NBER WORKING PAPER SERIES

# MECHANISM REFORM FOR TASK ALLOCATION

E. Jason Baron
Richard Lombardo
Joseph P. Ryan
Jeongsoo Suh
Quitze Valenzuela-Stookey

Working Paper 32369
http://www.nber.org/papers/w32369

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2024, Revised September 2025

We thank Peter Arcidiacono, Pat Bayer, Chris Campos, Sylvain Chassang, Allan Collard-Wexler, Michael Dinerstein, Laura Doval, Joseph Doyle, Federico Echenique, Natalia Emanuel, Maria Fitzpatrick, Ezra Goldstein, Aram Grigoryan, Peter Hull, Jacob Leshno, Chris Mills, Matt Pecenco, Michael Pollmann, Canice Prendergast, Katherine Rittenhouse, Alex Teytelboym, and seminar participants at the ASSA Annual Meeting, BYU, CREST, Duke, ESSET 2024, MIT Sloan, NBER Market Design Meeting (Fall 2024), Opportunity Insights, Princeton, PSE, SAET 2024, SEA 2024, SITE 2024, UC Berkeley, UCL, UIUC, University of Chicago Booth, USC, Yale University, and the Workshop on Economics of Education (Valle Nevado, Chile) for helpful comments and suggestions. A previous version of this paper was titled “Mechanism Reform: An Application to Child Welfare.” The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Mechanism Reform for Task Allocation
E. Jason Baron, Richard Lombardo, Joseph P. Ryan, Jeongsoo Suh, and Quitze Valenzuela-Stookey
NBER Working Paper No. 32369
April 2024, Revised September 2025
JEL No. D82, H75, J13, J45

## ABSTRACT

Reforming an existing system for allocating tasks among agents introduces additional political and institutional constraints relative to designing one in isolation. We develop a general mechanism-design framework for using data on agents' performance to improve outcomes while ensuring that no agents are made worse off relative to the status quo. As an illustration, we apply our results to the assignment of Child Protective Services investigators to maltreatment cases. Simulations show the mechanism reduces false positives (unnecessary foster care placements) by up to 14% while also lowering false negatives (missed maltreatment cases) and overall placements.

E. Jason Baron
Duke University
Economics Department
and NBER
jason.baron@duke.edu

Richard Lombardo
Harvard University
Department of Economics
richardlombardo@fas.harvard.edu

Joseph P. Ryan
University of Michigan
joryan@umich.edu

Jeongsoo Suh
Duke University
jeongsoo.suh@duke.edu

Quitze Valenzuela-Stookey
University of California, Berkeley
quitze@berkeley.edu

The need to divide heterogeneous tasks among agents arises in a wide range of high-stakes settings. Examples include the assignment of teachers to students, doctors to patients, and judges to cases. In such settings, advances in data availability and empirical methodology can enable better measurement of agents' individual performance on different tasks, information which may be valuable for improving the task-allocation system. However, reforms do not occur in a vacuum: the status-quo system may impose political and institutional constraints on the new design. For example, agents might resist the adoption of a new mechanism that they expect will make them worse off. This raises a question: how can a designer leverage the data generated under the current system while respecting the constraints that it imposes on the new proposal? We call this a problem of *mechanism reform*.

To fix ideas, consider a setting with two types of tasks, hard or easy, in equal proportion. Each task must be assigned to one of several agents. Under the current system, each agent is assigned 10 tasks at random, so that in expectation 5 are hard and 5 are easy. Suppose now that a designer figures out how to estimate each agent's output on each type of task. Denote the outputs of agent $j$ by $\pi_j^H$ and $\pi_j^E$. To what extent are there gains to be had, in terms of total output, by using these performance measures to assign tasks?

To answer this question, the designer must compare the current system to some plausible alternative which utilizes the output measures. A natural candidate is to re-allocate tasks while fixing the number assigned to each agent at 10. Under this *nominal* workload constraint, it is optimal to assign all hard tasks to the agents with higher "comparative advantage" $\pi_j^H - \pi_j^E$, and the easy tasks to the remaining agents.

This approach, while intuitive, may fail the test of plausibility. What matters to agents is their *real* workload, as determined by their own relative cost of handling hard versus easy tasks. If hard tasks are in general more costly to handle then the proposed mechanism makes the agents asked to specialize in such tasks worse off in real terms, and thus does not provide an apples-to-apples comparison with the existing system. Implementing such a mechanism in practice could lead to protests or resignations by such agents, and potentially necessitate broader changes to the organizational structure.[1] On the other hand, agents who handle only easy cases, and thus see a reduced workload in real terms, are under-utilized. As a result, the mechanism could either over- or under-estimate the potential gains.

Resolving the discrepancy between nominal and real workloads would be straightforward if

[1] In particular, changes in compensation schemes could be used to address workload concerns. However, such changes are infeasible in many contexts, such as public-sector settings where salaries are generally governed by civil service rules that impose rigid pay scales (e.g., see Biasi et al. (2025)). This study focuses on reforms in which the compensation scheme is fixed.

all agents had the same relative cost (known to the designer) of handling hard versus easy tasks. However, this is typically not the case. In most applications, agents' preferences are their private information, unobservable to the designer, and there is no basis for assuming that their preferences are identical. The question, then, is how to construct a plausible alternative mechanism (defined here as one which makes no agent worse off in real terms) that exploits the performance measures to maximize output. Such a mechanism can provide a basis for measuring potential gains and serve as a starting point for practical reforms.

We solve this mechanism-design problem. In the context of the simple example above, the output-maximizing mechanism can be succinctly described. Assume for the moment that there is a continuum of agents.[2] Each agent is first offered an assignment of 5 hard tasks and 5 easy tasks, i.e., their expected assignment under the status-quo system. Agent $j$ is then given the option to take on additional hard tasks, giving up $p_1^j$ easy tasks in exchange for each hard task, or give up hard tasks in exchange for $p_2^j$ additional easy tasks per hard task, where $p_2^j \geq p_1^j$ are pre-specified by the designer. Since every agent has the option to retain the assignment $(5, 5)$, they can be no worse off than they were in expectation under the status quo. The personalized rates $p_1^j$ and $p_2^j$ are used to steer agents towards the tasks they are relatively good at, while allowing the allocation to respond to their preferences. For suitably chosen rates, this mechanism is optimal among all Bayesian incentive compatible (BIC) mechanisms that respect the real subjective workload constraint, and we provide an efficient procedure for calculating the optimal rates (Theorems 1 through 3).

The general model we study goes beyond the binary-task model above by allowing for an arbitrary number of task types, alternative status-quo assignments, a finite number of agents, and dynamic arrival of tasks. Each of these extensions introduces its own set of complications.

To motivate these extensions, and to provide a concrete illustration of our results, we frame the analysis in terms of assigning Child Protective Services (CPS) workers to investigate reported cases of child maltreatment. This is an important application, and we provide empirical evidence supporting our approach in this context. Moreover, the CPS setting is representative of a wide range of task-allocation problems to which our results can be applied.

Contact with CPS is surprisingly common in the U.S.: 37% of children are the subject of a maltreatment investigation by age 18 and 5% spend time in foster care (Doyle et al., 2025; Kim et al., 2017). Moreover, foster care placement is one of the most far-reaching government interventions and previous work shows that it has large effects on children's later-in-life outcomes (see Bald et al. (2022) for a review). Thus, understanding whether there

[2] This assumption simplifies the market-clearing conditions, which facilitates the description of the mechanism. We also study the extension to an arbitrary finite number of agents.

are simple mechanisms to improve the efficacy of CPS investigations is of policy importance.

At a high level, the system operates as follows: Cases are initiated through calls to a state-level hotline. After initial screening, cases that require further investigation are allocated to a regional office based on the child's location. The case is then promptly assigned to one of several investigators through a rotational system: a new case gets assigned to the investigator at the top of the queue, and that investigator moves to the end of the queue. The investigator probes the allegations and determines whether the child should be placed in foster care. Under CPS guidelines, this decision should be based on the assessed probability that the child will experience subsequent maltreatment if left in the home (Baron et al., 2024).

Our starting point in tackling this application is a set of identification results which allow us to estimate each investigator's performance on cases with different observable characteristics. These results, which may be of independent interest, are discussed in detail in Section IV.A. For the current purposes of describing the mechanism-design problem, it suffices to think of performance as the "output" from assigning different cases to each investigator.[3] Naturally, we would like to use these performance measures to more efficiently allocate cases.

However, handling cases is costly—requiring time, energy, and imposing emotional and psychological burdens—and we provide empirical evidence below that these burdens vary across case types and investigators and translate into significant turnover. In CPS agencies, where turnover and shortages are already chronic challenges (Casey Family Programs, 2023), these dynamics are especially salient: high turnover not only imposes heavy recruitment and training costs, but also risks a vicious cycle in which departing staff increase the caseloads of those who remain, further fueling exits. Absent a detailed understanding of this unraveling dynamic, it is prudent to approach investigator welfare with caution when considering reforms.

In response to concerns about investigator welfare, we impose the *status-quo constraint* that no investigator be made worse off under the new mechanism relative to the current rotational system.[4] Relative to the stylized example discussed at the outset (with cases as tasks and investigators as agents) the CPS setting features a number of complications. As in the example, the rotational system produces what is effectively a random assignment of cases to investigators. However, here cases arrive over time and must be assigned as they come, without knowledge of which cases will arrive in the future. Moreover, cases differ along

[3] In a bit more detail: investigator performance is determined by their propensity to make accurate foster care placement decisions (e.g., placing a child if and only if they would experience subsequent maltreatment in their home).

[4] It is straightforward to extend the results to allow each investigator to be at most $x\%$ worse off, for any $x \in [0, 100]$.

more than two dimensions.

In the language of mechanism design, the problem we face is one of dynamic combinatorial allocation with a type-dependent participation constraint and no transfers, where an investigator's type represents their privately-observed preference over bundles of cases.[5] We focus on reforms in which the compensation scheme is fixed, and thus do not consider transfers as part of the mechanism design.[6] The problem features several well-known technical challenges. To gain tractability, we focus on a class of *binary-classification mechanisms.* These mechanisms partition cases into two groups, which we refer to as "high" and "low," and are random conditional on each group (our results apply for any given partition of the set of cases). Focusing on binary-classification mechanisms allows us to first develop insights in a simpler model with two types of cases, and then apply these to the settings with an arbitrary number of case types. Even with only two types of cases, however, the optimal mechanism has not been identified in the literature. Our primary technical contribution is a solution to this problem.

We first solve a version of the problem which is relaxed along two dimensions. The first relaxation is to a static setting, in which we have a fixed set of cases to allocate among the investigators. The second is that we require only that each case be assigned in expectation (where the expectation is taken over the profile of investigator types), rather than ex-post, i.e., conditional on every realized type profile. Equivalently, this model represents a market with a large number of agents, and we therefore refer to it as the Large-Market Static (LMS) problem. The primary technical results characterize the optimal mechanism in the LMS problem (Theorems 1 through 3). The solution takes the form of personalized two-part pricing, as described above.

We then take the optimal mechanism for the LMS problem and convert it into an approximately optimal mechanism for the original (small-market, dynamic) problem. This is accomplished in two steps. First, we reimpose the constraint that every case be assigned ex-post. This defines what we call the Small-Market Static (SMS) problem. We show how to modify our mechanism from the LMS problem to approximately solve the SMS problem (Proposition 2). This mechanism is obviously strategy-proof (Li, 2017). Moreover, it is robust to the form

[5]It is worth reiterating the distinction between preferences, which are unobserved by the designer and determine an agent's subjective workload; and performance, which is observable and defines the designer's objective function.

[6]As mentioned above, workload concerns could perhaps be addressed with additional reforms, such as changes in investigator compensation or termination. However, as with many public sector employees, compensation and employment schemes tend to be rigid, and termination is not practical when agencies suffer from major staff shortages. We view a full-scale overhaul of the CPS system as beyond the scope of the current study, and instead opt for a more minimalist market-design approach (Sönmez, 2023).

of investigators' preferences and to misspecification of the preference distribution. We also show that this mechanism incentivizes agents to improve their performance (Section II.C).

Finally, we convert this mechanism into one that works in the dynamic setting, which we refer to as the Small-Market Dynamic (SMD) problem (Proposition 4). This gives us a mechanism which can be applied in practice. It is approximately optimal and strategy-proof, where the approximation improves in the number of investigators and the time horizon. Furthermore, this mechanism can be implemented without ex-ante knowledge of the distribution of investigators' preferences. Thus, this simple and easily-implementable mechanism provides a realistic lower bound on the potential gains from exploiting our measures of investigator performance.

To conclude, we apply our mechanism-design results in the CPS context. Using an administrative dataset from Michigan covering all child maltreatment investigations between 2008 and 2016, we exploit our identification results to estimate investigator performance and simulate counterfactual outcomes under our mechanism. To complement these simulations, we also conducted a statewide survey of investigators, which reveals substantial heterogeneity in how investigators perceive the burden of cases with different characteristics.

This analysis shows that there are potential gains to be had by exploiting performance data: in our baseline specification, assigning investigators to cases according to our mechanism lowers investigators' false positives (children placed in foster care who would have been safe in their homes) by 14%, while also reducing false negatives (children left at home who are subsequently maltreated) and overall placements.[7] This result is robust to several alternative specification assumptions and measures of subsequent maltreatment, and the gains are even larger for the largest counties in our sample. Importantly, the mechanism involves reallocating only existing resources and ensures that no investigator is made worse off. In fact, it improves welfare relative to the status quo by at least 10% for 26% of investigators in our sample. In contrast, the "naive" mechanism that simply fixes the total number of cases handled by each investigator generates better aggregate performance, but reduces investigator welfare by at least 10% for 27% of investigators. Thus, failing to consider preferences leads to an incorrect estimate of the true potential gains, and implementing such a mechanism could significantly harm recruitment and increase turnover in an already strained system.

**Related literature:** This study's primary contribution is a general mechanism-design framework which can be applied in a wide range of task allocation problems where (i) agents' performance

[7] Because we have restricted attention to binary classification mechanisms, and as the mechanism is only approximately optimal with a finite number of agents and dynamic arrival of cases, these estimates provide a lower bound on the potential gains from reassignment. Moreover, we provide evidence showing that the additional benefits of considering more complicated mechanisms may be limited in our context.

can be estimated, (ii) the designer seeks to reform an existing assignment mechanism to optimize some aggregate performance measure, and (iii) the designer aims to ensure that no agent is made worse off compared to the status-quo mechanism. This constraint is especially relevant in public sector settings like CPS, where high turnover and staff shortages make additional turnover infeasible. It may also bind in environments with hold harmless clauses (Dinerstein and Smith, 2021) or union contracts that explicitly prevent agents from being made worse off compared to the status quo. More generally, solving the design problem under this constraint provides a useful benchmark with which to empirically assess the potential of performance-based assignments.

In developing the framework, we contribute to several related literatures within economics. First and foremost, this paper belongs to the theoretical mechanism design literature. Consider first the static versions of the problem (LMS/SMS). Broadly, these are problems of organizational economics in which tasks are allocated to agents whose cost for performing the task is unknown (Spence, 1973; Holmstrom and Milgrom, 1987; Grossman and Hart, 1983; Holmstrom, 1989; Baker et al., 2001). Unlike the bulk of this literature, agents' private information matters not because we hope to influence their effort levels or minimize total input cost, but because we need to guarantee that investigators are not made worse off.

In the mechanism-design problem, the status-quo constraint is equivalent to a type-dependent participation constraint, which makes this a problem of countervailing incentives (Lewis and Sappington, 1989; Maggi and Rodriguez-Clare, 1995; Jullien, 2000; Dworczak and Muir, 2023). Our solution uses an ironing approach similar to that in Dworczak and Muir (2023). However, the multi-item, multi-agent allocation problem that we study here has not been addressed in this literature and requires the development of new techniques. Our work also relates to the growing literature on mechanism design with generalized social objectives (e.g., Dworczak et al. (2021) and Akbarpour et al. (2024a)).

Within the literature on combinatorial allocation problems, in which indivisible objects must be assigned to agents with unknown preferences, the natural benchmark is the class of market-type mechanisms based on ideas from competitive equilibrium. Seminal contributions include Varian (1973) and Hylland and Zeckhauser (1979). This literature focuses on identifying mechanisms with various desirable properties, usually including a notion of Pareto efficiency or fairness with respect to agents' preferences. In contrast, the current study is concerned with maximizing a social objective based on performance, and the preferences of the agents (investigators) enter only as a constraint. This is an important conceptual distinction, and leads to an optimal mechanism involving non-linear, personalized exchange rates for cases.[8]

[8]Classical market-type mechanisms belong to what Budish (2012) calls the "good-properties" approach

The problem that we ultimately address (SMD) is inherently dynamic, in that cases must be assigned as they arrive. As with the static problem, we differ from the literature on dynamic combinatorial allocation (Combe et al., 2021; Nguyen et al., 2023) and matching (Karp et al., 1990; Mehta et al., 2007; Aggarwal et al., 2011; Baccara et al., 2020) in both the constraints that we face and the objective.

Finally, the empirical application relates to a structural and empirical literature on matching in foster care systems. Most of this work focuses on assigning children to foster or adoptive families (e.g., Baccara et al. (2014), MacDonald (2024), Robinson-Cortes (2019), and Slaugh et al. (2016)). Our work complements this literature by analyzing a different allocation margin—matching investigators to cases—and by taking a mechanism-design approach to propose a data-driven improvement to the existing assignment process.

# I Model

We first introduce the static version of the model in which there is a fixed set of cases (tasks) $\mathcal{I} = \{1, \ldots, I\}$, with typical element $i$, to be allocated among a set of investigators (agents) $\mathcal{J} = \{1, \ldots, J\}$ with typical element $j$.[9] In Section III.B we study the dynamic extension in which cases arrive over time and must be assigned as they come, without knowledge of which cases will arrive in the future, as in the CPS setting. As discussed above, while we frame the model in terms of this application, the formal analysis is more broadly applicable.

The input side of the problem concerns the preferences of investigators. Importantly, each case is unique, and investigators may differ in their preferences over cases. For each investigator, we assume that preferences over cases are represented by a function $p_j : \mathcal{I} \to \mathbb{R}$, where $p_j(i)$ is the cost to $j$ of handling case $i$. The total cost of assigning a set of cases $X \subset \mathcal{I}$ to $j$ is then $\sum_{i \in X} p_j(i)$, which we refer to as $j$'s *workload*. We adopt the normalization that investigators prefer a lower workload. Since we do not impose $p_j(i) \geq 0$, it is possible for investigators to prefer more of certain cases.

The output side of the problem concerns the objective of the designer. The primary directive of CPS is to make accurate foster care placement decisions, ensuring that children are

of identifying mechanisms with various desirable properties (other examples include Deferred Acceptance and Top Trading Cycles mechanisms), in contrast to the "mechanism-design approach", adopted here, of maximizing an objective subject to constraints. For instance, Combe et al. (2022) adopt a good-properties approach in a two-sided matching context to study the assignment of teachers to schools. As Budish (2012) notes, however, the distinction between good-properties and mechanism-design approaches is not sharp. For example, Abdulkadiroğlu and Grigoryan (2023) study how to translate a designer's general distributional preferences into "good properties."

[9] Since cases are assigned within CPS offices, the problem is separable across offices. The theoretical analysis describes the assignment mechanism for a given office.

removed from the home only if they would otherwise experience subsequent maltreatment. Thus, the expected social cost of assigning case $i$ to investigator $j$ is

$$c(i,j) := \mathrm{FN}_{ij} \cdot c_{\mathrm{FN}} + \mathrm{FP}_{ij} \cdot c_{\mathrm{FP}} + \mathrm{TN}_{ij} \cdot c_{\mathrm{TN}} + \mathrm{TP}_{ij} \cdot c_{\mathrm{TP}}$$

where $\mathrm{FN}_{ij}$ is the probability of a false negative; that is, investigator $j$ leaves child $i$ in the home and $i$ subsequently experiences maltreatment. The social cost associated with this outcome is $c_{\mathrm{FN}}$. The other three outcomes are defined analogously. Thus, $(\mathrm{FN}_{ij}, \mathrm{FP}_{ij}, \mathrm{TN}_{ij}, \mathrm{TP}_{ij})$ describes the joint distribution of $j$'s potential decision and the latent variable indicating the maltreatment potential of case $i$, and $(c_{\mathrm{FN}}, c_{\mathrm{FP}}, c_{\mathrm{TN}}, c_{\mathrm{TP}})$ is the designer's Bernoulli utility function over the potential outcomes of $i$. We take the standard utilitarian approach to aggregating across individual cases: let $Z \in \mathbb{R}^{I \times J}$ be an assignment, where $Z_{ij} = 1$ if case $i$ is assigned to investigator $j$, and denote the social cost of Z by

$$C(Z) := \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} Z_{ij} c(i,j). \tag{1}$$

The parameters $(c_{\mathrm{FN}}, c_{\mathrm{FP}}, c_{\mathrm{TN}}, c_{\mathrm{TP}})$ are taken as given for the purposes of designing the mechanism. These must ultimately be chosen by the designer (CPS), and we calibrate them in the empirical application below. The challenge lies in identifying $(\mathrm{FN}_{ij}, \mathrm{FP}_{ij}, \mathrm{TN}_{ij}, \mathrm{TP}_{ij})_{i \in \mathcal{I}, j \in \mathcal{J}}$. While we can observe reliable proxies for subsequent maltreatment among children who remain at home—enabling us to identify $\mathrm{FN}_{ij}$ and $\mathrm{TN}_{ij}$ under the random assignment of investigators—$\mathrm{FP}_{ij}$ and $\mathrm{TP}_{ij}$ are not non-parametrically identified. This is because it is impossible to observe what would have happened in the home for children placed in foster care. Nonetheless, we show in Section IV.A that the difference between any two investigators' false positive and true positive rates is non-parametrically identified. Since these differences are sufficient to identify the designer's objective, we can treat the function $c$ as observable when constructing the mechanism.

Although the application focuses on the CPS context, our theoretical results are directly applicable to other assignment problems in which the designer's objective takes the form in eq. (1) for some known function $c$. In fact, many of the results also extend to more general objectives.[10]

## I.A Binary-classification mechanisms

To ensure that no investigator is made worse off, the mechanism must respond to their preferences over cases. The challenge is that these preferences are investigators' private information and must be elicited by the designer. If $I > 2$, as is the case in our application and

[10]In particular, the "inner problem" that we solve in Section III.B is independent of the designer's objective, and so the qualitative features of the optimal mechanism are preserved.

many other settings, the problem that we want to solve belongs to a class of multi-dimensional screening problems which are known to be extremely difficult to solve both analytically (e.g., Pavlov (2011); Hart and Nisan (2019)) and computationally (formally $\#P$-hard Daskalakis et al. (2014)). A full solution to this problem is beyond the scope of the current study. Moreover, even if such a solution could be obtained, previous studies suggest that it would likely be unworkably complex (Daskalakis et al., 2015), whereas the ability to describe the mechanism in simple terms is desirable from a policy perspective.

In order to derive a practical policy, we therefore focus on a restricted class of mechanisms, which we refer to as *binary-classification mechanisms*. In such mechanisms, the designer first specifies a binary partition of the set of cases into two categories, which we refer to as *high-type*, or type-$H$, and *low-type*, or type-$L$. We then consider mechanisms which are measurable with respect to this partition; that is, for which the probability of assigning cases $i$ and $i'$ to investigator $j$ is the same if $i$ and $i'$ are of the same type. In the empirical application, we define this partition based on the predicted likelihood that a case would result in subsequent maltreatment. However, from a theoretical perspective this partition can be arbitrary.

