use the same covariance shrunk estimated fixed effects. I cluster caseworker-month observations by manager. This regression implicitly weights managers by the number of caseworker-month observations. I assume that $\beta = 0.5$ when summarizing the correlation between accuracy and other explanatory factors. Certain information for manager tenure, salary, and demographics are obtained through a linkage to the quarterly employee records data available for about 40% of managers and is not available for caseworkers.

D.1 Manager Experience and Demographics

First I look at how manager experience correlates with manager impacts. For the 40% of managers linked to the employee records files, I observe a manager's tenure at Texas HHS, tenure as a manager, and gross salary for a subset of managers. In addition, I can observe how many quarters the manager appears in my 6 year period as a proxy for tenure available for all managers in the data. Appendix Figure 21 shows that manager accuracy and decision-making are uncorrelated with tenure, suggesting that experience does not explain differences in decision-making across managers. It is important to remember that prior to promotion, managers have been caseworkers for many years. Hence, they are very familiar with how decision-making works. For throughput, there is a weak positive correlation between throughput and experience; a one s.d. increase in experience is correlated with up to a 0.1 s.d. increase in the estimated manager log throughput fixed effect. Overall, this suggests that what explains differences across managers is not explained by differences in experience.[96]

96. This could reflect improvements in performance with experience being counteracted by negative selection of managers at longer tenures that don't get promoted or leave for a higher paying job. Using the

Next, I turn to manager demographics. For the 40% of managers linked to the employee records files, I observe manager gender, ethnicity and race jointly categorized, and education. However, education is often missing. Focusing on gender and race, Appendix Figure 22 shows that Black managers are 0.2 standard deviations more permissive than white mangers, though they have similar throughput. Given the relative difference between Black and white manager miss rate are not that different, this may imply that white managers are more accurate, but this result is sensitive to assumptions about the relative cost of errors and the measured magnitude of the miss rate.

D.2 Benefit Generosity and Targeting

I investigate if there are differences in the composition of the applicants that more and less permissive managers grant benefits to. I have limited information about applicant composition for Medicaid cases where I do not observe realized medical expenditure and there is very limited participants with non-zero income in Texas. For SNAP cases, I observe income, household size, and the awarded monthly benefit amount, which is a function of income and household size. Existing literature often focuses on how well different policies target benefits based on the characteristics of the marginal applicants the policy screens in or out of the program.[97] However, the awarded SNAP benefit amount in my setting may also reflect intensive margin differences in generosity across managers. Appendix Figure 23 shows that there is a positive correlation between permissiveness and the average SNAP benefit amount awarded for both managers and caseworkers. For managers, this is characterized by a weak positive correlation with extremes at either end of the distribution. Managers at the 90th percentile awarded 3.4% higher SNAP benefits per case permitted than managers at the 10th percentile. This suggests either that more permissive managers on the extensive margin are also more generous on the intensive margin or that increases in manager permissiveness are well targeted.

D.3 Other Factors

I look at a series of other factors that may explain differences in outcomes across managers. To start, I find that team composition, managerial assistance, when the manager

limited time frame of my data, I do not see clear evidence of improvements in throughput or accuracy within manager over time.

97. Neoclassical theory predicts that marginal applicants for application screens will be relatively low income and receive low SNAP benefits, so increasing permissiveness will reduce average benefit amounts. However, existing evidence suggests that behavioral biases can lead marginal applicants to have substantial gains from program participation. In addition, in my setting it is unclear why more and less permissiveness managers differ in their decision-making, which may or may not relate to applicant need.

was active between 2018-2023, and case composition are generally not correlated with any manager impacts.[98] Managers with a large team size may have marginally lower accuracy driven by higher permissiveness, but team size is not correlated with manager throughput.

In Appendix Figure 26, I show that there is notable variation in manager impacts across Texas HHS' 11 administrative regions. In addition, I show that managers located in more urban, densely populated areas with lower democratic vote share in the 2020 election have lower throughput and lower accuracy, but similar permissiveness. This suggests that differences in regional performance are important factors. Notably wages for managers are relatively standardized across the state, but labor markets vary significantly. In addition, higher-level regional administration could drive differences.

Finally, I discuss denial reasons. A caseworker that denies an application notes the reason for denial. These reasons are not mutually exclusive and may be applied sequentially.[99] Failure to provide information means that a case was not complete enough to make a determination on its status, meaning it was missing relevant information. SNAP and TANF cases often have interviews, which if missed is a reason for denial. A case being denied for "not eligible" means that the composition of the household members made them ineligible for the benefit. A retracted case is retracted by the applicant, but this is very rare. Denials based on having too many resources (assets) or too much income are as expected.