Fixing the partition, let $n^h$ and $n^l$ be the per-investigator number of type-$H$ and type-$L$ cases, respectively, so the total number of cases of each type is $Jn^h$ and $Jn^l$. Define $c^k(j) := \hat{E}[c(i,j)|i \text{ is type } k]$ for $k \in \{H, L\}$, where $\hat{E}$ denotes the empirical expectation. We refer to $c^k(j)$ as $j$'s *performance* on type-$k$ cases, so that a lower $c^k(j)$ means better performance. Similarly, let $p_j^k := \hat{E}[p_j(i)|i \text{ is type } k]$ for $k \in \{H, L\}$. Then, in any binary-classification mechanism in which investigator $j$ is assigned $H^j$ high-type cases and $L^j$ low-type cases, the expected social cost is $\sum_{j \in \mathcal{J}} c^h(j)H^j + c^l(j)L^j$ and the expected workload of investigator $j$ is $p_j^h H^j + p_j^l L^j$. The status-quo rotational mechanism amounts, in the long run, to a random allocation of cases. In other words, in expectation each investigator receives $n^h$ high-type and $n^l$ low-type cases. The constraint that investigator $j$ be made no worse off under the new assignment $(H^j, L^j)$ is thus given by $p_j^h H^j + p_j^l L^j \leq p_j^h n^h + p_j^l n^l$.

Observe that no binary-classification mechanism will be able to distinguish between types $(p_j^h, p_j^l), (\hat{p}_j^h, \hat{p}_j^l)$ such that $\frac{p_j^h}{p_j^l} = \frac{\hat{p}_j^h}{\hat{p}_j^l}$, as these agents have identical preferences over assignments. Thus, it is without loss of generality to consider mechanisms which elicit only the relative cost of high-type cases, $p_j := \frac{p_j^h}{p_j^l}$.[11] Henceforth, we refer to this ratio as the investigator's type and write mechanisms simply as a function of the one-dimensional type. We maintain

[11]This also means that the designer does not benefit from assigning different caseloads to agents with the same preferences; for a formal proof of this claim, see Dworczak et al. (2021) Theorem 8, where the same observation on the reduction of a two-dimensional to a one-dimensional type appears, albeit in a different setting.

the following assumption on types:

**Assumption.** The type $p_j := \frac{p_j^h}{p_j^l}$ has a full-support distribution on a bounded interval $[\underline{p}_j, \bar{p}_j]$ with CDF $F_j$ and continuous density $f_j$. Types are independent across investigators.

In the baseline model, we assume that the distributions $(F_j)_{j \in \mathcal{J}}$ are known to the designer. In Section III, we show how our solution can be used to construct a mechanism which does not depend on knowledge of these distributions.

In summary, under the restriction to binary classification mechanisms the problem becomes isomorphic to one with just two types of cases. This makes it possible to solve for the optimal mechanism, which we then show has desirable properties and comparative statics. Moreover, we demonstrate empirically that this class of mechanisms is rich enough to generate meaningful welfare gains, and that the gains from considering more complex mechanisms are modest in our setting. As discussed above, going beyond binary-classification introduces well-known technical challenges. Nonetheless, we discuss some potential avenues in Appendix B.3.

*Remark* 1. The definition of binary classification mechanisms makes no use of the fact that the status-quo mechanism is a uniform random allocation of cases. Indeed, our approach can also be applied if the status quo involves non-uniform randomization and if it differs across investigators. This more general specification complicates the market-clearing conditions in the mechanism-design problem (e.g., ($H$-capacity) and ($L$-capacity) below) but does not substantively affect the results. However, for simplicity, and as it is relevant for the CPS application and many others with rotational or quasi-random allocations, we focus here on the uniform-randomization case.

## I.B Discussion of the modeling assumptions

**Linearity of preferences:** We maintain throughout the linear specification of investigators' preferences. We find this restriction palatable in this context. First, the status-quo constraint ensures that aggregate workloads do not increase, which is conceptually similar to allowing the social cost of increasing an investigator's workload to be highly convex. Second, in reality, the problem is dynamic: an investigator's caseload consists of cases assigned at different points in time. Indeed, if each case were resolved before the next one began, to assume linear costs would just be to assume that payoffs are separable across periods. While there are certainly valid critiques of time separability, it is a standard assumption on preferences in dynamic settings. Of course, cases for a given investigator may overlap, so linear costs are not precisely equivalent to time separability in our context.

Moreover, the mechanism that we ultimately propose is robust to this linearity assumption.

The mechanism can be implemented regardless of the form that investigators' preferences take. It still has the potential to improve upon the status quo, and at worst achieves the same social welfare. See Section III for details.

**Investigators' performance:** For the purposes of designing the mechanism, we treat investigators' performance, as captured by the function $c$, as fixed. A concern is that changes to the mechanism might affect performance. One channel for this effect is that, if investigators' workload increases, they may perform worse on each individual case. However, our status-quo constraint ensures that this does not occur. Another channel would be through the investigators' incentives for effort. This would be especially concerning if investigators could reduce their workload by degrading their performance. Fortunately, we show that in our proposed mechanism the scope for such manipulation is limited (Theorem 4).[12] Moreover, our mechanism can accommodate performance which evolves over time, as well as the arrival of new investigators, because we can continue to update the performance estimates even after the mechanism has been implemented (see Appendix B.2).

# II Large-Market Static problem

We first consider a relaxed problem that abstracts from the combinatorial dimension of the original SMD problem. We do this by allowing for fractional assignments of cases and by requiring only that each case be assigned to some investigator in expectation (taken over investigators' types) rather than ex-post (for every realized type profile). Formally, in the LMS program the designer chooses a $(H^j, L^j) : [\underline{p}_j, \bar{p}_j] \to \mathbb{R}^2_+$ to minimize

$$\sum_{j \in \mathcal{J}} \mathbb{E}_{p_j \sim F_j} \left[ c^h(j) H^j(p_j) + c^l(j) L^j(p_j) \right]$$

subject to

$$pH^j(p) + L^j(p) \leq pH^j(p') + L^j(p') \qquad \forall j \in \mathcal{J} \quad p, p' \in [\underline{p}_j, \bar{p}_j] \tag{IC}$$

$$pH^j(p) + L^j(p) \leq pn^h + n^l \qquad \forall j \in \mathcal{J} \quad p \in [\underline{p}_j, \bar{p}_j] \tag{SQ}$$

$$\sum_{j \in \mathcal{J}} \mathbb{E}_{p_j \sim F_j} \left[ H^j(p_j) \right] = Jn^h \tag{$H$-capacity}$$

$$\sum_{j \in \mathcal{J}} \mathbb{E}_{p_j \sim F_j} \left[ L^j(p_j) \right] = Jn^l \tag{$L$-capacity}$$

[12] This property of the mechanism is also intimately related to the perceived fairness of the mechanism, in that investigators who receive more favorable assignments are precisely those whose performance is higher. Formally, our mechanism possesses an *envy-freeness* property that can help justify to investigators why their caseloads are no longer identical (see section II.C).

The fact that the capacity constraints are required to hold only in expectation is what makes this a "large-market" relaxation; this formulation is equivalent to assuming that each investigator $j$ is actually a unit-mass population of agents with identical performance and preference types distributed according to $F_j$.[13] The IC constraint says that investigators are better off (receive a lower workload) if they report their type truthfully (the standard revelation principle of Myerson (1981) applies here). The SQ, or status-quo, constraint ensures that every investigator is weakly better off relative to the current system in which cases are divided equally across investigators.[14]

## II.A Solution preview

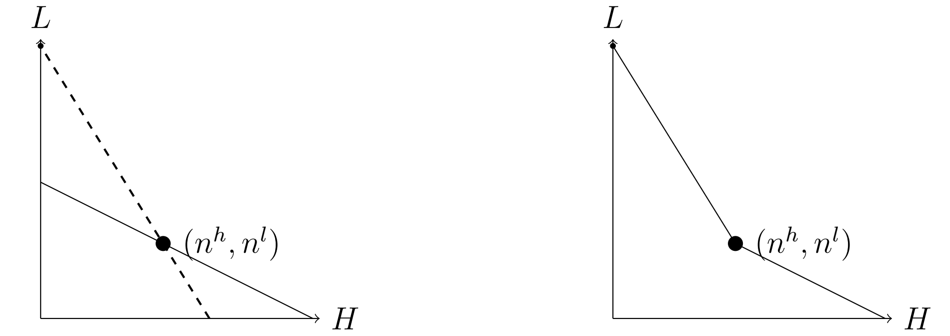
Before proceeding to the solution, it is useful to take a step back. The problem that we ultimately want to solve is one of combinatorial allocation. For such problems, the natural benchmark is the class of market-based mechanisms built on the idea of competitive equilibrium (CE) (Varian, 1973; Hylland and Zeckhauser, 1979; Budish, 2011; Nguyen and Vohra, 2021; Prendergast, 2022; Nguyen et al., 2023). To build intuition for our solution, we first evaluate a standard market-based solution applied to this context. This comparison is instructive for understanding the distinctive features of the current problem.

To apply a CE mechanism to the current setting we would grant each investigator an "endowment" equal to the expected status-quo assignment, denoted by $(n^h, n^l)$, and set a "price" $p$ for high-type cases in terms of low-type cases. We then allow investigator $j$ to choose their favorite bundle from the budget set $\{(\hat{n}_j^h, \hat{n}_j^l) : p\hat{n}_j^h + \hat{n}_j^l \geq pn^h + n^l\}$. The price $p$ should be set so that the market clears, i.e., all cases are assigned. Assuming such a market-clearing price exists and that agents behave as price takers, the allocation is efficient and fair; no investigator can be made better off without making some other investigator worse off, and no investigator would prefer another's assignment to their own (Varian, 1973). In finite markets, agents may be able to influence the price by distorting their demand, but this incentive disappears as the market grows (Roberts and Postlewaite, 1976).[15] By construction,

[13] Note that in this formulation the mechanism specifies each investigator's assignment as a function of their own type, rather than the entire type profile. In other words, $H^j$ and $L^j$ can be thought of as the interim allocation rules. Because of the large-market assumption, it is without loss of generality to work directly with the interim allocation rules; there are no additional constraints needed to ensure feasibility of these rules, à la Border (1991), as there would be in a "small market" in which all cases must be assigned ex post.

[14] An alternative would be to study the "profit maximization" problem: given a weight on investigator welfare relative to social welfare on the output side, maximize the sum of social welfare and investigator costs. This approach (or the dual of minimizing cost subject to a social welfare constraint) may be suitable in some task-allocation settings. However, here this would require the policymaker to take a stand on the relative weights of outcomes for children and burdens for investigators; a difficult, not to mention politically fraught, exercise. Our approach, in addition to the benefits already discussed, avoids such comparisons.

[15] Existence of market clearing price is not guaranteed with indivisible goods. However, the CE mechanism




(a) Two one-price budget sets

(b) One two-price budget set

Figure 1: Budget sets

the status quo is always affordable, so no investigator is worse off.

The downside of this market mechanism is that while it respects the preferences of investigators, it ignores those of the designer: the efficiency of the CE mechanism concerns the preferences of investigators, while the designer's objective concerns investigators' performance, which does not enter into the construction. This is the key distinction between our setting and those to which market-type mechanisms are typically applied (e.g., Budish et al. (2017); Prendergast (2017)).

To incorporate the designer's objective, the natural modification is to introduce personalized prices. Suppose that investigator $j$ performs well on high-type cases and poorly on low-type ones. Intuitively, we might try to steer $j$ towards the former by increasing $p^j$, the number of additional low-type cases $j$ must take on in exchange for one fewer high-type case, and allowing $j$ to choose from the budget set $\{(\hat{n}_j^h, \hat{n}_j^l) : p^j \hat{n}_j^h + \hat{n}_j^l \geq p^j n^h + n^l\}$.

Since $j$'s budget is determined by the value of the endowment $(n^h, n^l)$, increasing $p^j$ rotates the budget set around this point. Figure 1a depicts the budget line, where the high-type caseload is on the horizontal axis. An increase in $p^j$ corresponds to a rotation from the solid to the dotted budget set. The higher is $p^j$, the less attractive it is for $j$ to take on additional low-type cases. However, a larger $p^j$ also means that $j$ will perform fewer additional high-type cases for each low-type case they give up. Thus we face a trade-off between increasing the probability that $j$ specializes in high-type cases and ensuring that $j$ handles their fair share of the overall workload.

This trade-off arose because when we increase $p^j$ to make specializing in low-type cases less attractive to $j$, we simultaneously reduce the number of additional high-type cases that $j$

can be well approximated in a way that preserves its desirable efficiency and incentive properties (Budish, 2011; Azevedo and Budish, 2019).

can be asked to take on. This suggests that non-linear pricing could be useful. Suppose that in order to trade *away* a high-type case, $j$ is forced to take on an additional $p_2^j$ low-type cases, while if $j$ wants to trade *for* a high-type case they can give up $p_1^j$ low-type cases, where $p_1^j < p_2^j$. The induced budget set is depicted in Figure 1b. By increasing $p_2^j$ above $p_1^j$ we make it less attractive for $j$ to give up high-type cases, without affecting the rate at which they can give up low-type cases. Instead, the kink in the budget set at $(n^h, n^l)$ increases the likelihood that $j$ simply opts to retain the status quo.

Given the potential value of non-linear pricing, we can consider even more flexible schemes. In the extreme we could allow the exchange rate between high- and low-type cases to vary continuously in the space of case bundles. Surprisingly, additional flexibility is not needed. Type distributions are called (Myerson) regular if $\phi_j(p) := p - \frac{1-F_j(p)}{f_j(p)}$ is non-decreasing for all $j$, and strictly regular if $\phi_j$ is strictly increasing. We say that $F$ is *strongly regular* if $\phi_j$ is strictly increasing and, moreover, $\phi(\underline{p}) \geq 0$ and $F(\bar{p})/\bar{p} > F(p)/p$ for all $p \in [\underline{p}, \bar{p}]$.[16]

**Theorem 1.** If type distributions are strongly regular, then a personalized two-price mechanism, i.e., one in which each investigator receives a budget set as in Figure 1b, is optimal in the LMS problem, and the solution is unique if no two investigators have the same performance for either case type. Absent strong regularity, at most four prices are needed for each investigator (i.e., three kinks in the budget set).

This result is implied by Theorem 2 below. Theorem 1 describes only the qualitative features of the optimal mechanism. The final crucial step, which we describe below, is to then derive the optimal prices for each investigator.[17]

## II.B Solving the LMS program

To solve the LMS program, we make use of the fact that both the objective and the market-clearing conditions depend only on the expected caseloads for each investigator. Thus, we can solve the problem in two parts. First, in an "inner problem" we characterize for each investigator $j$ the expected caseloads, $(\mathbb{E}_j[H^j(p)], \mathbb{E}_j[L^j(p)])$, that $j$ can be assigned by some IC and SQ mechanism. We refer to this as the set of *incentive-feasible* pairs, denoted by $\mathcal{F}_j$. We then solve an "outer problem" in which for each $j$ we choose an incentive-feasible expected caseload $(\hat{n}_j^h, \hat{n}_j^l) \in \mathcal{F}_j$ to minimize the designer's objective, subject to the market-clearing conditions. In other words, the inner problem deals with IC and SQ, and the outer problem with market clearing.

[16] A sufficient condition for $F$ to be strongly regular is that it satisfy strict Myerson regularity, and $f(p)p \geq \max\{1 - F(p), F(p)\}$ for all $p \in [\underline{p}, \bar{p}]$.

[17] In practice it may be easier to implement the mechanism directly by eliciting type reports, rather than indirectly allowing agents to choose from a budget set. See Section V for further discussion of implementation.

It turns out to be convenient in the inner problem to characterize the sets $(\mathcal{F}_j)_{j\in\mathcal{J}}$ via their support functions. The dual to the outer problem then has a convenient formulation in terms of these support functions. This dual formulation is computationally useful and also facilitates analytical comparative statics.

**Step 1: LMS inner problem:** The inner problem concerns the design of a mechanism for a single investigator, and so we drop the dependence on $j$ in the notation here. It is easy to see that the set of incentive-feasible pairs, $\mathcal{F}$, is convex, since the IC and SQ constraints are linear in the mechanism $(H, L)$. This set can thus be described by its support function $S : \mathbb{R}^2 \to \mathbb{R}$ defined by $S(a, b) := \max\{a\hat{n}^h + b\hat{n}^l : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}\}$, which yields the dual characterization $\mathcal{F} = \{(\hat{n}^h, \hat{n}^l) : a\hat{n}^h + b\hat{n}^l \leq S(a, b)\ \forall\ (a, b) \in \mathbb{R}^2\}$. In order to calculate the support function, we need to maximize over precisely the set we wish to characterize. The way to do this is to maximize directly over the set of IC and SQ mechanisms. That is, for arbitrary $(a, b) \in \mathbb{R}^2_+$, we solve

$$S(a,b) = \max_{H,L\geq 0} \quad a\int H(p)dF(p) + b\int L(p)dF(p) \tag{2}$$

$$s.t \qquad -pH(p) - L(p) \geq -pH(p') - L(p') \quad \forall\ p, p' \in [\underline{p}, \bar{p}] \tag{IC}$$

$$-pH(p) - L(p) \geq -pn^h - n^l \qquad \forall\ p \in [\underline{p}, \bar{p}] \tag{SQ}$$

Note that while there are no transfers in our setting, we can think of $H$ as playing the role of the physical allocation and $L$ that of transfers. Thus, this program shares many similarities with the classic monopoly pricing problem of Myerson (1981). The two essential distinctions between the LMS-within program in eq. (2) and the monopoly-pricing problem are *(i)* a non-negativity constraint on $L$, and *(ii)* a type-dependent participation constraint determined by the need to respect the status quo.[18]

Lower bounds on transfers, $L$ in the current context, are studied in a similar problem by Loertscher and Muir (2021). However, their problem does not feature a type-dependent participation constraint. Such constraints are studied in the literature on countervailing incentives (e.g., Maggi and Rodriguez-Clare (1995), Jullien (2000), and Dworczak and Muir (2023)). However, a status-quo constraint of this form in a multi-item allocation problem without transfers has not, to our knowledge, been studied. Nonetheless, for solving the inner problem similar ironing techniques can be used. The main challenge lies in identifying for which types the SQ constraint should bind. As in Jullien (2000), it turns out that the status-quo constraint binds for an intermediate interval of types.

[18] The objective in the LMS-within program also differs from that in monopoly pricing, but this distinction is conceptually less important. For an example of recent work on mechanism design with general designer objectives see Akbarpour et al. (2024b).

**Theorem 2.** For any $(a, b) \in \mathbb{R}^2$ there is an optimal mechanism $(H^*, L^*)$ defining the support function $S(a, b)$ which takes the following form: there exist three thresholds $\underline{p} \leq p_1 \leq p_2 \leq p_3 \leq \bar{p}$ and a level $H_2 \geq n^h$ such that

$$H^*(p) = \begin{cases} n^h + \frac{n^l - (p_2 - p_1)H_2}{p_1} & \text{if } p \in [\underline{p}, p_1) \\ H_2 & \text{if } p \in [p_1, p_2] \\ n^h & \text{if } p \in (p_2, p_3] \\ 0 & \text{if } p \in (p_3, \bar{p}] \end{cases}$$

where $H_2$ must satisfy $n^h + \frac{n^l - (p_2 - p_1)H_2}{p_1} \geq H_2$. Under the optimal mechanism $L(p) = n^l$ for $p \in [p_2, p_3]$ and, as noted above, $L(p) = 0$ for $p \in [\underline{p}, p_1)$. Moreover, the IC constraint of type $p_3$ implies that $L(p) = p_3 n^h + n^l$ for $p \in (p_3, \bar{p}]$.

*Proof.* Proof in Appendix A.1. □

Theorem 2 says that the mechanism maximizing a weighted sum of expected $H$ and $L$ caseloads takes on no more than four distinct values; one intermediate set of types (between $p_2$ and $p_3$) who retain the status-quo assignment, one set above who get only low-type cases, and two sets below. The reason there are two assignment levels below the status quo, as opposed to only one above, is that in this region the non-negativity constraint on $L$ may bind, and ironing under this additional constraint can give rise to an additional assignment level. However, under the standard Myerson regularity condition on $F$ it is without loss to consider only two-part mechanisms when maximizing a non-decreasing objective.

**Corollary 1.** If the virtual value $\phi(p) = p - \frac{1-F(p)}{f(p)}$ is increasing, then for any $(a, b) \geq 0$ there is an optimal mechanism defined by thresholds $p_1 \leq p_2$ such that

$$(H^*(p), L^*(p)) = \begin{cases} (n^h + \frac{1}{p_1} n^l, 0) & \text{if } p \leq p_1 \\ (n^h, n^l) & \text{if } p \in (p_1, p_2) \\ (0, p_2 n^h + n^l) & \text{if } p \geq p_2. \end{cases}$$

If $\phi$ is strictly increasing, then the solution is unique (up to zero-measure perturbations).

*Proof.* Proof in Appendix A.1. □

As discussed above, such a mechanism has a simple indirect implementation, in which the investigator is presented with a kinked budget set as in Figure 1b and allowed to exchange cases. Without regularity, the budget set features two kinks: one at the endowment, and one at a point of more high-type and fewer low-type cases.

Theorem 2 greatly simplifies the problem of solving for the value $S(a,b)$. Moreover, it tells us what a mechanism that achieves the value $S(a,b)$ will look like, which allows us to characterize $N^*(a,b) := \arg\max\{a\hat{n}^h + b\hat{n}^l : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}\}$. Define the *frontier* to be the set $\{N^*(a,b) \ : \neg\{(a,b) \le 0\}\}$, i.e., the set of solutions when at least one of $a,b$ is strictly positive. The set $\mathcal{F}$ is just the subset of the positive orthant that lies within the frontier. Abusing terminology, we say that $\mathcal{F}$ is *strictly convex* if the mixture of any two points on the frontier lies in the interior of $\mathcal{F}$. Say that $\mathcal{F}$ is *downward closed* if $(x,y) \in \mathcal{F}$ and $0 \le (x',y') \le (x,y)$ implies $(x',y') \in \mathcal{F}$. So $\mathcal{F}$ is downward closed if and only if the frontier is downward sloping.

**Corollary 2.** If $F$ is strongly regular then $\mathcal{F}$ is strictly convex and downward closed.

*Proof.* Proof in Appendix A.2. □

*Remark* 2. Theorem 2 states that the value of $S(a,b)$ can be obtained with a mechanism involving at most 4 allocations. However, Theorem 1 allowed that the optimal mechanism could involve a mechanism involving 4 prices, and thus 5 allocations, for some investigators. The discrepancy arises because while the value $S(a,b)$ can be obtained with a 3-price mechanism, 4 prices might be needed to implement some points in $\mathcal{F}$. This occurs precisely when the frontier of $\mathcal{F}$ has linear segments. However, our primary focus is on the strongly regular case, where $\mathcal{F}$ is strictly convex.

**Step 2: LMS outer problem:** Theorem 2 shows us how to characterize the incentive-feasible set for any type distribution. In particular, it allows us to easily compute the support function for this set. We now use this characterization to identify the optimal mechanism for the LMS problem. We begin with a convex set $\mathcal{F}_j \subset \mathbb{R}^2_+$ with a support function $S_j : \mathbb{R}^2 \to \mathbb{R}_+$ for each $j$. Let $N_j^*(a,b) := \arg\max\{a\hat{n}^h + b\hat{n}^l : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}_j\}$. We study the optimal division of cases among the investigators, such that the caseload for each $j$ is an element of $\mathcal{F}_j$.[19] That is, we want to solve:

$$\min_{(\hat{n}_j^h, \hat{n}_j^l)_{j=1}^J} \sum_{j=1}^J c^h(j)\hat{n}_j^h + c^l(j)\hat{n}_j^l \qquad s.t. \ \ (\hat{n}_j^h, \hat{n}_j^l) \in \mathcal{F}_j \quad \forall\, 1 \le j \le J \tag{3}$$

$$\sum_{j=1}^J \hat{n}_j^h = Jn^h \quad , \quad \sum_{j=1}^J \hat{n}_j^l = Jn^l$$

[19] An interesting subtlety can arise when $F$ is not regular and the frontier has linear segments. Theorem 2 tells us that for any $(a,b)$, the *value* $S(a,b)$ can be achieved with a mechanism such that $H$ takes on no more than four distinct values. But this does not mean that every *point* on the frontier can be implemented with such a mechanism. For any point $(\hat{n}^h, \hat{n}^l)$ that lies on a linear segment of the frontier, it may in fact be necessary to use a mechanism that takes five distinct values. The details are omitted since we focus on the regular case.

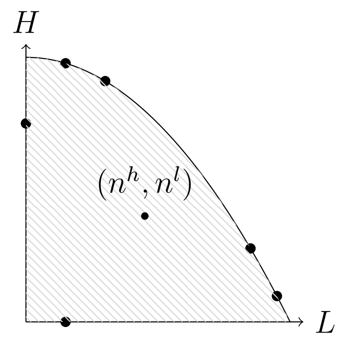



Figure 2: Outer problem with identical (strongly regular) type distributions

If type distributions are symmetric, so that $\mathcal{F}_j = \mathcal{F}$ for all $j$, then we can think of this as the problem of choosing a set of points $\mathcal{F}$ with barycenter equal to $(n^h, n^l)$, as illustrated in Figure 2, where $\mathcal{F}$ is depicted as the shaded region.

We solve the outer problem by studying its dual. Let $\lambda_h, \lambda_l$ be the dual variables for the market-clearing constraints. Then, the previous program is equivalent to

$$\min_{(\hat{n}_j^h, \hat{n}_j^l)_{j=1}^J} \max_{\lambda_h, \lambda_l} \sum_{j=1}^J c^h(j)\hat{n}_j^h + c^l(j)\hat{n}_j^l + \lambda_h \left( Jn^h - \sum_{j=1}^J \hat{n}_j^h \right) + \lambda_l \left( Jn^l - \sum_{j=1}^J \hat{n}_j^l \right)$$

$$s.t. \qquad (\hat{n}_j^h, \hat{n}_j^l) \in \mathcal{F}_j \quad \forall\, 1 \leq j \leq J \qquad \text{(incentive-feasible)}$$

Strong duality holds (see Appendix A.3), so this is equivalent to

$$\max_{\lambda_h, \lambda_l} \ \lambda_h J n^h + \lambda_l J n^l - \sum_{j=1}^J \max \left\{ \left( \lambda_h - c^h(j) \right) \hat{n}^h + \left( \lambda_l - c^l(j) \right) \hat{n}^l \quad : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}_j \right\} \tag{4}$$

In words, the dual variables $\lambda_h, \lambda_l$ are the marginal social costs of adding high- and low-type cases, respectively. Alternatively, we can view $\lambda_h, \lambda_l$ as the "benefit" paid to the designer for completing each case. Then, the dual program can be interpreted as a competitive market in which firms seek to maximize cost minus benefit, and $\lambda_h, \lambda_l$ adjust to clear the market.

Using the definition of the support function, we can rewrite the dual as

$$\max_{\lambda_h, \lambda_l} \ \lambda_h J n^h + \lambda_l J n^l - \sum_{j=1}^J S_j \left( \left( \lambda_h - c^h(j) \right), \left( \lambda_l - c^l(j) \right) \right) \tag{5}$$

Support functions are always convex and continuous. Thus, the objective in (5) is concave in $(\lambda_h, \lambda_l)$. Using this formulation, we can simplify the outer problem of choosing incentive-feasible pairs $(\hat{n}_j^h, \hat{n}_j^l)$ for each investigator, to the much simpler two-dimensional dual. Moreover, this formulation allows us to identify quantitative features of the solution and perform comparative statics. Recall that we defined $N_j^*(a, b) := \arg\max\{a\hat{n}^h + b\hat{n}^l : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}_j\}$.

**Theorem 3.** Let $(\lambda_h, \lambda_l)$ solve the dual program in (5). Then, there exist selections from $N_j^*\left(\lambda_h - c^h(j), \lambda_l - c^l(j)\right)$ such that:

$$\sum_{j=1}^{J} \hat{n}_j^h = Jn^h \quad \text{and} \quad \sum_{j=1}^{J} \hat{n}_j^l = Jn^l,$$

and these constitute a solution to the social-cost minimization program. If $c^k(j) \neq c^k(j')$ for all $j, j' \in \mathcal{J}$ and $k \in \{h, l\}$, then in any solution at most one investigator is off the boundary of $\mathcal{F}_j$, and at most two investigators have non-zero allocations that are off the frontier. If, moreover, every investigator also has a strongly regular type distribution then there is a unique solution, $(\hat{n}_j^h, \hat{n}_j^l)_{j=1}^J$, to the social-cost minimization problem.

*Proof.* Proof in Appendix A.3. □

The optimal mechanism is determined by the performance parameters (through the objective function), and the type distributions (which determine the incentive-feasible sets $\mathcal{F}_j$). The allocation rule for investigator $j$ is the solution to the LMS inner problem for $j$ for weights $(a, b) = (\lambda_h - c^h(j), \lambda_l - c^l(j))$. As discussed above, if each $F_j$ is strongly regular then this allocation is characterized by a pair of prices $(p_1^j, p_2^j)$ at which $j$ is allowed to trade given their induced kinked budget set. We refer to the optimal mechanism under strong regularity as the *LMS two-price (LMS-TP) mechanism*, and focus primarily on this case.

We refer to investigators whose assignment is off the frontier as *remedial*. Such agents belong to one of four categories. If $\lambda_h - c^h(j) = 0$ and $\lambda_l - c^l(j) < 0$ then $j$ receives only type-$h$ cases, if $\lambda_l - c^l(j) = 0$ and $\lambda_h - c^h(j) < 0$ then they receive only type-$l$ cases, and if both $\lambda_l - c^l(j) < 0$ and $\lambda_h - c^h(j) < 0$ then $j$ receives no cases. The remaining remedial agents are those with $\lambda_l - c^l(j) = \lambda_h - c^h(j) = 0$, and only these can receive allocations in the interior of $\mathcal{F}_j$. Note that except for this last group, the assignment of all other remedial agents is independent of their type.

By the envelope theorem (Milgrom and Segal, 2002), $S$ is differentiable almost everywhere, and if $N_j^*(a, b) = (x^*, y^*)$ then the right derivative of $S_j(a, b)$ with respect to the first argument is $\max\{x^*\}$, and with respect to the second argument is $\max\{y^*\}$. Along with Theorem 3, having access to this derivative facilitates efficient computation of the optimal mechanism. Moreover, the dual formulation of the outer problem allows us to perform comparative statics and study investigators' incentives for effort.

*Remark* 3. The same approach of dividing the problem into inner and outer components can be applied regardless of the size of the partition used to define case types. Indeed, the outer problem is not fundamentally different in higher dimensions. The challenge lies in solving the inner problem when preferences are multi-dimensional, as discussed above. However an approximate solution to the inner problem could be used to characterize a lower bound on the support function. This in turn could be used to identify an approximate solution to the outer problem, and thus to the full program. We leave this extension for future work.

## II.C Fairness and incentives for effort

Our approach treats $c^h(j)$ and $c^l(j)$ as policy-invariant parameters. However, a natural concern in any performance-based assignment mechanism is whether it gives agents the right performance incentives. This concern is inherently dynamic: agents might intentionally perform worse today if they expect their performance data to be used in the future to re-design the mechanism, particularly if low-performing workers are rewarded with lower caseloads. Fortunately, we show that in our mechanism the scope for such manipulation is limited.

So far, we fixed $(c^h(j), c^l(j))_{j\in J}$ and defined a mechanism as a function of the type profile. To talk about the agents' incentives to perform, we need to make explicit the mechanism's dependence on the performance parameters. We thus think of the mapping from $(c^h(j), c^l(j))_{j\in J}$ to the LMS-TP mechanism as itself a meta-mechanism mapping performance parameters and type reports to allocations. We refer to this simply as the *optimal LMS-TP mechanism*.

The question of whether the optimal LMS-TP mechanism delivers the correct incentives for effort is fundamentally about its comparative statics in $(c^h(j), c^l(j))_{j\in J}$. Consider, for example, an investigator who improves their performance on type-$h$ cases, holding that on type-$l$ cases fixed. Intuitively, the mechanism should try to assign this investigator to more type-$h$ cases in expectation. From an ex-ante perspective, i.e. without knowing the investigator's type, there are two ways to do this: $(i)$ ensure that the investigator receives a large number of high-type cases in the event that they report a low type, or $(ii)$ increase the size of this event, i.e. the probability that the investigator specializes in high-type cases. These two objectives are at odds: we must lower $p_1^j$ to achieve $(i)$, and raise it to achieve $(ii)$.[20] Notice, however, that if the second objective dominates, the investigator receives better terms for exchanging for high-type cases the better their performance on these cases. Indeed, this is the case under the optimal LMS-TP mechanism under a slight strengthening of strong regularity.[21]

**Proposition 1.** Assume $F_j$ is strongly regular and $F(p)/p$ is increasing. Then $p_1^j$ is decreasing

[20] The intuition here is incomplete, since we also need to consider changes in $p_2^j$. The proof in Appendix C.1 deals with this additional complication.

[21] Recall that strong regularity is defined as strict regularity, non-negative virtual values, and $F_j(\bar{p})/\bar{p} > F(p)/p \ \forall \ p$.

in $c^h(j)$ and $p_2^j$ is increasing in $c^l(j)$ for any $j$ on their frontier.

In other words, conditional on specializing in type-$k$ cases, $j$ benefits from reducing $c^k(j)$. Proposition 1 is proven along with Theorem 4 in Appendix B.1, using the dual in eq. (5). The intuitive connection between Proposition 1 and effort incentives is clear. Appendix B.1 formalizes this relationship and clarifies an additional connection to fairness.

# III Small-Market Static and Dynamic problems

## III.A Small market static

In the previous section, we solved a relaxed problem in which we only required that all cases be assigned in expectation. The problem is more difficult if we require that all cases be assigned ex-post, i.e., conditional on each realized type profile—rather than just in expectation. While it may be possible to solve for the optimal Bayesian incentive compatible (BIC) mechanism in the SMS problem directly, doing so would sacrifice the simplicity of the LMS-TP mechanism.[22] We opt instead to modify the LMS solution to accommodate the ex-post market-clearing condition. This approach has the additional benefit that we obtain a mechanism which is strategy proof (in fact Obviously Strategy-Proof as in Li (2017)) and robust to misspecification of the type distribution. This would not be the case for the optimal BIC mechanism. The trade-off is that our solution is only approximately optimal.

Assume that all agents' type distributions are strongly regular.[23] Let $\mathcal{E}$ be the set of non-remedial agents in the optimal LMS-TP mechanism, i.e. those agents whose expected allocation is on the frontier of their incentive-feasible set. Let $(p_1^j, p_2^j)_{j\in\mathcal{E}}$ be the prices defining the optimal LMS-TP mechanism for the non-remedial agents. Let $P = (p_j)_{j=1}^J$ be a type profile. Fixing the mechanism, $j \in \mathcal{E}$ is a *buyer* (of type-$h$ cases) if $p_j \leq p_1^j$, a *seller* if $p_j > p_2^j$, and retains the status quo otherwise. Let $\mathcal{B}$ be the set of buyers and $\mathcal{S}$ the set of sellers.

We wish to approximate the LMS-TP mechanism in the small market, while ensuring ex-post market clearing. The basic idea is to offer each non-remedial investigator the same prices $(p_1^j, p_2^j)$ that they would be given in LMS-TP, and ask them whether they would like to trade for high-type cases, trade away high-type cases, or retain the status quo. They are then guaranteed an allocation that lies on their budget set, in the direction that they choose to move. However, whereas in the LMS-TP mechanism all agents either retain the status quo or

[22]The main challenge is that in the small-market problem there are additional constraints on the interim allocations and the standard characterization of Border (1991) and Che et al. (2013) does not apply to this multi-item setting. More recent developments in Valenzuela-Stookey (2023) can be used to provide a characterization of interim allocations for this setting, but nonetheless how to design the optimal mechanism remains an open question.

[23]We can apply a similar approach without strong regularity.

go to a corner, here we may restrict the volume of trade. For remedial agents the assignment is independent of their reported type. Subject to these constraints, and market clearing, the allocation is chosen to minimize social cost. Concretely, the mechanism is the following.

**Small-market static two-price (SMS-TP) mechanism.**

- Assign $(n^h, n^l)$ to all $j \in \mathcal{E} \setminus (\mathcal{B} \cup \mathcal{S})$, i.e., the non-remedial agents who are neither buyers nor sellers.
- To allocate the remaining cases, we solve the linear program:

$$\min_{\{b_j\}_{j\in\mathcal{B}},\{s_j\}_{j\in\mathcal{S}},\{\dot{n}^h_j,\dot{n}^l_j\}_{j\notin\mathcal{E}}} \quad \sum_{j\in\mathcal{B}} b_j\left(c^h(j) - p_1^j c^l(j)\right) - \sum_{j\in\mathcal{S}} s_j\left(c^h(j) - p_2^j c^l(j)\right) + \sum_{j\notin\mathcal{E}} \dot{n}^h_j c^h(j) + \dot{n}^l_j c^l(j)$$

$$\begin{aligned} s.t. \quad & \bar{p}_j \dot{n}^h_j + \dot{n}^l_j \le \bar{p}_j n^h_j + n^l_j \quad \forall\, j \notin \mathcal{E} \\ & \underline{p}_j \dot{n}^h_j + \dot{n}^l_j \le \underline{p}_j n^h_j + n^l_j \quad \forall\, j \notin \mathcal{E} \\ & 0 \le \dot{n}^h_j,\ 0 \le \dot{n}^l_j \quad \forall\, j \notin \mathcal{E} \\ & 0 \le b_j \le \frac{n^l}{p_1^j} \quad \forall\, j \in \mathcal{B} \\ & 0 \le s_j \le n^h \quad \forall\, j \in \mathcal{S} \end{aligned}$$

$$\sum_{j\in\mathcal{B}}(n^h + b_j) + \sum_{j\in\mathcal{S}}(n^h - s_j) + \sum_{j\notin\mathcal{E}} \dot{n}^h_j + |\mathcal{E} \setminus (\mathcal{B}\cup\mathcal{S})| n^h = J n^h \quad (h\text{-capacity})$$

$$\sum_{j\in\mathcal{B}}(n^l - p_1^j b_j) + \sum_{j\in\mathcal{S}}(n^l + p_2^j s_j) + \sum_{j\notin\mathcal{E}} \dot{n}^l_j + |\mathcal{E} \setminus (\mathcal{B}\cup\mathcal{S})| n^l = J n^l \quad (l\text{-capacity})$$

  Let $(b_j^*)_{j\in\mathcal{B}}, (s_j^*)_{j\in\mathcal{S}}, (\dot{n}^h_j, \dot{n}^l_j)_{j\in\mathcal{E}}$ be the solution to this program. The assignment for $j \in \mathcal{E}$ is $(\dot{n}^h_j, \dot{n}^l_j)$. For $j \in \mathcal{B}$ the assignment is $(n^h + b_j^*, n^l - p_1^j b_j^*)$, and for $j \in \mathcal{S}$ it is $(n^h - s_j^*, n^l + p_2^j s_j^*)$.

To understand the asymptotic performance of our mechanism in finite markets, it is convenient to introduce a slight perturbation of SMS-TP, which we call SMS-TP'. The difference between the two mechanisms concerns only the allocation for investigators such that $\lambda^h - c^h(j) = \lambda^l - c^l(j) = 0$ (of which generically there is at most one). The modification ensures that the SMS-TP mechanism converges to LMS-TP as the market grows. However, the small perturbation from SMS-TP to SMS-TP' makes the mechanism cumbersome to state, and so a full description is deferred to Appendix C.2. In small markets SMS-TP and SMS-TP' are nearly identical, and for simplicity we focus primarily on the former.

Formally, consider a sequence of "replica economies" in which there are $y$ copies of each investigator. Let $V_{SMS}(\{F\}_{j=1}^J|y)$ be the expected social cost achieved by SMS-TP' in the $y$-replica economy, given the profile of type distributions $\{F\}_{j=1}^J$. Let $V_{OPT}(\{F\}_{j=1}^J|y)$ be the cost achieved by the (unknown) optimal SMS mechanism. The source of the divergence between the small market and large market is that in the former we do not know ex-ante the mass of investigators who will be buyers and sellers of high-type cases. Unsurprisingly, the SMS-TP' mechanism is a better approximation to the optimal mechanism as this aggregate uncertainty about the agents' types decreases, so that the small market approaches the large-market idealization. A mechanism is *strategy-proof* if truthful reporting is optimal for each agent, regardless of the type reports made by others.

**Proposition 2.** In the small-market static setting, SMS-TP and SMS-TP' are strategy-proof and respect the status quo. Moreover, assuming $F_j$ satisfies strong regularity for all $j \in \mathcal{J}$, the value $V_{SMS}(\{F\}_{j=1}^J|y)$ converges to $V_{OPT}(\{F\}_{j=1}^J|y)$ as either

i. $y \to \infty$, and/or

ii. $F_j$ converges in distribution to a constant for all $j$.

*Proof.* Proof in Appendix C.2. □

Strategy-proofness follows from linearity of the investigators' preferences and the fact that prices are fixed ex-ante. In fact, it is easy to see that the SMS-TP (and SMS-TP') is Obviously Strategy-Proof (Li, 2017). However, as an indirect mechanism SMS-TP (and SMS-TP') can be implemented even if investigators' preferences are not linear. To do this we can ask each investigator to choose between being a buyer, a seller, or retaining the status quo at the stated prices, rather than eliciting type reports directly. From the designer's perspective, the worst possible outcome is that all investigators report that they would like to retain the status quo. Thus, the designer can never do worse than the status quo, and does strictly better as long as some investigators are willing to be buyers and sellers.

Additionally, the mechanism is robust to misspecification of the type distribution.

**Proposition 3.** Regardless of the true type distributions, the SMS-TP and SMS-TP' mechanisms based on distributions $(F_j)_{j\in\mathcal{J}}$ are (obviously) strategy-proof and respect the status-quo constraint. Moreover, the expected social cost of the mechanism is no worse than that of the status quo.

Proposition 3 is an immediate implication of the fact that prices are fixed and the status quo is feasible in the linear program defining SMS-TP.

## III.B Small market dynamic

Ultimately, the setting in which we are interested is inherently dynamic: cases arrive over time and must be assigned "online" without knowledge of future arrivals. To go from the static to the dynamic setting, we develop a mechanism to approximately implement the SMS-TP mechanism, where the approximation in this case gets better the longer the time horizon.

The dynamic model is as follows. Time is continuous and runs from 0 to $T$.[24] At each instant one case may arrive. Let $\tau_t \in \{h, l, 0\}$ be the type of the case in period $t$, where $\tau_t = 0$ if no case arrives in period $t$. Denote by $N^k(t)$ the number of type-$k$ cases which have arrived up to and including time $t$, and let $\bar{n}^k(t) = \frac{1}{J} N^k(t)$.

Agents report their type only once, at time zero. The payoff of agent $j$ who receives a cumulative caseload of $(\hat{n}_j^h, \hat{n}_j^l)$ by time $T$ is $p_j \hat{n}_j^h + \hat{n}_j^l$. That is, agents care about their total undiscounted workload.[25] We start by estimating the total number of type $k$ cases that will arrive by time $T$; denote the estimate by $n^k$ for $k \in \{h, l\}$. Given $(n^h, n^l)$, we solve for the SMS-TP' assignment, which we denote by $(\dot{n}_j^h, \dot{n}_j^l)_{j=1}^J$.

Index each case by the time at which it arrives. Let $z_t$ be the investigator to which case $t$ is assigned. For each $j$ we keep track of their running case count $\hat{n}_j^k(t) := \sum_{i=1}^t \mathbb{1}[z_i = j, \tau_i = k]$. Define the *score* $r_j(t, k) = \hat{n}_j^k(t)/\dot{n}_j^k$, where $r_j(t, k) = \infty$ if $\dot{n}_j^k = 0$.

**SMD-TP mechanism.** For each time $t$ at which a case arrives, assign it to the investigator with the lowest value of $r_j(t, \tau_t)$ (using any tie-breaking rule).

Let $V_{SMD}((F_j)_{j=1}^J, A, T)$ be the value of SMD-TP mechanism given a sequence of case arrivals $A$. Abusing notation, let $V_{SMS}((F_j)_{j=1}^J, A, T)$ be the value of the SMS-TP' mechanism given the aggregate case counts from sequence $A$ over time horizon $T$. Say that a mechanism is $\varepsilon$-IC ($\varepsilon$-SQ) if for any agent the ratio of the expected payoff of truthful reporting to that of any deviation (to the status quo) is at least $1 - \varepsilon$.

For concreteness, assume that high- and low-type cases arrive at Poisson rates $\rho^h$ and $\rho^l$ respectively. In this case $n^k = \frac{T}{J} \rho^k$. This assumption is not necessary, but simplifies the exposition.[26]

[24] The assumption of continuous time simplifies the discussion here but has no bearing on the result. The algorithm in the empirical application is modified to run in discrete time.

[25] Ultimately, every case needs to be assigned as it comes, so there would be no scope for the designer to take advantage of agents' discounting of future payoffs by back-loading cases. Moreover, the mechanism we propose here smooths each investigator's workload evenly over time regardless of their type report, so discounting should not significantly affect incentives to report truthfully.

[26] What we need for the results is that $\frac{1}{T}\left(N^k(T) - T\rho^k\right) \xrightarrow{a.s.} 0$ as $T \to \infty$.

Intuitively, the SMD-TP algorithm tries to allocate a case of type-$k$ so as to move each agent towards their target caseload $\dot{n}_j^k$ in a way that smooths assignments over time. How well the algorithm can do this depends on how far $N^h(T)$ and $N^l(T)$ are from their expected values $\rho^h T$ and $\rho^l T$. The algorithm improves as $T$ increases, since by the strong law of large numbers, $\frac{1}{T}\left(N^k(T) - T\rho^k\right) \xrightarrow{a.s.} 0$ as $T \to \infty$.

**Proposition 4.** $\frac{V_{SMD}((F_j)_{j=1}^J, A, T)}{V_{SMS}((F_j)_{j=1}^J, A, T)} \xrightarrow{a.s.} 1$ as $T \to \infty$. Moreover, for any $\varepsilon > 0$ there exists $\bar{T}$ such that the SMD-TP mechanism is $\varepsilon$-IC and $\varepsilon$-SQ for any $T \geq \bar{T}$.

*Proof.* By the strong law of large numbers, $\frac{1}{T}\left(N^k(t) - T\rho^k\right) \xrightarrow{a.s.} 0$ as $T \to \infty$. Then, by construction, for each $j \in J$ and $k \in \{h, l\}$, we have $r_j(T,k) \xrightarrow{a.s.} 1$ as $T \to \infty$, and so the mechanism is $\varepsilon$-IC for large enough $T$. Note also that $\frac{1}{T}V_{SMD}((F_j)_{j=1}^J, A, T)$ is just a weighted sum of $(\hat{n}_j^h, \hat{n}_j^l)_{j=1}^J$, and $\frac{1}{T}V_{SMS}((F_j)_{j=1}^J, A, T)$ is a weighted sum of $(\dot{n}_j^h, \dot{n}_j^l)_{j=1}^J$. Convergence of the ratio of values follows from convergence of $r_j(T,k)$ for all $j \in \mathcal{J}$ and $k \in \{h, l\}$. □

In addition to targeting the aggregate caseloads $(\dot{n}_j^h, \dot{n}_j^l)$, the SMD-TP mechanism also attempts to smooth the arrivals over time. This is the benefit of using the ratio $r_j(t,k) = \frac{\hat{n}_j^k(t)}{\dot{n}_j^k}$ to assign cases, as opposed to the difference $\hat{n}_j^k(t) - \dot{n}_j^k$; the latter would front-load cases to investigators with high targets. On the other hand, this method is somewhat extreme in that it never assigns type-$k$ cases to an investigator $j$ with $\dot{n}_j^k = 0$. Assigning based on the difference between target and realized caseloads would ensure that this difference is small, even if it means giving a few type-$k$ cases to investigators with $\dot{n}_j^k = 0$. In Appendix B.4, we discuss finite-sample adjustments to the assignment rule which move between these extremes.