The first observation from Appendix Figure 24 is that manager throughput is uncorrelated with all denial reasons. This is a nice secondary check confirming that differences in decision making (i.e. denial reasons) are largely uncorrelated with differences in throughput, suggesting that managers are not achieving higher throughput by changing how they review cases. This is especially true for "failure to provide information," which is anecdotally what a worker would be more likely to list as a denial reason if they were fully reviewing case materials or rushing. The second observation is that managers that are more permissive have similar rates of denial for almost all denial reasons with the exception of failure to provide information. This makes sense because denials based on household composition, income, and assets are largely not subjective decisions. Instead, it seems like the differences in manager permissiveness relate to whether or not the manager's team decides the application is complete or not such that a full determination can be made. This could suggest that more permissive managers are more lenient about the provision of information. There is anecdotal evidence for this, since applicants often need to submit dozens of documents with their case to prove their eligibility. Another explanation is that some manager teams

98. See Appendix Figures 25 to 27.

99. For example, failure to provide information often means that something from the application was missing that was needed in order to make a determination. Even if the case was ineligible for other reasons like having too much income, it may be that only failure to provide information was flagged.

are more helpful in getting the case to the point of completion (e.g., Cook and East 2023).[100] The third observation is that higher manager accuracy is positively correlated with more denials for not being eligible and retracting as opposed to failure to provide information. Denial for retraction is rare; the effect is large in standard deviation terms but not in general terms. It isn't clear why more accurate managers deny more cases for being not eligible. One explanation is that applicants denied for not being eligible may be less likely to consider reapplying, or would reapply to a different benefit not included in the miss rate.

100. This could reflect behavior during interviews or for example reaching out to contact an employer to confirm prior employment when documentation is incomplete.

E Calculating Differences in Program Costs Across Managers

In Section 7, I show that shifting the least permissive manager quartile to the 75th percentile of permissiveness would increase organization-wide permit rates by 0.77 p.p. (1.4%).[101] This translates into over 200,000 additional permitted cases (1,000 per manager) from 2018-2023. I estimate that shifting the permissiveness of these 200 managers would generate $400 million ($0.94 million per manager-year) in additional program costs for tax payers from 2018-2023. In this section I explain the details behind this calculation

First, I determine whether it is reasonable to assume that differences in permissiveness due to managers is similar across case types. Appendix Table 18 shows variance decomposition results done separately for the four largest groups of cases: SNAP initial applications, SNAP recertifications, Medicaid initial applications, and Medicaid recertifications. These account for about 70% of all cases. I find that differences across managers explain a similar 8-10% of the variation in permissiveness for each of the most prominent case types. Therefore, I assume increases in permitted cases are split according to overall case composition. Overall, 42% of cases are Medicaid (or MEPD) and 54% of cases are SNAP. I do not include TANF cases in the calculation given their very limited number of cases.

Second, I estimate the causal impact of manager permissiveness on program participation. The impact of managers on participation and permit rates could deviate for several reasons. It could be the case that caseworker decisions in the data are not actually implemented or are later modified. In addition, differences in permissiveness will be attenuated as denied applicants reapply and approved applicants reach recertification. Appendix Table 10 uses a difference in differences version of my event study to show that a one standard deviation change in manager permissiveness increases the permit rate by 0.7 p.p. (.105 s.d.), 3-month participation by 0.5 p.p. (0.088 s.d.), and 6-month participation by 0.3 p.p. (0.071 s.d.).[102] This is a participation impact estimated using all programs and case types. This shows that differences in manager permissiveness have persistent, long-run impacts on program participation and cost. Based on this, I assume the manager's impact on participation decays linearly for the first six months from 100% of the permit effect (0.7 p.p.) to 43% of the permit effect (0.3/0.7).[103] Then I assume this linear rate continues such after month 10

101. This omits the first percentile permissiveness managers to avoid potential outliers.

102. The difference in difference compares the effect in event time periods $k = \{2, 3, 4\}$ to event time periods $k = \{-2, 3, 4\}$.

103. In the month the case is permitted, participation is 100% of the manager causal impact on permissiveness. Then this decays about 10 p.p. per period so that 3 months and 6 months it aligns with the estimated manager impacts for participation.

there is no remaining participation impacts. This is a conservative assumption given that many households participate for longer than 10 months and many Medicaid participants are certified for a year. In addition, during the pandemic emergency households were not required to recertify for Medicaid, which would lead to even more persistent effects.

Third, I discuss the average monthly program costs for participants. For SNAP monthly program costs, I show that more permissive managers if anything have a higher average awarded SNAP benefit amount. This suggests that either managers that are more permissive on the extensive margin are more generous on the intensive margin or that greater manager permissiveness is well-targeted.[104] Therefore, I assume that shifting manager permissiveness would not change the average SNAP benefit amount awarded, and use the average SNAP benefit amount awarded in my data of $334 per month.[105] I do not observe any information on applicant Medicaid costs, so I assume marginal applicants incur average participant program costs of $541 per month.[106]

This leads to an increase in cost from greater SNAP participation of $180 million dollars and from greater Medicaid participation of $227 million dollars, leading to a total increase of $406 million dollars in program costs. This is the increased cost from shifting the least permissive quartile of managers to the 75th percentile of permissiveness. This implies on average that the difference in program costs from a manager being in the bottom quartile versus at the 75th percentile is $0.94 million per manager-year.

104. See Appendix Figure 23 as well as Appendix Section D for an in-depth discussion.

105. This is an average across application, recertifications, and incomplete reviews.

106. The average Medicaid spending per enrollee in 2021 in Texas was $6,500. KFF State Health Facts https://www.kff.org/medicaid/state-indicator/medicaid-spending-per-enrollee/.