The dynamic mechanism inherits, up to the approximation error associated with predicting aggregate case arrivals, the robustness properties of the SMS-TP mechanism (Proposition 3). In the dynamic setting, this allows us to implement our mechanism without relying on distributional assumptions. We can do this by implementing the mechanism first under some initial specification of the type distribution for some period, say a year. The types reported in the first year can then be used to update the estimate of the type distribution used in the second year, and so on. This updating need not interfere with incentives, as we can base the estimate of $F_j$ only on the reports of other agents. Under mild assumptions, the estimated distributions converge to their true values.

# IV Empirical Application

Having characterized the optimal mechanism, we now turn to an empirical application that quantifies its potential impact. Our goal is to use data to simulate how the mechanism would perform relative to the status quo. We use administrative data on investigator assignments and child outcomes to identify cross-investigator differences in performance. We then combine these estimates with survey evidence on investigator preferences to estimate the type distribution, which allows us to simulate the implementation of the mechanism.[27]

## IV.A Identification

**Social Cost Parameters.** Thus far, we have assumed that the designer directly observes each investigator's cost parameters $c^k(j)$. However, these parameters are not point-identified, even under quasi-random assignment of investigators to cases. This reflects the standard "selective labels" problem: for children placed in foster care, we do not observe whether maltreatment would have occurred had they remained at home, making it impossible to separately identify true and false positives.

Our strategy is to instead identify differences in costs across investigators, rather than levels. These differences are sufficient to construct social preferences over mechanisms and hence to evaluate welfare comparisons. Although $c^k(j)$ is not non-parametrically identified, the difference $c^k(j) - c^k(j')$ is, provided that investigators $j$ and $j'$ are quasi-randomly assigned to cases and an exclusion restriction holds.

**Setup.** Consider a CPS office where cases are quasi-randomly assigned to investigators. Let $D_{ij} \in \{0, 1\}$ denote the potential placement decision of investigator $j$ on case $i$ ($D_{ij} = 1$ if the child is placed in foster care). Let $Y_i^* \in \{0, 1\}$ be the child's maltreatment potential ($Y_i^* = 1$ if maltreatment occurs). Then, the *potential* outcome for subsequent maltreatment if case $i$ were assigned to $j$ is $Y_{ij} := (1 - D_{ij})Y_i^\star$. Note that $Y_i^\star$ is selectively observed based on the assigned investigator and their potential decision: we observe $Y_i^\star$ if and only if case $i$ is assigned to an investigator $j$ satisfying $D_{ij} = 0$.

Normalizing $c_{\text{TN}} = 0$ (without loss of generality) we would ideally identify $(\text{FN}_{ij}, \text{FP}_{ij}, \text{TP}_{ij})$ and thus $c(i, j)$. Under random assignment, false negatives $\text{FN}_{ij}$ and true negatives $\text{TN}_{ij}$ are identified because $Y_i^*$ is observed when the child is not placed. However, when $D_{ij} = 1$, $Y_i^*$ is unobserved, so $\text{FP}_{ij}$ and $\text{TP}_{ij}$ are not nonparametrically identified. Fortunately, while this means we cannot identify the social cost function $c(i, j)$ without further assumptions, we next provide identification results which demonstrate how one can identify *social preferences*, i.e., the ranking over the set $\mathcal{Z}$ of possible assignments, under the random assignment assumption and an additional exclusion restriction.

**Identification of Differences.** Let $I \subset \mathcal{I}$ be a subset of cases. We say that assignment $Z$ is *random conditional on* $I$ if $(D_{ij}, Y_i^*) \perp\!\!\!\perp Z_{ij}$ conditional on $i \in I$, for all $j \in J$. Investigator

[27]See Appendix F for additional background on the CPS system and foster care processes.

$j$'s assignment is *supported on I* if $Pr(\{i \in I, Z_{ij} = 1\}) \neq 0$. Let $\text{FN}_j^I := \mathbb{E}[\text{FN}_{ij}|i \in I]$ and similarly for $\text{TN}_j^I$, $\text{TP}_j^I$, and $\text{FP}_j^I$.

**Lemma 1.** Assume that the observed assignment is random conditional on $I$. Then, for any $j, j' \in \mathcal{J}$ whose assignments are supported on $I$, the following three differences are identified: $\text{FP}_j^I - \text{FP}_{j'}^I$, $\text{TP}_j^I - \text{TP}_{j'}^I$, and $\mathbb{E}[c(i,j) - c(i,j')|i \in I]$.

*Proof.* First, recall that $Y_{ij}$ and $D_{ij}$ are observed for the set of cases when $Z_{ij} = 1$. Then, under random assignment conditional on $I$, we have

$$\text{FN}_j^I := \mathbb{E}[\text{FN}_{ij}|i \in I] = \mathbb{E}[Y_i^\star(1 - D_{ij})|i \in I] = \mathbb{E}[Y_{ij}|i \in I] = \mathbb{E}[Y_i|i \in I, Z_{ij} = 1]$$

and $P_j^I := \mathbb{E}[D_{ij}|i \in I] = \mathbb{E}[D_i|i \in I, Z_{ij} = 1]$. Moreover, we can express $\text{TN}_j^I$ as $\text{TN}_j^I = 1 - \left(\text{TP}_j^I + \text{FP}_j^I\right) - \text{FN}_j^I = 1 - P_j^I - \text{FN}_j^I$. Thus, $\text{FN}_j^I$, $\text{TN}_j^I$, and $P_j^I$ are identified if $j$'s assignment is supported on $I$. Let $S_j^I = \text{TP}_j^I + \text{FN}_j^I$. Note that, under random assignment conditional on $I$, $S_j^I = S_{j'}^I = \mathbb{E}[Y_i^\star|i \in I]$ for all $j, j' \in \mathcal{J}$. Then,

$$\begin{aligned}\text{FP}_j^I - \text{FP}_{j'}^I &= \left(1 - \text{TP}_j^I - \text{FN}_j^I - \text{TN}_j^I\right) - \left(1 - \text{TP}_{j'}^I - \text{FN}_{j'}^I - \text{TN}_{j'}^I\right) \\ &= \left(1 - S_j^I - \text{TN}_j^I\right) - \left(1 - S_{j'}^I - \text{TN}_{j'}^I\right) = -\left(\text{TN}_j^I - \text{TN}_{j'}^I\right).\end{aligned}$$

Similarly, $\text{TP}_j^I - \text{TP}_{j'}^I = -\left(\text{FN}_j^I - \text{FN}_{j'}^I\right)$. This is sufficient to also identify the cost differences. □

Lemma 6, presented in Appendix E.3, generalizes Lemma 1 beyond binary $Y_i^*$.

**Example.** To build intuition for Lemma 1, consider two investigators $j$ and $j'$ in the same office, with cases assigned as if at random. Random assignment implies that both investigators face the same underlying rate of true maltreatment: $\mathbb{E}[Y_i^* \mid Z_{ij} = 1] = \mathbb{E}[Y_i^* \mid Z_{ij'} = 1] = S$. Suppose investigator $j$ has a false negative rate $\text{FN}_j = 0.09$ and placement rate $P_j = 0.02$, while investigator $j'$ has $\text{FN}_{j'} = 0.08$ and $P_{j'} = 0.03$. Investigator $j'$ therefore places one percentage point more children and has one percentage point fewer false negatives than $j$.

What can we infer about their false positive rates? Although $S$ itself is unobserved, differences in false positive rates can still be recovered using the result above that $\text{FP}_j - \text{FP}_{j'} = (P_j + \text{FN}_j) - (P_{j'} + \text{FN}_{j'})$. Plugging in the numbers, $\text{FP}_j - \text{FP}_{j'} = 0.11 - 0.11 = 0$.

The interpretation is that the additional placements by $j'$ exactly offset the reduction in false negatives. Because both investigators see cases with the same underlying risk $S$, the two must have the same false positive rate. This illustrates the point of the lemma: while the absolute level of $S$ is not observed, *differences* in error rates across investigators are identified under random assignment.

**Assumptions.** The identification results rest on two key assumptions. First, investigators must be quasi-randomly assigned to cases within a CPS office. This assumption has been carefully examined in the Michigan CPS context (e.g., see Baron et al. (2024)). As discussed in Appendix F, investigators are assigned cases rotationally, following a queue of who is "next up," rather than through matching of specific investigators to specific families. Second, note that $(D_{ij}, Y_i^*) \perp\!\!\!\perp Z_{ij}$ also embeds an exclusion restriction: investigators can only "reveal" $Y_i^*$ through their placement decisions, and cannot otherwise directly affect it. This restriction has also been probed in this prior work in our context. The intuition is that investigators' primary influence on children's outcomes is the foster care decision. Other actions, such as referrals to preventive services, have been shown to have little impact on child maltreatment risk. Thus, conditional on placement, investigators are unlikely to directly shift $Y_i^*$.

**Computing Welfare Differences Across Mechanisms.** We next show how Lemma 1 can be used to recover social preferences over assignment mechanisms.[28]

First, note that the parameter $c^k(j)$ is a natural measure of the performance of investigator $j$ on cases of type $k$. While $c^k(j)$ is not non-parametrically identified, the difference $c^k(j)-c^k(j')$ is identified for any $j, j'$ and is given by:

$$c^k(j) - c^k(j') = c_{FP}\left(P_j^{I_k} - P_{j'}^{I_k}\right) + (c_{FP} + c_{FN} - c_{TP})\left(\text{FN}_j^{I_k} - \text{FN}_{j'}^{I_k}\right).$$

Defining $\gamma_j^k := c_{FP}P_j^{I_k} + (c_{FP} + c_{FN} - c_{TP})\text{FN}_j^{I_k}$, we have that $c^k(j) \leq c^k(j')$ if and only if $\gamma_j^k \leq \gamma_{j'}^k$. Intuitively, $\gamma_j^k$ tells us the position of investigator $j$ in the distribution of investigator performance among cases of type $k$. We therefore refer to $\gamma_j^k$ as $j$'s *performance score* on type-$k$ cases, where a lower score corresponds to greater performance. As we show next, $\gamma_j^k$ is a useful parameter because it can be used to compute social preferences.

Let $\{I_k\}_{k=1}^K$ be a partition of $\mathcal{I}$ into disjoint sets. Say that $Z$ is *conditionally random* given partition $\{I_k\}_{k=1}^K$ if it is random conditional on $I_k$ for all $K$. Say that it *has full support* if every agent's assignment is supported on $I_k$, for all $k$.

[28] We are certainly not the first to note that differences in false positive rates can be expressed in terms of placement rates, baseline maltreatment risk $S$, and false negative rates. Our contribution in this section is to show that, for the purpose of identifying social preferences over assignment mechanisms, it is sufficient to focus on *differences* in false positives and true positives across investigators. This is in contrast to prior work, which has aimed to directly identify the *levels* of false and true positives—typically by structurally estimating $S$ (Chan et al., 2022) or by relying on alternative strategies such as contraction (Kleinberg et al., 2018) or identification at infinity (Arnold et al., 2022; Angelova et al., 2023).

**Corollary 3.** Let $Z$ be an observed assignment that is conditionally random and has full support given a partition $\{I_k\}_{k=1}^K$ or $\mathcal{I}$. Then $\mathbb{E}[C(Z)] - \mathbb{E}[C(Z')]$ is identified for any other assignment $Z'$ that is conditionally random given the same partition, and equal to:

$$\sum_{k=1}^{K} \sum_{j \in \mathcal{J}} \gamma_j^k \sum_{i \in I_k} Z_{ij} - \sum_{k=1}^{K} \sum_{j \in \mathcal{J}} \gamma_j^k \sum_{i \in I_k} Z'_{ij}.$$

That is, under the conditions of Corollary 3, the cardinal ranking over $\mathcal{Z}$ is non-parametrically identified. Note that from Lemma 1 we can also identify the expected difference in false negatives, false positives, and the placement rate across the two assignments.

*Remark* 4. One useful application of Lemma 1 is to pick an arbitrary investigator, $j'$, and define $\tilde{c}(i,j) = c(i,j) - c(i,j')$. Then, we can replace the objective of the designer, $C(Z') := \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} Z'_{ij} c(i,j)$, with $\tilde{C}(Z') := \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} Z'_{ij} \tilde{c}(i,j) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} Z'_{ij} c(i,j) - \sum_{i \in \mathcal{I}} c(i,j')$, where the last equality follows from the fact that under each $Z' \in \mathcal{Z}$, each case is assigned to exactly one investigator. Since $\sum_{i \in \mathcal{I}} c(i,j')$ does not depend on $Z'$, $\tilde{C}$ and $C$ represent the same preferences over $\Delta(\mathcal{Z})$. Moreover, if the observed assignment $Z$ is conditionally random and has full support given a partition $\{I_k\}_{k=1}^K$ then $\mathbb{E}[\tilde{c}(i,j) | i \in I_k]$ is identified for all $k$, and $\mathbb{E}[\tilde{C}(Z')]$ is identified if $Z'$ is conditionally random given the same partition.

Finally, we consider an investigator's *relative advantage* across case types, $c^l(j) - c^h(j) = \delta_j$. While relative advantage is not identified, differences in relative advantage, or the *comparative advantage* of $j$ relative to $j'$, is: $D(j,j') = \delta_j - \delta_{j'} = c^l(j) - c^l(j') - \big(c^h(j) - c^h(j')\big)$. When high- and low-type cases are equally costly for all investigators, comparative advantage is the sufficient statistic for the optimal assignment. Let $d_j := \gamma_j^l - \gamma_j^h$ be investigator $j$'s *comparative advantage score*, which can be used to rank investigators in terms of comparative advantage. Below, we use this result to document evidence of comparative advantage across investigators within offices in our data.

## IV.B Data sources and analysis sample

We obtained the universe of child maltreatment investigations in Michigan between January 2008 and November 2016 from the Michigan Department of Health and Human Services. The dataset includes the allegation report date as well as child and investigation traits including the child's zip code, age, gender, race, relationship to the alleged perpetrator, and maltreatment type (e.g., physical abuse versus neglect). It also includes indicators for whether the child was placed in foster care following the investigation and investigator numeric identifiers.

To construct the analysis sample, we begin with the set of child maltreatment investigations that did not involve either sexual abuse or repeat reports since these cases are not quasi-randomly assigned to investigators. Given that foster care placement rates are low, we drop cases assigned to investigators who handled fewer than 200 investigations to minimize noise in our estimates of investigator placement rates ($N = 152,686$). We then drop observations in rotations (zip code by year pairs) with fewer than four investigators to compare investigators in a given "office" by year ($N = 22,201$). Furthermore, we drop cases for which we cannot observe subsequent child welfare outcomes for at least six months after the focal investigation ($N = 20,462$), as this will be the primary outcome of interest. We next drop a relatively small number of cases with missing child zip code information ($N = 4,856$), since quasi-random assignment of investigators is conditional on a zip code by year fixed effect. Finally, we limit to investigators assigned to at least 50 high- and low-risk cases, defined below, to limit noise in estimates of investigator comparative advantage ($N = 50,386$).

The final sample consists of 322,758 investigations involving 261,021 children assigned to 908 investigators. 3.2% of investigations result in foster care placement. Table A1 presents summary statistics: 60% of children in our sample are white, 48% are female, 45% have had a CPS investigation prior to their focal one, and the average child is nearly seven years old (Panel A). Investigations in our sample tend to include at least one allegation of improper supervision (53%), physical neglect (44%), and physical abuse (29%). In 77% of investigations, at least one of the alleged perpetrators of maltreatment is the child's mother or step-mother (Panel B).

Panel C summarizes rates of "subsequent maltreatment" for children left at home following the focal investigation. Our primary maltreatment measure considers whether a child was re-investigated within six months of the focal investigation. This is a common proxy for subsequent maltreatment in the child welfare literature (e.g., Baron et al. (2024); Putnam-Hornstein et al. (2021)). Nevertheless, re-investigations are imperfect proxies for actual child maltreatment, as they only account for cases that are re-reported to CPS. While there are other potential proxies, such as a subsequent *substantiated* investigation, we prefer re-investigation because re-investigations within a few months may be assigned to the initial investigator who will again make substantiation decisions. In contrast, both the decision to re-report and to screen-in a case, the two steps required for a re-investigation, are outside of the initial investigator's control. Still, we show below that our findings are robust when using these alternative proxies for subsequent maltreatment. With these caveats in mind, we refer to a re-investigation within six months as "subsequent maltreatment" throughout the manuscript for ease of exposition. Note that this maltreatment outcome is mechanically missing for children placed in foster care, which is the primary identification challenge in this study. 16.4% of children experience subsequent maltreatment in the home within six months of the focal investigation.[29]

## IV.C Estimation strategy

A key question for the empirical simulation is how to partition cases. The theory allows for any binary partitioning of cases. In the context of CPS, this could mean categorizing cases as high- or low-risk, distinguishing between abuse and neglect, or separating cases based on the gender or age of the children involved. Given CPS's primary focus on preventing further child maltreatment and guided by discussions with local agencies in Michigan, we partition cases based on the predicted risk of subsequent maltreatment. We construct this measure by training a machine learning algorithm to predict the risk of subsequent maltreatment in the home, which we discuss further in Appendix E.1. We define high-risk cases as those in the top quartile of predicted algorithmic risk, and low-risk cases as all other cases.[30]

The results of Lemma 1 allow us to identify differences in social cost in low- and high-risk cases between investigators. Our estimation approach accounts for (*i*) over-fitting concerns and measurement error in investigator moment estimates, and (*ii*) the fact that investigators are quasi-randomly assigned only within offices. Define $\tilde{c}^k(j) := c^k(j) - c^k(j^0)$, where $j^0$ is the benchmark investigator used for all social cost comparisons.[31] Then $\tilde{c}^k(j)$ is identified and equal to $\tilde{c}^k(j) = \gamma_j^k - \gamma_{j^0}^k$.

To avoid concerns that our estimates of the benefits of reassignment are overstated due to over-fitting, we follow a split-sample strategy. Specifically, we randomize within the set of cases that each investigator was assigned into a "training" set (50%) and an "evaluation" set (50%). We use the training set to derive the optimal investigator assignment mechanism, and then test its effectiveness on the evaluation set.

We first estimate $j$'s performance scores across case types: $\gamma_j^l$ for low-risk cases and $\gamma_j^h$ for

[29] Moreover, while we maintain the assumption that $Y_i^*$ is binary throughout, our identification and theoretical results can readily accommodate richer partitions of $Y_i^*$ (see Appendix E.3).

[30] Predictive risk modeling and binary risk partitions are also used in recent CPS policy efforts. For example, the Los Angeles County Risk Stratified Supervision Model uses ML techniques to notify supervisors when a new case, classified by the model as "complex-risk," has been assigned to their office, so that they can devote additional time and attention to these cases. The binary partition in these settings has been justified by the need for simplicity when explaining the practical implications of high and low risk to supervisors and investigators.

[31] The results of Section IV.A apply if $j$ and $j^0$ are in the same office-by-year. However, under an additional linearity assumption introduced below, we can use a single reference investigator across all offices. The choice of $j^0$ has no impact on the mechanism results. We choose $j^0$ as the investigator with greatest caseload over our sample.

high-risk cases. This requires investigator-specific estimates of placement and false negative rates. Let $D_i = \sum_i D_{ij} Z_{ij}$ and $\text{FN}_i = \sum_i \text{FN}_{ij} Z_{ij}$, so that $D_i$ is an indicator for whether case $i$ resulted in placement, and $\text{FN}_i$ an indicator for whether the case is a false negative.

Investigators in Michigan are rotationally assigned to cases within CPS offices. Typically, each county in the state has its own office, but some large counties have multiple offices, and many offices split investigators into geographic-based teams (Baron and Gross, 2022). As such, we define an "office" throughout based on the child's zip code. Our identification results apply separately to each office. To compare investigators across offices, we use a linear adjustment to estimate investigator placement and false negative rates.[32] That is, we estimate regressions of the form:

$$D_i = \sum_j \beta^D_{j1} Z_{ij} + \beta^D_{j2} \text{ High-Risk}_i Z_{ij} + \mathbf{X}_i' \alpha^D + u_i \tag{6}$$

$$\text{FN}_i = \sum_j \beta^{\text{FN}}_{j1} Z_{ij} + \beta^{\text{FN}}_{j2} \text{ High-Risk}_i Z_{ij} + \mathbf{X}_i' \alpha^{\text{FN}} + \epsilon_i \tag{7}$$

where $\text{High-Risk}_i$ is an indicator equal to one if case $i$ is high-risk. We estimate Equations 6 and 7 separately in the training and evaluation samples. $\mathbf{X}_i$ is a vector of office-by-year fixed effects to account for the level of randomization. We demean $\mathbf{X}_i$ so that $\beta_{j1}$ represents strata-adjusted investigator-specific estimates of each outcome for low-risk cases and $\beta_{j1} + \beta_{j2}$ represents strata-adjusted investigator estimates of each outcome for high-risk cases. We use our estimates of these parameters to estimate performance scores among low- and high-risk cases, separately in the training and evaluation dataset, as:

$$\widehat{\gamma^l_j} = c_{FP} \widehat{\beta^D_{j1}} + (c_{FN} + c_{FP} - c_{TP}) \, \widehat{\beta^{\text{FN}}_{j1}}$$
$$\widehat{\gamma^h_j} = c_{FP} \left( \widehat{\beta^D_{j1}} + \widehat{\beta^D_{j2}} \right) + (c_{FN} + c_{FP} - c_{TP}) \left( \widehat{\beta^{\text{FN}}_{j1}} + \widehat{\beta^{\text{FN}}_{j2}} \right)$$

Following Chan et al. (2022), we assume $c_{TP} = c_{TN} = 0$, so that the welfare measure is focused only on prediction mistakes. As mentioned above, the value of $c_{FN}, c_{FP}$ must ultimately be chosen by the agency. To bring our mechanism to data, we assume that $c_{FP} = 1$ and $c_{FN} = 0.25$, though we show below that our results are robust to this choice of parameter values. To motivate this choice, note that CPS investigators in our context place 3.2% of children but 16.4% of children face subsequent maltreatment when left at home. This mismatch may imply that CPS views $c_{FN} < c_{FP}$. Normalizing $c_{FP} = 1$ suggests that $c_{FN} \in (0, 1)$. For our benchmark estimates, the ratio between placement rates and subsequent maltreatment rates suggests that $c_{FN}$ is roughly 0.25. We explore

[32] As discussed in Arnold et al. (2022), this approach tractably incorporates the large number of zip code-by-year fixed effects, under an additional assumption that placement and false negative rates are linear in the zip code-by-year effects for each investigator and case type.

robustness to this assumption in Figure A1, where we show that the ranking of investigators is well-preserved if we instead assign, for example, $c_{FN} = 0.12$ or $c_{FN} = 0.5$. To reduce noise in the estimates of the investigator moments, we follow Arnold et al. (2022) and use empirical Bayes shrinkage estimates of $\widehat{\gamma_j^l}, \widehat{\gamma_j^h}$ to adjust for finite sample error. We then estimate $\tilde{c}^k(j)$ as $\widehat{\gamma_j^k} - \widehat{\gamma_{j0}^k}$ and $d_j$ as $\widehat{\gamma_j^l} - \widehat{\gamma_j^h}$.[33]

## IV.D Main empirical results

### IV.D.1 Motivating empirical facts

We begin by presenting empirical facts that motivate the use of our mechanism in this setting. We first show considerable variation in investigator performance and comparative advantage within CPS offices across high- and low-risk cases. Second, we show that assignment to high-risk cases is costly for investigators, as greater exposure to these cases leads to significantly higher turnover. Finally, we present evidence from a statewide survey of CPS investigators of substantial heterogeneity in preferences over high- and low-risk cases.

**Considerable variation in performance and comparative advantage:** Intuitively, gains from investigator reassignment in our proposed mechanism can only occur if there is heterogeneity in investigators' relative performance across cases and within offices. That is, it is not enough for investigators to differ in the level of their performance, they must also differ in their comparative advantage scores.

We find considerable variation in performance and comparative advantage. Table A2 estimates the relationship between performance and comparative advantage metrics ($\gamma_j$ and $d_j$) on the training dataset and prediction error rates in the evaluation dataset. Panel A shows that investigators with a one standard deviation $\gamma_j$ below the office mean achieve a 1.1pp [7.1%] reduction in false negatives and 1.6pp [70.4%] reduction in false positives. Panels B and C further regress prediction error rates on comparative advantage scores, $d_j$. Investigators with a one standard deviation greater comparative advantage score achieve 2.0pp [8.5%] lower false negative rates and 2.3pp [57.3%] lower false positive rates in high-risk cases, but 0.04pp [0.3%] *higher* false negative rates and 0.2pp [13.6%] *higher* false positive rates in low-risk cases. That is, investigators with greater comparative advantage in high-risk cases achieve lower prediction error rates in high-risk cases but higher error rates in low-risk cases—providing evidence of investigator specialization across case types within offices.

**High-risk cases are costly to investigators:** Under the status-quo rotational system, the composition of caseloads in expectation is equal across investigators within an office.

[33] We also estimate $\widehat{\gamma}_j$, the average performance of investigator $j$ across all cases, by following this same methodology but omitting the interaction terms in Equations 6 and 7.

In practice, however, there may be time periods in which some investigators receive larger numbers of high- or low-risk cases by random chance. In Table A3, we leverage this variation to examine whether greater exposure to high-risk cases leads to increased investigator turnover. To do so, we use survival analysis techniques to measure the effect of caseload risk composition on investigator career length. Column 1 shows that a one standard deviation increase in the mean predicted risk of an investigator's caseload increases turnover risk by 149%. Column 2 reports that being assigned to an above-median share of high-risk cases increases turnover risk by 54%. Thus, exposure to a greater share of high-risk cases leads to large increases in investigator turnover, suggesting that these cases are costlier to investigators.

**Heterogeneity in perceived costs of high-risk cases:** To complement the administrative data, we conducted a survey of Michigan CPS investigators. The survey probed how investigators perceive high- versus low-risk cases and elicited their marginal rates of substitution (MRS) between case types. For a detailed report of the responses, see Appendix E.2.

The survey provides direct evidence that investigators themselves perceive high-risk cases as disproportionately costly. Respondents rated high-risk cases as much more time-consuming (mean = 8.8 on a 0–10 scale), more emotionally taxing (8.6), and only modestly more satisfying (5.3) than low-risk cases. Open-ended responses reinforced these findings, with investigators frequently linking complex cases to burnout and turnover.

The survey also revealed significant heterogeneity in preferences. Investigators reported widely varying MRS's between high- and low-risk cases. On average, investigators were willing to trade one high-risk case for about 3.6 low-risk cases, with substantial heterogeneity across individuals (Figure A6). Importantly, investigators reported similar MRS's when asked about their coworkers' preferences—an average MRS of 3.7.

Altogether, these patterns motivate the potential for the SMD-TP mechanism to achieve welfare gains in this context. There is evidence in the data of significant comparative advantage across high- and low- risk cases within offices. However, investigators themselves recognize sharp tradeoffs between case types. Additional high-complexity cases are experienced as disproportionately costly, both in terms of job satisfaction and turnover, and these preferences are heterogeneous. This creates scope for welfare gains through reallocation that better matches investigators to cases, while accounting for their preferences.

### IV.D.2 The role of investigator type distributions

As is standard in the Bayesian mechanism-design literature, our theoretical analysis takes the preference distributions, $F_j$, as an input. However, our mechanism can in fact be implemented without prior knowledge of these distributions. As shown above, if $F_j$ is misspecified, the

SMD-TP mechanism retains its incentive properties and continues to respect the status quo. The downside of using incorrect type distributions is that the mechanism may not converge to the optimal outcome in the large market. In other words, there are potential social welfare gains to be realized by improving our understanding of type distributions. Fortunately, this information can be realistically obtained in practice, unlike knowledge of each investigator's realized type. For example, the MRS distribution from our survey can be used to inform an initial specification for $F_j$. We can then run the SMD-TP mechanism for an initial trial period, using this estimated distribution. The trial run can then generate further data on individual investigators' preferences, allowing refinement of estimates of their preference distributions.[34] Therefore, the SMD-TP mechanism can be implemented without prior knowledge of the investigators' preference distribution.

**Simulating welfare gains:** While it is feasible to eventually gather information about $F_j$, we would like to understand now whether the mechanism could generate welfare gains across a range of distributions. Our survey provides a natural baseline for specifying investigator type distributions. We use the distribution of MRS's reported by investigators as the baseline specification of $F_j$. To apply the results in Corollary 1, we estimate hazards from a kernel-smoothed empirical distribution from the survey and impose monotone hazards (Myerson regularity) by choosing the nondecreasing hazard sequence that minimizes the discrete sum of squared deviations in hazard rate. We refer to this as the "regularized own-type empirical distribution" and apply the same procedure to the coworker-reported responses (see Appendix E.2 for details).[35] In our baseline analysis, we assume a common type distribution across investigators; in Appendix E.5 we also consider an extension that allows for correlation between types and performance, in which we continue to see significant welfare gains.

While we do not claim that the survey responses reflect the "true" distribution of types, they provide a useful benchmark to illustrate the potential welfare gains achievable under the

[34] To avoid introducing additional agency problems by making the mechanism in future periods dependent on type reports in the trial period, we cannot use the type report of agent $j$ to learn about $j$'s own type distribution. In fact, we can only use information about $j$'s type to learn about the distribution for agents in other offices, whose assignments do not interact with those of $j$. Still, given the large number of investigators involved, this should be sufficient to generate significant learning. This also raises the question of how the trial period should be chosen, trading off the benefits of experimentation versus exploitation. Resolving this trade-off is beyond the scope of the current paper, but recent work in Nguyen et al. (2023) provides a potential path forward.

[35] The regularized empirical distribution may not satisfy strong regularity. As a result, the fully-optimal LMS mechanism could involve a four-price mechanism for some investigators (see Theorem 2 for details). Nonetheless, for small-market implementation we restrict attention to two-price mechanisms. As discussed above, in the small market these mechanisms have desirable incentive properties (OSP) and their simplicity means that they could plausibly be implemented in practice. Moreover, as we show below, two-price mechanisms generate sizable welfare gains.

SMD-TP mechanism in a data-driven way. To assess robustness, however, we also simulate outcomes under several alternative distributions that seem like natural candidates, such as uniform and truncated normal specifications. Across all of these cases, we consistently find welfare improvements relative to the status quo, demonstrating that the mechanism's potential gains do not depend on any single distributional assumption.

#### IV.D.3 Social welfare gains

Corollary 3 shows that the difference in social cost between assignments is identified using investigator performance scores and their caseload composition in the two mechanisms. Using this result, we report differences between the SMD-TP mechanism and a "status-quo counterfactual" which splits high- and low-risk cases equally across investigators within counties.[36]

Our strategy to estimate welfare gains accounts for (i) uncertainty in investigator type distributions and (ii) over-fitting concerns. Under an initial distributional assumption for $F_j$, we use a split-sample strategy that combines investigator performance measures with their $p_j$ draws to compute the assignment generated by the SMD-TP mechanism for that draw in the training set. We then calculate the realized welfare gains for the given type profile in the evaluation set. We summarize welfare gains for a given specification of the type distribution as the average welfare gain across 100 draws of types. Finally, we estimate standard errors, clustered by investigators, using a bootstrapping procedure that accounts for uncertainty in estimates of both investigator performance and their type draw.

Table 1 presents estimated welfare gains across different distributional assumptions. Each column compares outcomes when investigators are assigned to cases using the SMD-TP mechanism versus a counterfactual in which high- and low-risk cases are evenly split. For example, when $F_j$ is specified using the regularized own-type empirical distribution, we estimate a decline in social costs of $1,189$ [5.9%] driven by a reduction in both 755 false negatives [1.5%] and $1,017$ false positives [13.9%] (Column 1).[37] The mechanism also reduces the number of total placements by 262 [2.6%]. Results are nearly identical when $F_j$ is instead based on investigators' reports of their coworkers' preferences (Column 2).

To estimate the importance of investigator private information for welfare gains, in Column 3 we again assume that $F_j$ is distributed as in Column 1, but that the designer observes each investigator's type directly and implements the first-best assignment for each realized

[36]The equal split of cases is the expected assignment under the rotational system. Table A7 shows that we obtain similar welfare gains when we benchmark the SMD-TP mechanism to the observed status quo.

[37]Baseline false positive counts are unobserved. To express welfare changes in percent terms, we use extrapolation-based estimates of the false positive rate, which we describe in Appendix E.4.

Table 1: Gains from Investigator Reallocation

| | (1) Own Type | (2) Coworker Type | (3) Known Type, Own Type |
|---|---|---|---|
| Social Costs | -1,189.2*** | -1,111.2*** | -3,351.7*** |
| | (300.7) | (274.3) | (454.9) |
| | [-5.9%] | [-5.5%] | [-16.7%] |
| False Negatives | -754.9*** | -713.3*** | -2,209.0*** |
| | (219.5) | (213.4) | (372.9) |
| | [-1.5%] | [-1.4%] | [-4.3%] |
| False Positives | -1,017.0*** | -950.1*** | -2,865.3*** |
| | (239.7) | (226.8) | (385.6) |
| | [-13.9%] | [-13.0%] | [-39.2%] |
| Placements | -262.1** | -236.7** | -656.3*** |
| | (106.4) | (97.8) | (190.2) |
| | [-2.6%] | [-2.4%] | [-6.6%] |

**Notes.** This table reports the welfare gains derived from the SMD-TP mechanism. Each column corresponds to a different distributional assumption for $p_j$. Column 1 presents gains from the regularized own-type empirical distribution: we estimate hazards from a kernel-smoothed empirical distribution from the survey and impose monotone hazards (Myerson regularity) by choosing the nondecreasing hazard sequence that minimizes the discrete sum of squared deviations in hazard rate (see Appendix E.2 for details). Column 2 performs the same procedure but using investigator reports of their coworkers' preferences. Column 3 uses the same type distribution as in Column 1, but assumes that $p_j$ is known to the designer. We estimate welfare changes and robust standard errors (in parenthesis), clustered by investigator, using the approach described in the text. Percent effects relative to the baselines are presented in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

type profile. In this simulation, the welfare gains increase dramatically—social costs decline by 16.7%, driven by a false negative decline of 4.3% and false positive decline of 39%. The comparison with Column 1 shows that information rents are significant when types are unobserved. As discussed in Appendix B.2, over time the designer will be able to use the data generated by the mechanism to reduce uncertainty about investigators' preferences. Column 3 represents an upper bound on welfare gains as the designer learns about individual investigators' types.

**Robustness checks:** We conduct two exercises to assess the robustness of our main findings. First, Table A4 demonstrates that the welfare gains persist across a wide range of distributional assumptions.[38] The largest gains arise when $p_j$ follows a degenerate distribution. This gap highlights that a naive analysis which ignores investigators' private information, and the attendant information rents, would significantly overstate welfare gains. Second, Table A6 shows that our results are robust to alternative proxies for maltreatment. While we use a subsequent investigation within six months as our preferred measure, since it is external to

[38]Figure A2 further illustrates this point by showing reductions in social costs under truncated normal distributions with varying means $\mu$ and standard deviations $\sigma$. The mechanism consistently achieves welfare improvements relative to the status quo.

the local CPS office, we obtain similar results when contracting the time horizon or when restricting to subsequent *substantiated* investigations.[39]

### IV.D.4 Investigator preferences

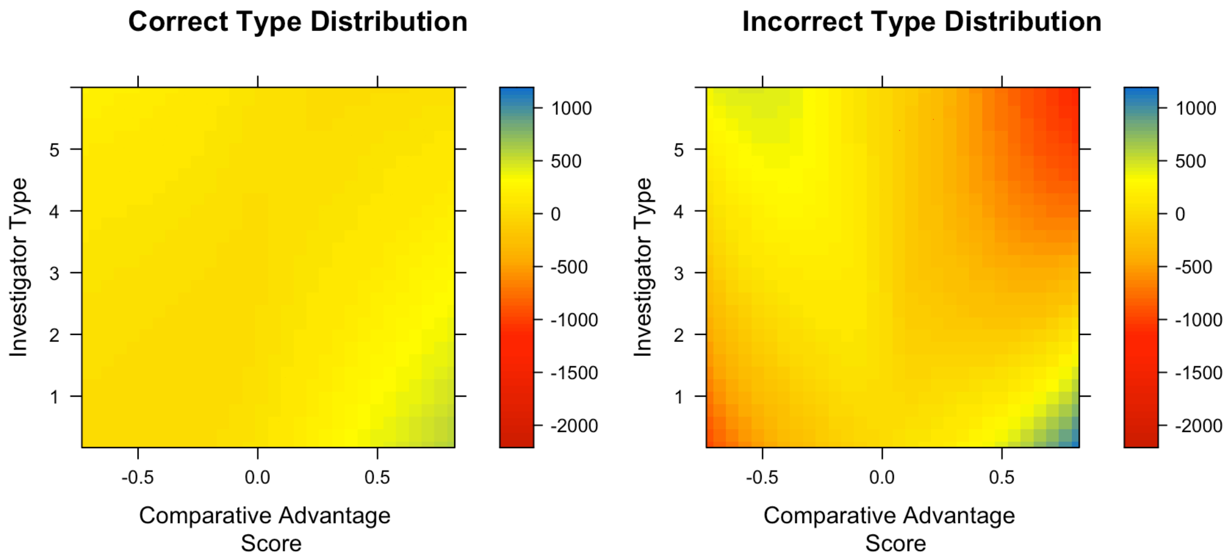
Figure 3 demonstrates the importance of considering investigators' heterogeneous preferences in the SMD-TP mechanism. We define investigator welfare for an investigator with type $p_j$ and caseload $(n^h, n^l)$ as $-(p_j n^h + n^l)$, the negative of their price-weighted caseload. In the left panel of Figure 3, we derive the optimal allocation of cases assuming that the true distribution of types is the regularized own-type empirical distribution. We then compute the difference between investigator welfare under the SMD-TP mechanism relative to the status quo. Under the SMD-TP mechanism, which accounts for investigator type heterogeneity, investigator welfare is improved by approximately 51 price-weighted cases, on average. Average investigator welfare is $-562$ in the equal-split counterfactual, so this represents a modest welfare improvement. Importantly, the reassignment makes no investigator substantively worse off: no investigator experiences a welfare loss greater than 10 price-weighted cases, and the 1st percentile of investigator welfare change is a loss of 3.4 cases.[40] In fact, 36% of investigators experience welfare gains under the correct SMD-TP mechanism of greater than 10 price-weighted cases and 26% of investigators experience welfare gains of at least 10%, which could in turn improve recruitment and retention.

The right panel of Figure 3 instead assigns cases without considering heterogeneity in investigator preferences. Formally, the mechanism assigns cases assuming that $p_j = 1$ for all investigators. But, when computing investigator welfare, we assume that their types truly follow the regularized own-type empirical distribution. Under this scenario, 27% of investigators experience welfare losses of at least 10%. Moreover, Figure 3 shows that there is significant heterogeneity in investigator welfare loss by comparative advantage and type. The investigators experiencing the largest losses are those with a large comparative advantage on high-risk cases as well as high $p_j$—investigators above the median in both their comparative advantage score, $d_j$, and $p_j$ experience an average welfare loss of 214 price-weighted cases (38%), and those in the top quartile of both experience an average welfare loss of 340 (60%). On the other hand, investigators with low comparative advantage on high-risk cases and high $p_j$ are made better-off under the mechanism that ignores preference heterogeneity.

[39] For computational purposes, we compare the welfare gains across different proxies for subsequent maltreatment in the LMS-TP mechanism, for which the SMD-TP is an approximation.

[40] The constraint that no investigator is made worse off by the mechanism is imposed exactly in the static model. The SMD-TP mechanism approximates the SMS-TP mechanism, and this approximation improves as the time horizon grows. Thus, in the dynamic version, some investigators can be made slightly worse off than the equal-split counterfactual.

Figure 3: The Importance of Accounting for Investigator Preferences




**Notes.** This figure plots the effect of reassignment according to the SMD-TP mechanism on investigators' welfare by their comparative advantage score, $d_j$ and their type, $p_j$. Investigator welfare for an investigator type $p_j$ and assigned to caseload $(n^h, n^l)$ is $-(p_j n^h + n^l)$. We report the difference between investigator welfare under the SMD-TP mechanism and a counterfactual in which cases are equally-split within counties. The left panel assumes that the true distribution of investigator types is the regularized own-type empirical distribution. The right panel calculates changes in investigator welfare under an assignment that assumes $p_j = 1 \ \forall j \in \mathcal{J}$, but where the true $p_j$ is distributed according to the regularized own-type empirical distribution. We present results averaged across the 100 investigator-type draws.

Figure 3 highlights why considering investigator preferences in the assignment problem is paramount. If the mechanism ignores types, investigators with large comparative advantage in high-risk cases receive more of these cases, but are made substantially worse off if assignment to high-risk cases is costly relative to low-risk cases. This would likely create greater turnover or worsened performance among such investigators, a particularly negative outcome in a system that already suffers from staff shortages.

### IV.D.5 Dynamic nature of the mechanism

Figure 3 considers investigators' welfare over their cumulative caseloads, but does not consider how their caseloads are spread over time. While smoothing caseloads over time does not directly enter the mechanism-design problem as a constraint, the SMD-TP solution attempts to do so by allocating cases based on the percent of the target level for each case type that each investigator has completed thus far. Thus at any point in time, each investigator should have completed approximately the same percentage of their high- or low-risk cumulative caseload. Figure A3 describes how cumulative price-weighted workloads vary over time. This figure shows that the SMD-TP mechanism is successful in spreading caseloads.

#### IV.D.6 Welfare comparisons across mechanisms

For reference, we also estimate welfare gains under the SMS-TP and LMS-TP mechanisms and compare these to outcomes under the SMD-TP mechanism in Table A5. Columns 1–3 report results when $p_j$ is drawn from the regularized own-type empirical distribution, while Columns 4–6 repeat the exercise assuming $p_j \sim \text{Unif}[1, 2]$. Comparing Columns 1 and 2, we find that moving from LMS-TP to SMS-TP significantly reduces welfare gains. The fact that LMS yields significantly larger gains than SMS in the baseline specification suggests that the benefits of reassignment may be greater in larger offices. Indeed, in smaller offices welfare gains are naturally limited regardless of the mechanism, since cases must ultimately be handled as they arrive and each investigator's workload cannot be too skewed over time. Consistent with this, Table A8, shows that focusing on the five largest counties in Michigan produces considerably higher welfare gains relative to our baseline estimates in Table 1. Finally, differences between SMS-TP and SMD-TP are small, indicating that the "online" nature of case assignment—the need to allocate cases without knowing which cases will arrive in the future—is not a first-order concern.

Finally, we investigate the potential gains from going beyond the binary-classification framework. Recall that our main binary split considered high-risk cases to be those in the top quartile of the predicted risk distribution, and low-risk as all other cases. Suppose instead that we allow the mechanism to condition on each of the four quartiles of this distribution. In general, solving for the optimal mechanism with four categories is beyond the scope of the current study. However, we can get a sense of the potential gains by considering the special case in which the designer observes investigators' preferences, in which case the SMS allocation problem reduces to a simple linear program. Table A9 reports this comparison. We find that the welfare gains from considering quartile mechanisms only increase social welfare gains from $1,506$ to $1,711$. This provides some evidence that the gains from considering more complicated mechanisms may be limited in this setting.

# V Conclusion

The ultimate objective of this work is a tractable framework for reforming institutions which allocate tasks among agents. Our main contribution is a mechanism-design analysis of this problem. Moreover, we have demonstrated empirically that the simple mechanism that we propose has the potential to significantly improve outcomes in the CPS context. The primary takeaway of this empirical analysis is that there are gains to be had, and further study is warranted on the use of performance data in assigning CPS investigators to cases. With the goal of providing a starting point for such reform (in this and other contexts) the theoretical analysis sought to address what we view as the primary challenges that a reforming mechanism must overcome:

1. *Unobservable agent preferences and status-quo constraints.* We wish to avoid negative impacts on agents' welfare. This concern may be motivated by recruitment and turnover considerations, as well as the need to convince stakeholders to adopt the proposed mechanism. Moreover, failure to impose the status-quo constraint may bias the empirical estimates regarding the value of performance data. Careful design of the mechanism is needed to deal with the fact that preferences are unobserved.

2. *Effort incentives.* While we do not explicitly model the decision to exert effort, we showed that under our mechanism agents' payoffs are improving in their performance, at least locally (Theorem 4). Thus, if the mechanism is implemented in successive periods and data from past performance is used to inform future assignments, the mechanism should provide agents with motivation to perform well.

3. *Perceived fairness of the mechanism.* Within our mechanism, it is possible for agents with the same preferences to receive different allocations. A concern is how agents will react to this disparity (even if every agent is better-off relative to the current system). Fortunately, we showed that disparate caseloads can be justified on the basis of performance: agents who receive fewer type-$k$ tasks for the same type-report are those who perform better on type-$k$ tasks (Theorem 4).

4. *Beyond binary case classifications.* We focused on mechanisms which condition assignments on a binary partition of cases. We emphasize that this is a restriction on the mechanism, not an assumption about the setting: the choice of how to partition the set of cases is itself a design choice. We discuss how the mechanism-design results can be extended to richer partitions (Appendix B.3). Moreover, our main identification results in Section IV.A do not depend on the binary partition assumption.

Before implementing the mechanism in the field, several practical considerations would need to be carefully addressed. A key challenge is effectively communicating the mechanism to agents and establishing a clear protocol for reporting their preferences. In the direct implementation of the mechanism, agents only need to report a single number—their MRS between high- and low-type cases. However, they would likely require guidance on how to interpret and understand this parameter (Budish and Kessler, 2022). Once types are elicited, the SMD-TP mechanism can operate with no further input from agents and requires minimal changes to current office procedures.

# A Proofs of main results

## A.1 Proof of Theorem 2 and Corollary 1

*Proof.* We begin, as in Myerson (1981), by using the envelope condition to simplify the IC constraints. First, note that in any IC mechanism $H$ must be non-increasing. Also, by the envelope theorem (Milgrom and Segal, 2002)

$$-pH(p) - L(p) = -\underline{p}H(\underline{p}) - L(\underline{p}) - \int_{\underline{p}}^{p} H(z)dz$$

in any IC mechanism. Moreover, if $H$ is non-increasing and $H, L$ satisfy the envelope condition, then the mechanism is IC. From the envelope condition and monotonicity of $H$, we then have that $L$ is non-decreasing. Thus non-negativity of $L(\underline{p})$ is sufficient for non-negativity of $L$. Note also that

$$\begin{aligned}\int L(p)dF(p) &= \underline{p}H(\underline{p}) + L(\underline{p}) - \int_{\underline{p}}^{\bar{p}} \left(pH(p) - \int_{\underline{p}}^{p} H(z)dz\right) dF(p) \\ &= \underline{p}H(\underline{p}) + L(\underline{p}) - \int_{\underline{p}}^{\bar{p}} H(p) \left(p - \frac{1-F(p)}{f(p)}\right) dF(p)\end{aligned}$$

We can use the above IC characterization to simplify the SQ constraint to

$$\underline{p}n^h + n^l - \underline{p}H(\underline{p}) - L(\underline{p}) - \sup_p \left\{\int_{\underline{p}}^{p} (H(z) - n^h)dz\right\} \geq 0.$$

Let $\phi(p) := p - \frac{1-F(p)}{f(p)}$ be the *virtual type* of $p$. Putting together our previous observations, the program defining the support function $S(a, b)$ becomes

$$S(a,b) = \max_{H,L} \; b\left(\underline{p}H(\underline{p}) + L(\underline{p})\right) + \int_{\underline{p}}^{\bar{p}} H(p)\left(a - b\phi(p)\right) dF(p) \tag{8}$$

$$s.t \qquad H \text{ is non-increasing} \qquad \text{(IC')}$$

$$\underline{p}n^h + n^l - \underline{p}H(\underline{p}) - L(\underline{p}) - \sup_p \left\{\int_{\underline{p}}^{p} (H(z) - n^h)dz\right\} \geq 0 \qquad \text{(SQ')}$$

$$H(p) \geq 0, \qquad L(p) \geq 0 \qquad \forall\, p \in [\underline{p}, \bar{p}]$$

By inspection of the program in eq. (8), it is optimal to choose $L(\underline{p})$ so that the (SQ') constraint binds. Then the program becomes

$$\max_{H \geq 0} \; b\left(\underline{p}n^h + n^l - \sup_p \left\{\int_{\underline{p}}^{p} (H(z) - n^h)dz\right\}\right) + \int_{\underline{p}}^{\bar{p}} H(p)\left(a - b\phi(p)\right) dF(p)$$

$$s.t \qquad H \text{ is non-increasing} \qquad \text{(IC')}$$

$$\underline{p}n^h + n^l - \underline{p}H(\underline{p}) - \sup_p \left\{ \int_{\underline{p}}^{p} (H(z) - n^h)dz \right\} \geq 0 \qquad \text{(non-negative } L\text{)}$$

Now notice that since $H$ is non-increasing, $\sup_p \left\{ \int_{\underline{p}}^{p} (H(z) - n^h)dz \right\} = \int_{\underline{p}}^{p^*} (H(z) - n^h)dz$, for any $\sup\{p : H(z) > n^h\} \leq p^* \leq \inf\{p : H(z) < n^h\}$. So we can solve the above program in two steps. First, for any fixed $\underline{p} \leq p^* \leq \bar{p}$ we solve

$$\max_{H \geq 0} \; b\left(\underline{p}n^h + n^l - \int_{\underline{p}}^{p^*} (H(z) - n^h)dz\right) + \int_{\underline{p}}^{\bar{p}} H(p)\,(a - b\phi(p))dF(p)$$

$$s.t \qquad H \text{ is non-increasing} \qquad \text{(IC')}$$

$$\underline{p}n^h + n^l - \underline{p}H(\underline{p}) - \int_{\underline{p}}^{p^*} (H(z) - n^h)dz \geq 0 \qquad \text{(non-neg. } L\text{)}$$

$$H(p) \geq n^h \quad \forall\, p \in [\underline{p}, p^*] \quad , \quad H(p) \leq n^h \quad \forall\, p \in [p^*, \bar{p}]$$

then we can optimize over $p^*$. We can solve this program separately for $H$ on $[\underline{p}, p^*]$ and $H$ on $[p^*, \bar{p}]$. First, fix $H$ on $[\underline{p}, p^*]$. Then we choose $H$ on $[p^*, \bar{p}]$ to solve

$$\max_{H \geq 0} \int_{p^*}^{\bar{p}} H(p)\,(a - b\phi(p))dF(p)$$

$$s.t \qquad H \text{ is non-increasing} \qquad \text{(IC')}$$

$$H(p) \leq n^h \quad \forall\, p \in [p^*, \bar{p}]$$

This looks exactly like a standard monopoly pricing problem. The extreme points of the set of feasible functions are step functions taking values in $\{0, n^h\}$. Since the objective is linear, there are always solutions in this set.[41]

Now consider the other half of the problem, choosing $H$ on $[\underline{p}, p^*]$. Rearranging the non-negative $L$ constraint, we have

$$\max_{H} \; b\left(p^* n^h + n^l\right) + \int_{\underline{p}}^{p^*} H(p)\,(a - b - b\phi(p))dF(p)$$

$$s.t \qquad H \text{ is non-increasing} \qquad \text{(IC')}$$

$$p^* n^h + n^l - \underline{p}H(\underline{p}) \geq \int_{\underline{p}}^{p^*} H(z)dz \qquad \text{(non-neg. } L\text{)}$$

$$H(p) \geq n^h \quad \forall\, p \in [\underline{p}, p^*]$$

[41]There may also be solutions which take intermediate values. For the problem of of maximizing $\hat{n}^l$ subject to a minimum requirement on $\hat{n}^h$, it may be necessary to use functions which takes values in $\{0, x, n^h\}$ for some $x \in (0, n^h)$. This would imply a four-price mechanism.

Fix $H(\underline{p}) > n^h$. The standard ironing argument implies that the optimal mechanism takes at most three values in $\{n^h, x, H(\underline{p})\}$ for some $x \in (n^h, H(\underline{p}))$. To be precise, if we ignore the non-negative L constraint then the extreme points of the feasible set are step functions taking values in $\{n^h, H(\underline{p})\}$, and to satisfy the non-negative L constraint we need to take a mixture between at most two such functions. It takes on only values in $\{n^h, H(\underline{p})\}$ if $\phi$ is strictly increasing and $b > 0$.

Consider now the choice of $H(\underline{p})$. If the non-negative $L$ constraint is slack, it is optimal to increase the value of $H(\underline{p})$ since doing so relaxes the monotonicity constraint (IC'). More explicitly, by the ironing argument we know that whenever the optimal mechanism given a fixed $H(\underline{p})$ takes three values, it must be that the non-negative $L$ constraint binds. Thus (given the fixed $H(\underline{p})$) the non-negative $L$ constraint is slack if and only if it is satisfied when we maximize over simple step functions, which means

$$(H(\underline{p}) - n^h) \min \left\{ \underset{z \in [\underline{p}, p^*]}{\arg\max} \left\{ \int_{\underline{p}}^{z} (a - b - b\phi(p)) dF(p) \right\} \right\} < n^l$$

However if this holds then it would be optimal to increase $H(\underline{p})$. Thus the non-negative $L$ constraint always binds (meaning $L(\underline{p}) = 0$) under the optimal mechanism.

Combining the solutions above and below $p^*$ yields the general solution described in Theorem 2. When $\phi$ is strictly increasing, we have that any optimal mechanism must use only two prices. Moreover, since the mixture of any two distinct two-price mechanisms is not itself a two-price mechanism, the solution must be unique. □

## A.2 Proof of Corollary 2

*Proof.* We want to show that under strong regularity, the frontier of $\mathcal{F}$ is equal to $\{N^*(a, b) : (a, b) \geq\}$, i.e. it suffices to consider only non-negative weights. To prove this we show that if $a = 0$, $b > 0$ then $N^*(a, b) = (0, x)$ for $x > 0$, and if $a >$, $b = 0$ then $N^*(a, b) = (y, 0)$ for $y > 0$. This implies that the entire frontier is downward sloping, and is traced out by considering $(a, b) \geq 0$.

Consider first $a = 0, b > 0$. Then under strong regularity, $(a - b\phi(p))$ and $(a - b - b\phi(p)$ are negative for all $p$. Thus from the proof of Theorem 2, the optimal mechanism sets

$$H(p) = \begin{cases} n^h & \text{on } [\underline{p}, p^*] \\ 0 & \text{on } (p^*, \bar{p}] \end{cases}$$

for some $p^* \in [\underline{p}, \bar{p}]$. The value obtained is

$$b(\underline{p}n^h + n^l) - bn^h \int_{\underline{p}}^{p^*} \phi(p) dF(p),$$

so it is optimal to set $p^* = \underline{p}$.

Suppose instead that $a > 0, b = 0$. Then from the proof of Theorem 2 it is optimal to set

$$H(p) = \begin{cases} n^h + \frac{1}{p^*} n^l & \text{on } [\underline{p}, p^*] \\ n^h & \text{on } (p^*, \bar{p}] \end{cases}$$

for some $p^* \in [\underline{p}, \bar{p}]$. The value obtained is

$$a\left(n^h + \frac{1}{p^*} n^l\right) F(p^*) + an^h(1 - F(p^*)) = an^h + an^l \frac{F(p^*)}{p^*}$$

which is maximized at $p^* = \bar{p}$ under strong regularity.

We have shown that it suffices to consider $(a, b) \geq 0$. Then if $F$ is strictly regular, Theorem 2 tells us that for any point $(\hat{n}^h, \hat{n}^l)$ on the frontier, the only way to implement $(\hat{n}^h, \hat{n}^l)$ is with a two-price mechanism. Suppose there is a linear segment of the frontier which contains distinct points $(\hat{n}_1^h, \hat{n}_1^l)$ and $(\hat{n}_2^h, \hat{n}_2^l)$. Then the mixture $\alpha(\hat{n}_1^h, \hat{n}_1^l) + (1 - \alpha)(\hat{n}_2^h, \hat{n}_2^l)$ can be induced by the $\alpha$ mixture of the two-price mechanisms that induce $(\hat{n}_1^h, \hat{n}_1^l)$ and $(\hat{n}_2^h, \hat{n}_2^l)$. However since such a mixture is not itself a two-price mechanism, $\alpha(\hat{n}_1^h, \hat{n}_1^l) + (1 - \alpha)(\hat{n}_2^h, \hat{n}_2^l)$ cannot be on the frontier. □

## A.3 Proof of Theorem 3

We first verify that strong duality holds. We know $(n^h, n^l) \in \mathcal{F}_j$ for all $j$. If $(n^h, n^l)$ is, moreover, in the interior of $\mathcal{F}_j$ for all $j$ then Slater's condition is satisfied, so we are done. Otherwise, define a perturbed problem by letting $S_j^\varepsilon(a, b) = S_j(a, b)(1 + \varepsilon)$ for all $j$, and

$$\mathcal{F}_j^\varepsilon = \{(\hat{n}^h, \hat{n}^l) \in \mathbb{R}^2 \ : \ a\hat{n}^h + b\hat{n}^l \leq S_j^\varepsilon(a, b) \ \forall \ (a, b) \in \mathbb{R}^2\}$$

Then $(\hat{n}^h, \hat{n}^l)$ is in the interior of $\mathcal{F}_j^\varepsilon$ for all $j$, so Slater's condition is satisfied for any $\varepsilon > 0$. The value of the dual in the perturbed program is

$$\max_{\lambda_h, \lambda_l} \lambda_h J n^h + \lambda_l J n^l - (1 + \varepsilon) \sum_{j=1}^{J} S_j \left( \left(\lambda_h - c^h(j)\right), \left(\lambda_l - c^l(j)\right) \right)$$

which converges to the value of the original dual in eq. (5) as $\varepsilon \to 0$. Moreover $\varepsilon \mapsto \mathcal{F}_j^\varepsilon$ is upper- and lower-hemicontinuous; and compact and convex valued, so the value of the primal program is continuous in $\varepsilon$ by Berge's maximum theorem. Because strong duality holds for all $\varepsilon > 0$, we conclude that it holds in the limit.

The first part of the theorem, up to and including the claim that each investigator receives a caseload on the boundary of $\mathcal{F}_j$, is immediate from the dual formulation. Suppose now that no two investigators are identical, in the sense stated in the result. We first show that at most two investigators have non-zero allocations that are off of the frontier. Note that if $a, b < 0$ then $N_j^*(a,b) = \{(0,0)\}$, and $N_j^*(a,b)$ contains non-zero points that are off of the frontier if and only if either $a \leq b = 0$ or $b \leq a = 0$. If agents are not identical, for any $\lambda_h, \lambda_l$ there is at most one $j$ such that $\lambda_h - c^h(j) = 0$, and one $j'$ such that $\lambda_l - c^l(j') = 0$.

It remains to prove the stated implications of strong regularity. There are two cases to consider. First, suppose there exists an optimal mechanism such that some investigator receives a strictly positive quantity of both types of cases. Because $F_j$ admits a continuous density for all $j$, $S_j$ is strictly convex over the set of $(a,b)$ that such that $N_j^*(a,b)$ is on the interior of the frontier. Thus the solution to the dual, $(\lambda_h, \lambda_l)$ must be unique. Moreover, under strong regularity $N_j^*(\lambda_h - c^h(j), \lambda_l - c^l(j))$ is single-valued for all $j$ such that either $\lambda_h - c^h(j) \neq 0$ and/or $\lambda_l - c^l(j) \neq 0$. The solution for the remaining two agents is uniquely pinned down by market clearing.

Alternatively, suppose that there are no solutions such that some investigator receives a strictly positive quantity of both types of cases. Then in any solution there is a set $A \subset \mathcal{J}$ of investigators who receive no low-type cases, and a set $B \subset \mathcal{J}$ of investigators who receive no high-type cases. For each pair of sets $(A, B)$ there is a unique allocation of the cases (under the non-identical $c^k(j)$ assumption): among $A$ give as many cases as possible to the agents with lower $c^h(j)$, and similarly for $B$. Suppose that there two solutions in which these sets differ, say $(A, B)$ and $(A', B')$, such that $j \in A \cap B'$ and $j$ receives a non-zero allocation in both. Since the objective is linear, the half-half mixture of these two assignments must also be a solution. However in that case $j$ gets some of both types of cases, contradicting our initial assumption.

# References

Abdulkadiroğlu, Atila and Aram Grigoryan (2023) "Market Design for Distributional Objectives in Allocation Problems: An Axiomatic Approach," Working Paper.

Aggarwal, Gagan, Gagan Goel, Chinmay Karande, and Aranyak Mehta (2011) "Online vertex-weighted bipartite matching and single-bid budgeted allocations," in *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, 1253–1264, SIAM.

Akbarpour, Mohammad, Eric Budish, Piotr Dworczak, and Scott Duke Kominers (2024a) "An economic framework for vaccine prioritization," *The Quarterly Journal of Economics*, 139 (1), 359–417.

Akbarpour, Mohammad, Piotr Dworczak, and Scott Duke Kominers (2024b) "Redistributive allocation mechanisms," *Journal of Political Economy*, 132 (6), 1831–1875.

Angelova, Victoria, Will Dobbie, and Cyrstal Yang (2023) "Algorithmic Recommendations and Human Discretion," Working paper.

Arnold, David, Will S Dobbie, and Peter Hull (2022) "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 112 (9), 2992–3038.

Azevedo, Eduardo M and Eric Budish (2019) "Strategy-proofness in the large," *The Review of Economic Studies*, 86 (1), 81–116.

Baccara, Mariagiovanna, Allan Collard-Wexler, Leonardo Felli, and Leeat Yariv (2014) "Child-Adoption Matching: Preferences for Gender and Race," *American Economic Journal: Applied Economics*, 6 (3), 133–58.

Baccara, Mariagiovanna, SangMok Lee, and Leeat Yariv (2020) "Optimal dynamic matching," *Theoretical Economics*, 15 (3), 1221–1278.

Baker, George, Robert Gibbons, and Kevin J Murphy (2001) "Bringing the market inside the firm?" *American Economic Review*, 91 (2), 212–218.

Bald, Anthony, Joseph Doyle Jr., Max Gross, and Brian Jacob (2022) "Economics of Foster Care," *Journal of Economic Perspectives*, 36 (2), 223–246.

Baron, E. Jason, Joseph J. Doyle, Natalia Emanuel, Peter Hull, and Joseph Ryan (2024) "Discrimination in Multi-Phase Systems: Evidence from Child Protection," *The Quarterly Journal of Economics*, Forthcoming.

Baron, E Jason and Max Gross (2022) "Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators," NBER Working Paper 29922.

Biasi, Barbara, Chao Fu, and John Stromme (2025) "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage," *Working Paper*.

Border, Kim C (1991) "Implementation of reduced form auctions: A geometric approach," *Econometrica*, 1175–1187.

Budish, Eric (2011) "The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes," *Journal of Political Economy*, 119 (6), 1061–1103.

——— (2012) "Matching "versus" mechanism design," *ACM SIGecom Exchanges*, 11 (2), 4–15.

Budish, Eric, Gérard P Cachon, Judd B Kessler, and Abraham Othman (2017) "Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation," *Operations Research*, 65 (2), 314–336.

Budish, Eric and Judd B Kessler (2022) "Can market participants report their preferences accurately (enough)?" *Management Science*, 68 (2), 1107–1130.

Casey Family Programs (2023) "Turnover Costs and Retention Strategies," https://www.casey.org/turnover-costs-and-retention-strategies/, Accessed: 2024-08-27.

Chan, David C, Matthew Gentzkow, and Chuan Yu (2022) "Selection with Variation in Diagnostic Skill: Evidence from Radiologists," *The Quarterly Journal of Economics*, 137 (2), 729–783.

Che, Yeon-Koo, Jinwoo Kim, and Konrad Mierendorff (2013) "Generalized Reduced-Form Auctions: A Network-Flow Approach," *Econometrica*, 81 (6), 2487–2520.

Combe, Julien, Vladyslav Nora, and Olivier Tercieux (2021) "Dynamic assignment without money: Optimality of spot mechanisms," *Working Paper*.

Combe, Julien, Olivier Tercieux, and Camille Terrier (2022) "The design of teacher assignment: Theory and evidence," *The Review of Economic Studies*, 89 (6), 3154–3222.

Daskalakis, Constantinos, Alan Deckelbaum, and Christos Tzamos (2014) "The complexity of optimal mechanism design," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, 1302–1318, SIAM.

——— (2015) "Strong duality for a multiple-good monopolist," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 449–450.

Dinerstein, Michael and Troy D Smith (2021) "Quantifying the supply response of private schools to public policies," *American Economic Review*, 111 (10), 3376–3417.

Doyle, Joseph J, Natalia Emanuel, and Raimundo Eyzaguirre (2025) "Cumulative Risks of Foster Care Placement for US Children: Comparing Birth Cohort and Synthetic Cohort Approaches," *NBER Working Paper #w34069*.

Dworczak, Piotr, Scott Duke Kominers, and Mohammad Akbarpour (2021) "Redistribution through markets," *Econometrica*, 89 (4), 1665–1698.

Dworczak, Piotr and Ellen Muir (2023) "A mechanism-design approach to property rights," Working Paper.

Grossman, Sanford J and Oliver D Hart (1983) "An analysis of the principal-agent problem," *Econometrica*.

Hart, Sergiu and Noam Nisan (2019) "Selling multiple correlated goods: Revenue maximization and menu-size complexity," *Journal of Economic Theory*, 183, 991–1029.

Holmstrom, Bengt (1989) "Agency costs and innovation," *Journal of Economic Behavior & Organization*, 12 (3), 305–327.

Holmstrom, Bengt and Paul Milgrom (1987) "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica*, 303–328.

Hylland, Aanund and Richard Zeckhauser (1979) "The efficient allocation of individuals to positions," *Journal of Political economy*, 87 (2), 293–314.

Jullien, Bruno (2000) "Participation constraints in adverse selection models," *Journal of Economic Theory*, 93 (1), 1–47.

Karp, Richard M, Umesh V Vazirani, and Vijay V Vazirani (1990) "An optimal algorithm for on-line bipartite matching," in *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, 352–358.

Kim, Hyunil, Christopher Wildeman, Melissa Jonson-Reid, and Brett Drake (2017) "Lifetime Prevalence of Investigating Child Maltreatment Among US Children," *American Journal of Public Health*, 107 (2), 274–280.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018) "Human Decisions and Machine Predictions," *Quarterly Journal of Economics*, 133 (1), 237–293.

Lewis, Tracy R and David EM Sappington (1989) "Countervailing incentives in agency problems," *Journal of economic theory*, 49 (2), 294–313.

Li, Shengwu (2017) "Obviously strategy-proof mechanisms," *American Economic Review*, 107 (11), 3257–3287.

Loertscher, Simon and Ellen V Muir (2021) "Wage dispersion, minimum wages and involuntary unemployment: A mechanism design perspective," *Working Paper*.

MacDonald, Diana E. (2024) "Foster Care: A Dynamic Matching Approach," *Working Paper*.

Maggi, Giovanni and Andres Rodriguez-Clare (1995) "On countervailing incentives," *Journal of Economic Theory*, 66 (1), 238–263.

Mehta, Aranyak, Amin Saberi, Umesh Vazirani, and Vijay Vazirani (2007) "Adwords and generalized online matching," *Journal of the ACM (JACM)*, 54 (5), 22–es.

Milgrom, Paul and Ilya Segal (2002) "Envelope theorems for arbitrary choice sets," *Econometrica*, 70 (2), 583–601.

Myerson, Roger B (1981) "Optimal auction design," *Mathematics of operations research*, 6 (1), 58–73.

Nguyen, Thành, Alexander Teytelboym, and Shai Vardi (2023) "Dynamic Combinatorial Assignment," *Working Paper*.

Nguyen, Thành and Rakesh Vohra (2021) "Δ-Substitute Preferences and Equilibria with Indivisibilities," in *Proceedings of the 22nd ACM Conference on Economics and Computation*, 735–736.

Pavlov, Gregory (2011) "Optimal mechanism for selling two goods," *The BE Journal of Theoretical Economics*, 11 (1).

Prendergast, Canice (2017) "How food banks use markets to feed the poor," *Journal of Economic Perspectives*, 31 (4), 145–162.

——— (2022) "The allocation of food to food banks," *Journal of Political Economy*, 130 (8), 1993–2017.

Putnam-Hornstein, Emily, John Prindle, and Ivy Hammond (2021) "Engaging Families in Voluntary Prevention Services to Reduce Future Child Abuse and Neglect: A Randomized Controlled Trial," *Prevention Science*, 22 (7), 856–865.

Roberts, Donald John and Andrew Postlewaite (1976) "The incentives for price-taking behavior in large exchange economies," *Econometrica*, 115–127.

Robinson-Cortes, Alejandro (2019) "Who gets placed where and why? An empirical framework for foster care placement," *Working Paper*.

Slaugh, Vincent W, Mustafa Akan, Onur Kesten, and M Utku Ünver (2016) "The Pennsylvania Adoption Exchange improves its matching process," *Interfaces*, 46 (2), 133–153.

Sönmez, Tayfun (2023) "Minimalist market design: A framework for economists with policy aspirations," *arXiv preprint arXiv:2401.00307*.

Spence, Michael (1973) "Job Market Signaling," *The Quarterly Journal of Economics*.

Valenzuela-Stookey, Quitzé (2023) "Interim Allocations and Greedy Algorithms," Working Paper.

Varian, Hal R (1973) "Equity, envy, and efficiency," *Journal of Economic Theory*.

# Online Appendix

# B Extensions and additional properties

## B.1 Fairness and Incentives for Effort

**Definition.** A mechanism is ***locally effort-inducing*** for $j$ if the following holds: if $j$ reports their type truthfully and receives a type-$k$ case, then $j$'s payoff must be locally increasing in their performance on type-$k$ cases (i.e., decreasing in $c^k(j)$).

To understand this definition, consider a two-period model of mechanism design. In the first period, a mechanism is designed and implemented for the LMS problem, based on some initial estimates of $(c^h(j), c^l(j))_{j\in\mathcal{J}}$. In the second period, the outcomes from the first period are used update the performance estimates, and a new mechanism is designed and implemented. Suppose that agent $j$ expects the performance of other agents to remain unchanged from the first to the second period, but in the first period can choose to degrade their own performance when assigned a type-$k$ case, so as to increase $c^k(j)$. That the mechanism implemented in both periods is locally effort-inducing means precisely that $j$ cannot gain by such degradation (for small increases in $c^k(j)$), conditional on having truthfully reported their type in the first period. To be clear, this does not rule out the possibility that $j$ could profit from the double deviation of misreporting their type in period 1 and degrading their performance on the cases they are assigned. However, such deviations are costly, since $j$ must take on a less-preferred caseload in period 1 in order to potentially improve their assignment in period 2. We therefore view local effort-inducing as a real, albeit qualified, restriction on the potential gains from performance degradation.

A closely related condition concerns the fairness of a mechanism. Say that an agent $j$ with type $p_j$ *envies* agent $j'$ if $j'$ is not excluded (i.e., $j'$ receives some cases), and $j$ would prefer to be offered the mechanism $(H^{j'}, L^{j'})$ rather than $(H^j, L^j)$. We say that agent $j$'s envy is *justified* if, moreover, they have the same type distribution and either $H^j(p_j) > 0, L^j(p_j) = 0, c^h(j) < c^h(j')$, and $c^l(j) = c^l(j')$; or $H^j(p_j) = 0, L^j(p_j) > 0, c^l(j) < c^l(j')$, and $c^h(j) = c^h(j')$. To understand this definition, suppose $j$ and $j'$ are both asked to specialize in type-$h$ cases, but $j$ has justified envy for $j'$. This means that $j'$ handles fewer type-$h$ cases than $j$, despite the fact that $j$ performs better on these cases, and both perform the same on type-$l$

cases.[42] Such an outcome is arguably unfair to agent $j$.[43]

**Definition.** A mechanism is ***locally fair*** for agent $j$ with type $p_j$ if there exists $\epsilon > 0$ such that there is no $j'$ with $|c^h(j) - c^h(j')| + |c^l(j) - c^l(j')| < \varepsilon$ for which $j$ has justified envy.

**Theorem 4.** Assume $F_j$ is strongly regular and $F(p)/p$ is increasing.. Then, for each agent $j$, the optimal LMS-TP mechanism is locally fair for agent $j$, regardless of their type. Moreover, for all but at most two agents, the optimal LMS-TP mechanism is locally effort-inducing.

*Proof.* Proof in Appendix C.1. □

The only agents for whom the mechanism may not be locally effort-inducing are the remedial agents who are off the frontier. The result depends on restrictions on the type distributions. If fairness and effort concerns are important in practice, the designer can impose these conditions on the distribution. If the distributions are misspecified, the solutions to the SMS and SMD problems derived from the LMS-TP mechanism will be sub-optimal, but they remain feasible, IC, and IR.

*Remark* 5. Theorem 4 is stated for the LMS-TP mechanism, but these properties translate approximately to the SMS and SMD mechanisms described below. In fact, we can say a bit more: both the SMS and SMD mechanisms are based on taking the prices $(p_1^j, p_2^j)_{j=1}^J$ defined in the LMS-TP mechanism and using these prices to construct an allocation, and so Proposition 1 applies to these mechanisms as well.

A separate concern is that, although our mechanism is locally fair, if $j$ saw that $j'$ was receiving more favorable exchange rates for high-type cases, $j$ might be discouraged about their own performance on high-type cases. In general, however, the mapping from performance to exchange rates is difficult to invert, and so investigators are unlikely to be able to make detailed inferences about others' performance. For example, these exchange rates could also be consistent with investigator $j'$ performing poorly on low-type cases. How exactly to convey information about the mechanism to avoid discouraging agents is a question that would need to be addressed as part of a practical implementation of the mechanism.

[42] We require equal performance for cases that $j$ is not assigned to guarantee that $j$ and $j'$ are roughly comparable agents. Strict equality is not important, it would suffice for their performance on these cases to be similar. The reason we need the agents to be similar has to do with comparative advantage. Suppose $j$ is assigned only type-$h$ cases. If $j'$ performs worse than $j$ for both case types, but is significantly worse for the type-$l$ cases, then $j'$ may still have a comparative advantage for type-$h$ cases. Thus, it would be justifiable for the mechanism to match $j'$ with no type-$l$ cases and still give $j'$ relatively few type-$h$ cases.

[43] The definition of justified envy does not cover the case where $j$ retains the status quo. This is because the status quo is the same for all investigators, so $j$ cannot have justified envy for another investigator who also retains the status quo. This is the only relevant comparison for local fairness, which is our focus here.

## B.2 Learning about $c^h(j), c^l(j)$

In Section IV.A, we leverage the quasi-random nature of the observed assignment to identify the cost parameters $c^k(j)$. A natural concern is that if one were to implement the SMD-TP mechanism, we would lose the ability to continue to learn about the performance, $c^h(j)$ and $c^l(j)$, of investigators. Fortunately, what matters for identification is that the assignment be quasi-random *conditional on case type*, which the SMD-TP mechanism is. The only remaining challenge to continued learning about investigator performance is that the SMD-TP assignment may violate the full-support condition of Lemma 1. In other words, if investigator $j$ never receives any type-$k$ cases, then we cannot hope to learn about $c^k(j)$.

A simple way to solve this problem is to introduce some additional randomness into the mechanism, so that every agent receives at least some of each type of case. This is also how the proposed mechanism can accommodate the arrival of new investigators: by keeping them on the status-quo "track" until we have enough information to estimate their performance parameters. In essence, we face the familiar experimentation-exploitation trade-off (Weitzman, 1978; Bolton and Harris, 1999). A more sophisticated solution would involve explicitly modeling this trade-off as part of the mechanism design problem, as in Kasy and Teytelboym (2023).

## B.3 More than two case types

Thus far, we have maintained the assumption that cases are partitioned into two types. It is worth reiterating that this is a restriction on the mechanism, not an assumption about the setting: the binary-type restriction imposes that assignments are random conditional on case type, but this this does not mean that cases with the same type must be identical.

The mechanism designer here has the freedom to choose the partition of cases that is used by the mechanism. In theory, we could choose any finite partition of the cases as a function of observable characteristics, provided the partition satisfies the identification conditions in Lemma 1 and Corollary 3. The challenge when moving beyond the binary partition setting is that it becomes difficult to characterize the optimal mechanism. With only two types of cases we were able to reduce the investigator's type to a one-dimensional variable. With more than two types of cases this is no longer possible. Mechanism design with multi-dimensional types and allocations is in general significantly more challenging than the one dimensional case, and even simple instances of this problem remain unsolved (see for example Hart and Reny (2015)).

Given this difficulty, there are two options available if we allow for non-binary partitions.

First, we could look for computational solutions to the optimal mechanism within a restricted class of "pricing mechanisms" which nests the LMS-TP mechanism as a special case. Just like in the two-price mechanism, the idea would be to endow each investigator the status quo assignment and then allow them to "buy and sell cases" according to some (potentially non-linear) price schedule. While such a mechanism is likely sub-optimal in the space of all mechanisms, it would at least improve on the binary-partition specification.

A second option would be to allow for non-binary partitions of cases, but impose additional restrictions to allow us to characterize the optimal mechanism. One simple case would be to assume that we can partition cases in a way that is orthogonal to investigators' preferences. For example, suppose that in addition to being high- or low- risk, cases are either "left" or "right." If investigators care about whether a case is high- or low- risk, but not whether it is left or right, then the characterization of the optimal mechanism remains essentially unchanged. The only difference is that rather than each investigator getting an assignment which is random given risk type, we can now match left- and right-type cases with investigators according to their relative performance. Assuming that this dimension is indeed orthogonal to investigators' preferences, this yields a lower social cost to the designer without affecting investigators' payoffs. More generally, if we can restrict investigators' preferences to be one-dimensional given the partition of cases, it should be possible to characterize the optimal mechanism using techniques similar to those employed above.

The downside of both of these options, especially the computational approach, is that we lose some of the simplicity of the mechanism. Simplicity is not only useful for practical implementation purposes; it also allows us to establish theoretical properties of the mechanism, such as effort incentives (Theorem 4). Nonetheless, generalizations beyond binary partitions, particularly by pursuing the second approach above, are an interesting direction for future work.

## B.4 Finite-sample adjustments to SMD-TP mechanism

To move between the two extremes of assigning based on the difference between realized and target caseloads, versus assigning based on the ratio, we can modify the algorithm by adjusting the score as follows for some $\varepsilon > 0$

$$\tilde{r}_j(t,k) = \frac{\hat{n}_j^k(t) + \varepsilon}{\dot{n}_j^k + \varepsilon}.$$

For large $\varepsilon$ the assignments generated by using the ratio $\tilde{r}$ converge to those generated by using the difference $\hat{n}_j^k(t) - \dot{n}_j^k$.

**Lemma 2.** For any $\hat{n}_j^k, \hat{n}_m^k$ and $\dot{n}_j^k, \dot{n}_m^k$, there exists $x$ large enough such that

$$\frac{\hat{n}_j^k + \varepsilon}{\dot{n}_j^k + \varepsilon} < \frac{\hat{n}_m^k + \varepsilon}{\dot{n}_m^k + \varepsilon} \Leftrightarrow \dot{n}_j^k - \hat{n}_j^k > \dot{n}_m^k - \hat{n}_m^k$$

for all $\varepsilon > x$.

*Proof.*

$$\begin{aligned}
\frac{\hat{n}_j^k + \varepsilon}{\dot{n}_j^k + \varepsilon} < \frac{\hat{n}_m^k + \varepsilon}{\dot{n}_m^k + \varepsilon} &\Leftrightarrow (\hat{n}_j^k + \varepsilon)(\dot{n}_m^k + \varepsilon) < (\hat{n}_m^k + \varepsilon)(\dot{n}_j^k + \varepsilon) \\
&\Leftrightarrow \varepsilon(\hat{n}_j^k + \dot{n}_m^k) + \hat{n}_j^k \dot{n}_m^k < \varepsilon(\hat{n}_m^k + \dot{n}_j^k) + \hat{n}_m^k \dot{n}_j^k \\
&\Leftrightarrow \hat{n}_m^k \dot{n}_j^k + \varepsilon(\dot{n}_j^k - \hat{n}_j^k) > \hat{n}_j^k \dot{n}_m^k + \varepsilon(\dot{n}_j^m - \hat{n}_j^m).
\end{aligned}$$

Taking $\varepsilon$ large yields the result. □

Thus, by adjusting $\varepsilon$ we can smoothly move between the two extremes of assigning based on ratios and assigning based on differences. More generally, in finite samples we can balance the desire to smooth investigators caseloads over time on the one hand, accomplished by assigning based on the ratio, versus ensuring that the difference between target and realized caseloads is small, by using a generalized scoring rule of the form

$$\tilde{r}_j(t,k) = \frac{\hat{n}_j^k(t) + x(t)}{\dot{n}_j^k + x(t)}.$$

for some increasing function $f > 0$. The asymptotic properties of the SMD-TP mechanism are preserved, but it may be possible to adjust $f$ to improve finite sample performance. We leave this as a topic for future work.

# C Omitted proofs

## C.1 Proof of Theorem 4

*Proof.* We begin with some preliminary comparative statics observations.

**Lemma 3.** If $c^k(j)$ increases (fixing $c^{-k}(j)$) then in expectation $j$ receives fewer type-$k$ cases in the optimal LMS-TP mechanism (where the expectation is taken over $p_j$). Similarly, if $c^k(j) > c^k(j')$, $c^{-k}(j) = c^{-k}(j')$, and $F_j = F_{j'}$ then $j$ receives fewer type-$k$ cases than $j'$ in expectation.

*Proof.* The first case is easily seen by observing that the objective function in the program in eq. (3) has increasing differences in $c^k(j)$ and $\hat{n}_j^k$. The second case is immediate from eq. (5) and the definition of $S^j$. □

Given Lemma 3, the remaining question is how changes in the optimal expected caseloads translate into changes in the prices offered to each investigator.

Consider now the claim about local fairness. If $j$ is remedial then any agent with worse performance is excluded, so $j$ cannot have justified envy. Assume therefore that $j$ is on their frontier. The prices $p_1^j, p_2^j$ defining the optimal LMS-TP mechanism solve

$$\max_{\underline{p}_j \leq p_1 \leq p_2 \leq \bar{p}^j} (\lambda_h - c^h(j)) \left( F_j(p_2)n^h + \frac{F_j(p_1)}{p_1} n^l \right) + (\lambda_l - c^l(j)) \left( (1 - F_j(p_1))n^l + (1 - F_j(p_2))p_2 n^h \right).$$

Fixing $\lambda_h, \lambda_l$, the solution $p_1^j$ is decreasing in $c^h(j)$ if $p \mapsto F_j(p)/p$ is increasing. Similarly, $p_2^j$ is increasing in $c^l(j)$ if $p \mapsto (1-F_j(p))p$ is decreasing, which is equivalent to $f(p)p \geq (1-F(p))$. This holds under strong regularity.

Suppose $p_j < p_1^j$, so $H^j(p_j) = n^h + \frac{1}{p_1^j} n^l$ and $L^j(p_j) = 0$. If $c^h(j') > c^h(j)$ and $c^l(j') = c^l(j)$ then by the previous claim $p_1^j \geq p_1^{j'}$. Similarly, if $p_j > p_2^j$, $c^h(j') = c^h(j)$ and $c^l(j') > c^l(j)$ then $p_2^j < p_2^{j'}$.

The workload of agent $j$ is

$$\max_p \mathbb{1}[p \leq p_1^j] p_j (n^h + \frac{1}{p_1^j} n^l) + \mathbb{1}[p_1^j < p < p_2^j](p_j n^h + n^l) + \mathbb{1}[p \geq p_2^j](p_2^j n^h + n^l).$$

If $p_j \leq p_1^j$ then the marginal impact of increasing $-p_j^1$ is $(p_1^j)^{-2} p_j n^l < 0$. Similarly if $p_j \geq p_2^j$ then the marginal impact of reducing $p_2^j$ is $-n^h < 0$. Finally, if $p_j \in (p_1^j, p_2^j)$ then the agent's welfare is invariant to local perturbations of $c^h(j), c^l(j)$. This proves local fairness for $j$.

We now prove that the mechanism is locally effort inducing for non-remedial agents. Following the same argument used to prove local fairness above, it suffices to show that $p_j^1$ is decreasing in $c^h(j)$ and $p_j^2$ is increasing in $c^l(j)$. Unlike comparisons across agents, however, to evaluate changes in an individual agent's performance we need to know how $\lambda^h, \lambda^l$ respond to changes in $c^h(j), c^l(j)$.

To this end, recall the conclusions of Lemma 3. From eq. (4) we can see that in order for the expected allocation of type-$k$ cases for $j$ to strictly increase, it must be that $(\lambda_h - c^h(j)/(\lambda_l - c^l(j))$ increases. This follows from convexity of $\mathcal{F}_j$. Given this observation, the comparative statics above (for fixed $\lambda_h, \lambda_l$) go through unchanged. □

## C.2 Proof of Proposition 2

Because of the notational complexity involved in stating the SMS-TP' mechanism, for clarity we first prove a version of the result for the SMS-TP mechanism under an auxiliary assumption. We then state the SMS-TP' mechanism, and show that the proof technique extends to this case, dropping the auxiliary assumption.

## C.3 The SMS-TP mechanism

Strategy-proofness follows from linearity of preferences and the fact that prices are fixed: If $p_j \leq p_1^j$ then $j$ prefers the allocation $(n^h + b_j, n^l - p_1^j b_j)$ to $(n^h - s_j, n^l + p_2^j s_j)$ for any $b_j, s_j \geq 0$.

To prove convergence, we make the following auxiliary assumption. Let $\mathcal{Z} := \{j \in \mathcal{J} : \lambda_h - c^h(j) = \lambda_l - c^l(j) = 0\}$, and let $(\tilde{n}_j^h, \tilde{n}_j^l)$ be the expected allocation for $j$ in LMS-TP.

**Assumption.** $\bar{p}_j \tilde{n}_j^h + \tilde{n}_j^l \leq \bar{p}_j n^h + n^l$ and $\underline{p}_j \tilde{n}_j^h + \tilde{n}_j^l \leq \underline{p}_j n^h + n^l$ for all $j \in \mathcal{Z}$.

This assumption ensures that SMS-TP is able to replicate LMS-TP as the market grows.

Consider first the case of $y \to \infty$. First, notice that in the large-market problem, there is an optimal mechanism which gives all identical agents the same allocation. This follows from eq. (5). We focus on this mechanism, and show that SMS-TP approximates it as $y \to \infty$.

In the replica economy, we index the $k^{th}$ copy of agent $j$ as $(j, k)$, so for example $p_{j,k}$ is the type of this agent. In theory, we could treat each $(j, k)$ as an separate agent. However in order to obtain a lower bound for $V_{SMS}$, we assume that if $(j, k)$ and $(j, k')$ are both buyers (or both sellers) then they receive the same allocation. With a slight abuse of notation, denote the allocation for $j$'s copies who are buyers as $b_j$, and for those who are sellers as $s_j$.

Given a realized type profile $P$, let $\hat{F}_j(\cdot|P, y)$ be the empirical CDF of types among the $y$ copies of agent $j$. So $y \cdot \hat{F}_j(p_1^j|P, y)$ is the number of the $j$-replica agents who are buyers, and $y\left(1 - \hat{F}_j(p_2^j|P, y)\right)$ is the number of these agents who are sellers. Then for a given type profile $P$ we can write the program defining SMS-TP in the replica economy as

$$\begin{aligned}
\min_{\{b_j\}_{j\in\mathcal{B}}, \{s_j\}_{j\in\mathcal{S}}, \{\dot{n}_j^h, \dot{n}_j^l\}_{j\notin\mathcal{E}}} \quad & \sum_{j\in\mathcal{E}} \hat{F}_j(p_1^j|P, y) b_j \left(c^h(j) - p_1^j c^l(j)\right) \\
& - \sum_{j\in\mathcal{E}} \left(1 - \hat{F}_j(p_2^j|P, y)\right) s^j \left(c^h(j) - p_2^j c^l(j)\right) + \sum_{j\notin\mathcal{E}} \dot{n}_j^h c^h(j) + \dot{n}_j^l c^l(j) \\
s.t. \quad & \bar{p}_j \dot{n}_j^h + \dot{n}_j^l \leq \bar{p}_j n_j^h + n_j^l \quad \forall\ j \notin \mathcal{E} \\
& \underline{p}_j \dot{n}_j^h + \dot{n}_j^l \leq \underline{p}_j n_j^h + n_j^l \quad \forall\ j \notin \mathcal{E}
\end{aligned} \tag{9}$$

$$0 \le \dot{n}_j^h,\ 0 \le \dot{n}_j^l \quad \forall\ j \notin \mathcal{E}$$

$$0 \le b_j \le \frac{n^l}{p_1^j} \quad \forall\ j \in \mathcal{B}$$

$$0 \le s_j \le n^h \quad \forall\ j \in \mathcal{S}$$

$$\sum_{j\in\mathcal{B}} \hat{F}_j(p_1^j|P,y)b_j - \sum_{j\in\mathcal{S}} \left(1 - \hat{F}_j(p_2^j|P,y)\right) s_j + \sum_{j\notin\mathcal{E}} \dot{n}_j^h + |\mathcal{E}|n^h = Jn^h \qquad (h\text{-capacity})$$

$$-\sum_{j\in\mathcal{B}} \hat{F}_j(p_1^j|P,y)p_1^j b_j + \sum_{j\in\mathcal{S}} \left(1 - \hat{F}_j(p_2^j|P,y)\right) p_2^j s_j + \sum_{j\notin\mathcal{E}} \dot{n}_j^l + |\mathcal{E}|n^l = Jn^l \qquad (l\text{-capacity})$$

Now notice if we replace $\hat{F}_j$ with $F_j$ for all $j$ in this program, the solution is precisely the LMS-TP allocation. This holds because of the auxiliary assumption, which ensures that the LMS-TP allocation for $j \in \mathcal{Z}$ (which implies $j \notin \mathcal{E}$) is feasible in the SMS-TP program (this is the only point in the proof where the auxiliary assumption is used). Moreover, by the strong law of large numbers $\hat{F}_j(p_1^j|P,y) \xrightarrow{a.s.} F_j(p_1^j)$ and $\hat{F}_j(p_2^j|P,y) \xrightarrow{a.s.} F_j(p_2^j)$ as $y \to \infty$. Thus to show that $V_{SMS}$ converges to $V_{OPT}$, we just need to show that the value of the program in eq. (9) continuous in $\hat{F}_j$, which is what we proceed to demonstrate.

Let $\hat{F}_j^1 = \hat{F}_j(p_1^j|P,y)$ and $\hat{F}_j^2 = \hat{F}_j(p_2^j|P,y)$. Let $R\left((\hat{F}_j^1, \hat{F}_j^2)_{j=1}^J\right)$ be the set of feasible choice variables $\left((b_j)_{j\in\mathcal{B}}, (s_j)_{j\in\mathcal{S}}, (\dot{n}_j^h, \dot{n}_j^l)_{j\notin\mathcal{E}}\right)$ in the above program given parameters $(\hat{F}_j^1, \hat{F}_j^2)_{j=1}^J$.

**Lemma 4.** *$R$ is upper and lower hemicontinuous.*

*Proof.* Define $\varphi\left((\hat{F}_j^1, \hat{F}_j^2)_{j=1}^J\right)$ as the set of $\left((b_j)_{j\in\mathcal{B}}, (s_j)_{j\in\mathcal{S}}, (\dot{n}_j^h, \dot{n}_j^l)_{j\notin\mathcal{E}}\right)$ satisfying

$$\bar{p}_j \dot{n}_j^h + \dot{n}_j^l \le \bar{p}_j n_j^h + n_j^l \quad , \quad \underline{p}_j \dot{n}_j^h + \dot{n}_j^l \le \underline{p}_j n_j^h + n_j^l \quad , \quad 0 \le \dot{n}_j^h,\ 0 \le \dot{n}_j^l \quad \forall\ j \notin \mathcal{E}$$

$$0 \le b_j \le \frac{n^l}{p_1^j} \quad \forall\ j \in \mathcal{B} \quad , \quad 0 \le s_j \le n^h \quad \forall\ j \in \mathcal{S}$$

$$\sum_{j\in\mathcal{B}} \hat{F}_j(p_1^j|P,y)b_j - \sum_{j\in\mathcal{S}} \left(1 - \hat{F}_j(p_2^j|P,y)\right) s_j + \sum_{j\notin\mathcal{E}} \dot{n}_j^h = Jn^h - |\mathcal{E}|n^h \qquad (h\text{-capacity})$$

and $\eta\left((\hat{F}_j^1, \hat{F}_j^2)_{j=1}^J\right)$ as the set satisfying

$$\bar{p}_j \dot{n}_j^h + \dot{n}_j^l \le \bar{p}_j n_j^h + n_j^l \quad , \quad \underline{p}_j \dot{n}_j^h + \dot{n}_j^l \le \underline{p}_j n_j^h + n_j^l \quad , \quad 0 \le \dot{n}_j^h,\ 0 \le \dot{n}_j^l \quad \forall\ j \notin \mathcal{E}$$

$$0 \le b_j \le \frac{n^l}{p_1^j} \quad \forall\ j \in \mathcal{B} \quad , \quad 0 \le s_j \le n^h \quad \forall\ j \in \mathcal{S}$$

$$-\sum_{j\in\mathcal{B}} \hat{F}_j(p_1^j|P,y)p_1^j b_j + \sum_{j\in\mathcal{S}} \left(1 - \hat{F}_j(p_2^j|P,y)\right) p_2^j s_j + \sum_{j\notin\mathcal{E}} \dot{n}_j^l = Jn^l - |\mathcal{E}|n^l \qquad (l\text{-capacity})$$

so that $R = \varphi \cap \eta$. Both $\phi$ and $\eta$ are given by the intersection of the polytope defined by

$$\bar{p}_j \dot{n}_j^h + \dot{n}_j^l \leq \bar{p}_j n_j^h + n_j^l \quad , \quad \underline{p}_j \dot{n}_j^h + \dot{n}_j^l \leq \underline{p}_j n_j^h + n_j^l \quad , \quad 0 \leq \dot{n}_j^h,\ 0 \leq \dot{n}_j^l \quad \forall\ j \notin \mathcal{E}$$
$$0 \leq b_j \leq \frac{n^l}{p_1^j} \quad \forall\ j \in \mathcal{B} \quad , \quad 0 \leq s_j \leq n^h \quad \forall\ j \in \mathcal{S}$$

with a hyperplane the normal vector of which is a continuous function of $(\hat{F}_j^1, \hat{F}_j^2)_{j=1}^J$. Thus both $\varphi$ and $\eta$ are upper and lower hemicontinuous. Since both are also convex and compact valued, upper and lower hemicontinuity of $R = \varphi \cap \eta$ follows.[44] □

Given Lemma 4, Berge's Maximum Theorem implies that the value of the program defining SMS-TP for the replica economy is continuous in $(\hat{F}_j^1, \hat{F}_j^2)_{j=1}^J$.

Finally, by the strong law of large numbers $\hat{F}_j(p_1^j|P, y) \xrightarrow{a.s.} F_j(p_1^j)$ and $\hat{F}_j(p_2^j|P, y) \xrightarrow{a.s.} F_j(p_2^j)$ as $y \to \infty$. Combined with continuity of the program defining SMS-TP, this implies convergence of the expected cost to $V_{SMS}$.

The case of $F_j$ converging in distribution to a constant for all $j$ is similar. Let $n \mapsto (F_j^n)_{j=1}^J$ be a sequence of distributions which converge in distribution to a vector of constants $(x_j)_{j=1}^J \in [\underline{p}, \bar{p}]^J$. (Note that we maintain the assumption that each $F_j^n$ is regular.) In the limit, i.e. when each investigator's type is known, $V_{SMS}$ and $V_{OPT}$ coincide. We now make use of the following intermediate result.

**Lemma 5.** $(F_j)_{j=1}^J \mapsto V_{OPT}((F_j)_{j=1}^J|y)$ and $(F_j)_{j=1}^J \mapsto (p_1^j, p_2^j)_{j=1}^J$ are continuous.

*Proof.* Recall that $(p_1^j, p_2^j)_{j=1}^J$ are defined from the solutions to eq. (5). $S^j$ is continuous in $F_j$. Moreover, if $F^j$ satisfies strict regularity for all $j$ then the objective in eq. (5) is unique. The lemma follows from Berge's maximum theorem. □

Moreover, following the same argument as that of Lemma 4, we can show that $V_{SMS}$ is continuous in $(p_1^j, p_2^j)_{j=1}^J$. Combined with Lemma 5, this implies that $V_{SMS}$ converges to $V_{OPT}$ along any sequence of strictly regular $(F_j^n)_{j=1}^J$, as desired.

## C.4 The SMS-TP' mechanism

We first describe the SMS-TP' mechanism. Let $(\lambda_h, \lambda_l)$ be the multipliers from the LMS-TP program, and let $(\tilde{n}_j^h, \tilde{n}_j^l)$ be the expected allocation for $j$ in LMS-TP. For each $j \in \mathcal{Z}$ there

[44] See for example Border (2013) Proposition 24 (for upper hemicontinuity) and Lechicki and Spakowski (1985) (for lower hemicontinuity).

exist prices $(p_1^j, p_2^j)$ and a number $\beta_j \in [0,1]$ such that

$$\tilde{n}_j^h = \beta_j \left( F_j(p_2^j)n^h + F_j(p_1^j)\frac{1}{p_1^j}n^l \right)$$
$$\tilde{n}_j^l = \beta_j \left( (1 - F_j(p_1^j))n^l + (1 - F_j(p_2^j))p_2^j n^h \right)$$

To identify the SMS-TP' allocation:

1. Define the sets $\hat{\mathcal{B}}$, $\hat{\mathcal{S}}$, $\mathcal{B}$, $\mathcal{S}$ by

$$j \in \hat{\mathcal{B}} \Leftrightarrow j \in \mathcal{Z} \text{ and } p_j \leq p_1^j$$
$$j \in \hat{\mathcal{S}} \Leftrightarrow j \in \mathcal{Z} \text{ and } p_j > p_2^j$$
$$j \in \mathcal{B} \Leftrightarrow j \in \mathcal{E} \text{ and } p_j \leq p_1^j$$
$$j \in \mathcal{S} \Leftrightarrow j \in \mathcal{E} \text{ and } p_j > p_2^j$$

2. Let $\dot{n}_j^h = n^h$ and $\dot{n}_j^l = n^l$ for all $j \in \mathcal{E} \setminus (\mathcal{B} \cup \mathcal{S})$.
3. Let $\dot{n}_j^h = \beta_j n^h$ and $\dot{n}_j^l = \beta_j n^l$ for all $j \in \mathcal{Z} \setminus (\hat{\mathcal{B}} \cup \hat{\mathcal{S}})$.
4. To allocate the remaining cases, we solve the linear program:

$$\min_{\{\hat{b}_j\}_{j\in\hat{\mathcal{B}}},\{\hat{s}_j\}_{j\in\hat{\mathcal{S}}},\{b_j\}_{j\in\mathcal{B}},\{s_j\}_{j\in\mathcal{S}},\{\dot{n}_j^h,\dot{n}_j^l\}_{j\notin(\mathcal{E}\cup\mathcal{Z})}} \sum_{j\in\mathcal{B}} b_j \left(c^h(j) - p_1^j c^l(j)\right) - \sum_{j\in\mathcal{S}} s_j \left(c^h(j) - p_2^j c^l(j)\right)$$
$$+ \sum_{j\in\hat{\mathcal{B}}} \beta_j b_j \left(c^h(j) - p_1^j c^l(j)\right) + \sum_{j\in\hat{\mathcal{S}}} \beta_j s_j \left(c^h(j) - p_2^j c^l(j)\right) + \sum_{j\notin(\mathcal{E}\cup\mathcal{Z})} \dot{n}_j^h c^h(j) + \dot{n}_j^l c^l(j)$$
$$s.t. \quad \bar{p}_j \dot{n}_j^h + \dot{n}_j^l \leq \bar{p}_j n_j^h + n_j^l \quad \forall\, j \notin (\mathcal{E} \cup \mathcal{Z})$$
$$\underline{p}_j \dot{n}_j^h + \dot{n}_j^l \leq \underline{p}_j n_j^h + n_j^l \quad \forall\, j \notin (\mathcal{E} \cup \mathcal{Z})$$
$$0 \leq \dot{n}_j^h,\ 0 \leq \dot{n}_j^l \quad \forall\, j \notin (\mathcal{E} \cup \mathcal{Z})$$
$$0 \leq b_j \leq \frac{n^l}{p_1^j} \quad \forall\, j \in \mathcal{B} \cup \hat{\mathcal{B}}$$
$$\leq s_j \leq n^h \quad \forall\, j \in \mathcal{S} \cup \hat{\mathcal{S}}$$
$$|\mathcal{E} \setminus (\mathcal{B} \cup \mathcal{S})|n^l + \sum_{j\in\mathcal{Z}\setminus(\hat{\mathcal{B}}\cup\hat{\mathcal{S}})} \beta_j n^h + \sum_{j\in\hat{\mathcal{B}}} \beta_j(n^h + b_j) + \sum_{j\in\hat{\mathcal{S}}} \beta_j(n^h - s_j) + \sum_{j\in\mathcal{B}}(n^h + b_j)$$
$$+ \sum_{j\in\mathcal{S}}(n^h - s_j) + \sum_{j\notin(\mathcal{E}\cup\mathcal{Z})} \dot{n}_j^h = Jn^h \qquad (h\text{-capacity})$$
$$|\mathcal{E} \setminus (\mathcal{B} \cup \mathcal{S})|n^l + \sum_{j\in\mathcal{Z}\setminus(\hat{\mathcal{B}}\cup\hat{\mathcal{S}})} \beta_j n^l + \sum_{j\in\mathcal{B}} \beta_j(n^l - p_1^j b_j) + \sum_{j\in\mathcal{S}} \beta_j(n^l + p_2^j s_j) + \sum_{j\in\mathcal{B}}(n^l - p_1^j b_j)$$

$$+\sum_{j\in\mathcal{S}}(n^l + p_2^j s_j) + \sum_{j\notin(\mathcal{E}\cup\mathcal{Z})} \dot{n}_j^l = Jn^l \qquad (l\text{-capacity})$$

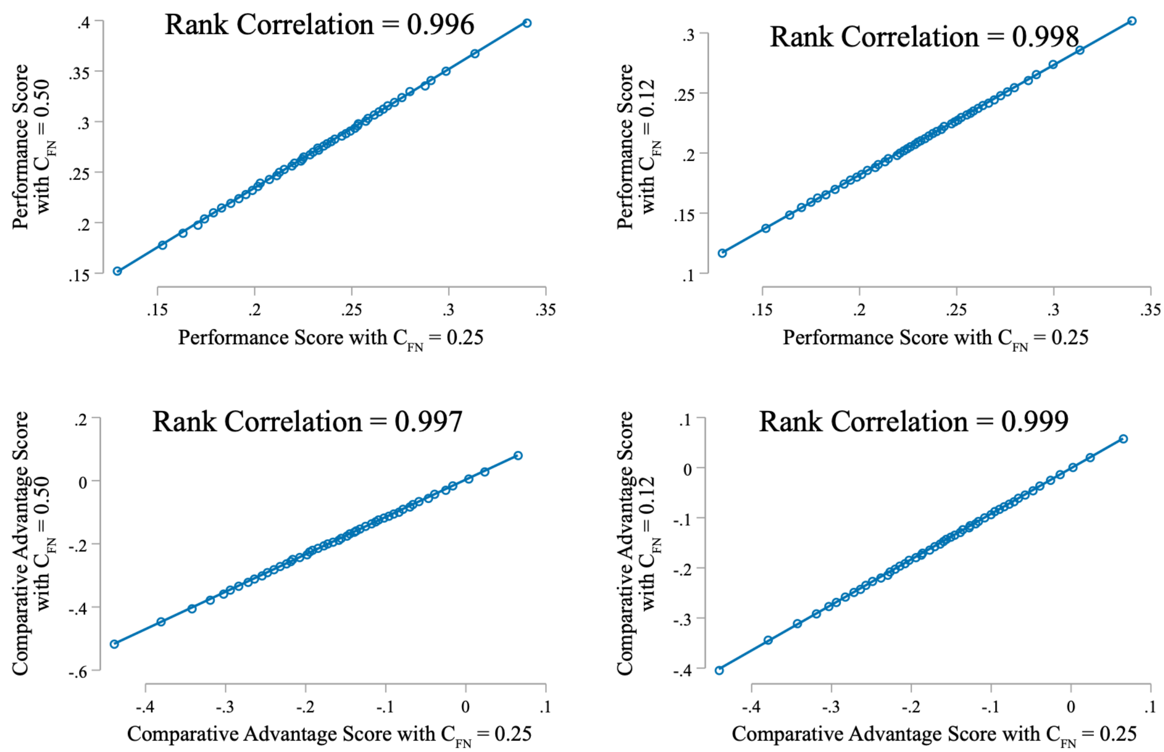
Let $(b_j^*)_{j\in\mathcal{B}\cup\hat{\mathcal{B}}}, (s_j^*)_{j\in\mathcal{S}\cup\hat{\mathcal{S}}}$ be the solution to this program.

5. For $j \in \mathcal{B}$ define $\dot{n}_j^h = n^h + b_j^*$ and $\dot{n}_j^l = n^l - p_1^j b_j^*$.
6. For $j \in \mathcal{S}$ define $\dot{n}_j^h = n^h - s_j^*$ and $\dot{n}_j^l = n^l + p_2^j s_j^*$.
7. For $j \in \hat{\mathcal{B}}$ define $\dot{n}_j^h = \beta_j(n^h + b_j^*)$ and $\dot{n}_j^l = \beta_j(n^l - p_1^j b_j^*)$.
8. For $j \in \hat{\mathcal{S}}$ define $\dot{n}_j^h = \beta_j(n^h - s_j^*)$ and $\dot{n}_j^l = \beta_j(n^l + p_2^j s_j^*)$.

Strategy-proofness holds for the same reason as SMS-TP. Moreover, if $\hat{F}_j = F_j$ for all $j$ then solution to the program defining SMS-TP' in the replica economy is precisely the LMS-TP allocation. This is because the SMS-TP' mechanism can replicate the LMS-TP assignment for $j \in \mathcal{Z}$, which was the only place we used the auxiliary assumption above. Thus we can drop the auxiliary assumption, and apply the same proof as above to show convergence of $V_{SMS}$ to $V_{OPT}$.

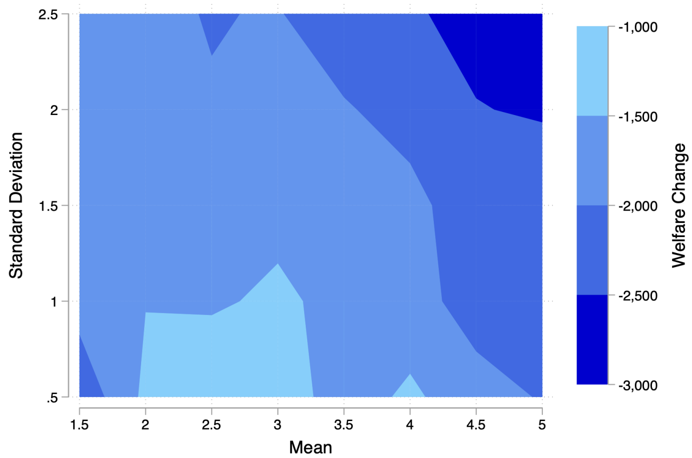
# D Supplemental Figures and Tables

Figure A1: Robustness to Different Choices of Social Costs



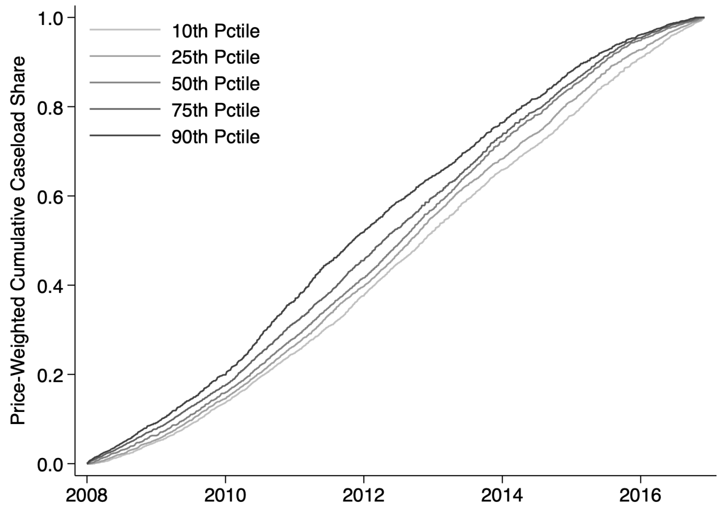

**Notes.** This figure plots the relationship between performance scores, $\gamma_j$, and comparative advantage scores, $d_j$, as we vary the choice of social costs. Benchmark estimates of $\gamma_j$ and $d_j$ are reported in the x-axis of each subfigure, with $(c_{TP}, c_{FN}, c_{FP}) = (0, 0.25, 1)$. In the left subfigures, we re-estimate $\gamma_j$ and $d_j$ with $c_{FN} = 0.50$. In the right subfigures, we re-estimate $\gamma_j$ and $d_j$ with $c_{FN} = 0.12$. Binned scatter plot estimates of the new performance score versus the benchmark performance score are displayed with 50 bins in each figure. We also report the Spearman's rank correlation coefficient between the new performance score measure and the benchmark measure. To minimize noise, for the comparative advantage estimates, the sample is limited to investigators that were assigned to at least 50 high-risk and low-risk cases across the sample. Investigator-specific and case type-specific estimates of subsequent maltreatment and placement rates are estimated via a regression adjustment for zipcode-by-year fixed effects.

Figure A2: Social Welfare Grains Across Distributional Assumptions




**Notes.** This figure presents the reduction in social costs (relative to the status quo) from implementing the LMS-TP mechanism under the assumption that $F_j$ is truncated normal with mean, $\mu$, and standard deviation, $\sigma$ (shown in the horizontal and vertical axis, respectively). The distribution is truncated to $[1, \mu + 2\sigma]$. For computational purposes, we implement this exercise in the LMS-TP mechanism, for which the SMD-TP is an approximation.

Figure A3: Smoothing Investigator Caseloads Over Time




**Notes.** This figure reports the distribution of cumulative price-weighted caseloads assigned by the SMD-TP mechanism over time. The sample is limited to the 34 counties that appear in every sample year. We re-estimate welfare gains under SMD-TP for this sample and find very similar results relative to those in Table 1. Cases are assigned according to the SMD-TP mechanism for one investigator type draw where types are drawn from the regularized own-type empirical distribution. We compute cumulative price-weighted caseloads in day $t$ for investigator $j$ as $\frac{\hat{n}_j^l(t)+p_j\hat{n}_j^h(t)}{\hat{n}_j^l(T)+p_j\hat{n}_j^h(T)}$, where $T$ is the last day of the sample period. We then report percentiles of this statistic for each day of our sample.

Table A1: Summary Statistics

| | |
|---|---|
| *Panel A: Child Socio-Demographics* | |
| White | 0.597 |
| Black | 0.266 |
| Female | 0.482 |
| Child had a previous investigation | 0.445 |
| Number of previous investigations | 1.024 |
| Age at investigation | 6.791 |
| | |
| *Panel B: Investigation Traits* | |
| Alleged perpetrator is the mother/stepmother | 0.772 |
| Alleged perpetrator is the father/stepfather | 0.328 |
| Alleged perpetrator is a non-parent relative | 0.053 |
| Investigation included a domestic violence allegation | 0.103 |
| Investigation included an improper supervision allegation | 0.530 |
| Investigation included a medical neglect allegation | 0.046 |
| Investigation included a physical abuse allegation | 0.290 |
| Investigation included a physical neglect allegation | 0.435 |
| Investigation included a substance abuse allegation | 0.170 |
| | |
| *Panel C: Outcome, if left at home* | |
| Re-investigated for child maltreatment within 6 months | 0.164 |
| | |
| Foster care rate | 0.032 |
| Number of investigations | 322,758 |
| Number of children | 261,021 |
| Number of investigators | 908 |

**Notes.** This table summarizes the analysis sample. The sample consists of maltreatment investigations of children in MI between 2008 and 2016, assigned to investigators who handled at least 200 cases during this period. The sample excludes repeat investigations and investigations of sexual abuse, as discussed in the main text. The final sample consists of 322,758 unique investigations of 261,021 children assigned to 908 investigators. Investigations can include multiple allegations and perpetrators, so these categories are not mutually exclusive.

Table A2: Estimates of Measures of Performance on Investigator Prediction Errors

| | (1) False Negative | (2) Foster Care Placement | (3) False Positive |
|---|---|---|---|
| *Panel A: Across all Cases* | | | |
| Standardized | 1.13*** | 0.49*** | 1.62*** |
| Performance Score | (0.13) | (0.09) | (0.14) |
| *Panel B: Across High-Risk Cases* | | | |
| Standardized Comparative | -2.04*** | -0.25 | -2.29*** |
| Advantage Score | (0.32) | (0.37) | (0.56) |
| *Panel C: Across Low-Risk Cases* | | | |
| Standardized Comparative | 0.04 | 0.23 | 0.26 |
| Advantage Score | (0.18) | (0.24) | (0.22) |

**Notes.** This table reports the results of OLS regressions of the investigator's false negative, foster care, and false positive rates on measures of their performance, $\gamma_j$ and comparative advantage on high-risk cases, $d_j$. The independent variables are standardized to mean 0 variance 1, and are estimated only in the randomized 50% training set. False negative rates and placement rates are estimated on the evaluation set, and are computed using a standard empirical Bayes shrinkage procedure. Implied false positive changes in Column 3 are estimated as the sum of coefficient estimates from Column 1 and Column 2, as $\text{FP}_j - \text{FP}_{j'} = (\text{FN}_j - \text{FN}_{j'}) + (P_j - P_{j'})$ by Lemma 1. In Panel A, we estimate this specification across all cases, in Panel B only among high-risk cases, and in Panel C among low-risk cases. All regressions are weighted by estimates of the inverse variance (clustered by investigator) of the investigator's performance or comparative advantage score. Baseline false negative and false positive rates are 15.9% and 2.3% over all cases, 24.0% and 4.0% over high-risk cases, and 13.1% and 1.9% over low-risk cases. False positive rates are identified via an identification-at-infinity strategy, described in Appendix E.4. Robust standard errors are reported in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Hazard Ratio Estimates of Risky Caseload Effect on Investigator Turnover

| | (1) Career Length | (2) Career Length |
|---|---|---|
| Mean Risk Level (Normalized) | 2.488*** (0.380) | |
| Above Median High-Risk Share | | 1.538*** (0.158) |
| Investigator Count | 908 | 908 |

**Notes.** This table reports the results of estimating a Cox proportional hazards model of investigator career length on caseload risk measures. We record an investigator's career length as the distance (in days) between their first and last observed CPS case assignment, and denote that this length is censored if the investigator is working in 2016 (the final year of the sample). Column 1 uses mean risk level—the average algorithmic predicted risk score across all of this investigator's cases, normalized to mean 0 variance 1 within each sample. Columns 2 uses an indicator recording whether the share of an investigator's cases that are high-risk is above the median for this sample. All estimates include a modal county fixed effect. We report the point estimates in terms of hazard ratios, with robust standard errors in parenthesis.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Gains from Investigator Reallocation, Alternative Type Distributions

| | (1) Unif[1,2] | (2) Unif[1,3] | (3) $\mathcal{N}(2, 0.5^2)$ | (4) $\mathcal{N}(2, 1^2)$ | (5) $p_j = 2$ | (6) Known type, Unif[1,2] |
|---|---|---|---|---|---|---|
| Social Costs | -947.1*** | -977.6*** | -925.5*** | -940.6*** | -1,716.5*** | -1,860.1*** |
| | (259.7) | (232.0) | (229.2) | (246.0) | (163.3) | (285.5) |
| | [-4.7%] | [-4.9%] | [-4.6%] | [-4.7%] | [-8.5%] | [-9.2%] |
| False Negatives | -624.3*** | -641.6*** | -618.0*** | -618.1*** | -1,140.2*** | -1,282.3*** |
| | (211.2) | (180.1) | (170.5) | (191.7) | (103.7) | (234.0) |
| | [-1.2%] | [-1.3%] | [-1.2%] | [-1.2%] | [-2.2%] | [-2.5%] |
| False Positives | -805.2*** | -828.5*** | -785.6*** | -798.5*** | -1,460.7*** | -1,571.4*** |
| | (224.4) | (197.1) | (179.3) | (202.8) | (124.7) | (248.7) |
| | [-11.0%] | [-11.3%] | [-10.7%] | [-10.9%] | [-20.0%] | [-21.5%] |
| Placements | -180.9* | -186.9** | -167.6** | -180.4** | -320.5*** | -289.1*** |
| | (95.7) | (81.3) | (78.5) | (83.3) | (62.2) | (100.1) |
| | [-1.8%] | [-1.9%] | [-1.7%] | [-1.8%] | [-3.2%] | [-2.9%] |

**Notes.** This table reports the welfare gains derived from the SMD-TP mechanism under different distributional assumptions. We estimate welfare changes and robust standard errors (in parenthesis), clustered by investigator, using the approach described in the text. Percent effects relative to the baselines are presented in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Importance of Small Market and Dynamic Considerations

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | Regularized MRS | | | Unif[1,2] | |
| | LMS-TP | SMS-TP | SMD-TP | LMS-TP | SMS-TP | SMD-TP |
| Social Costs | -2,363.7*** | -1,189.0*** | -1,189.2*** | -1,850.8*** | -948.4*** | -947.1*** |
| | (114.9) | (311.4) | (324.2) | (93.4) | (273.0) | (266.3) |
| | [-11.8%] | [-5.9%] | [-5.9%] | [-9.2%] | [-4.7%] | [-4.7%] |
| False Negatives | -1,516.1*** | -755.1*** | -754.9*** | -1,228.4*** | -625.1*** | -624.3*** |
| | (77.8) | (236.1) | (252.9) | (70.4) | (204.5) | (195.8) |
| | [-3.0%] | [-1.5%] | [-1.5%] | [-2.4%] | [-1.2%] | [-1.2%] |
| False Positives | -2,028.8*** | -1,016.8*** | -1,017.0*** | -1,563.2*** | -806.2*** | -805.2*** |
| | (93.4) | (245.2) | (262.9) | (77.4) | (229.2) | (220.6) |
| | [-27.7%] | [-13.9%] | [-13.9%] | [-21.4%] | [-11.0%] | [-11.0%] |
| Placements | -512.6*** | -261.7*** | -262.1** | -334.8*** | -181.1* | -180.9* |
| | (49.3) | (99.1) | (104.7) | (38.4) | (94.5) | (96.0) |
| | [-5.2%] | [-2.6%] | [-2.6%] | [-3.4%] | [-1.8%] | [-1.8%] |

**Notes.** This table reports the welfare gains derived from the LMS-TP, SMS-TP, and SMD-TP mechanisms. Columns 1-3 estimate welfare gains under the distribution assumption that investigator type is drawn from the regularized own-type empirical distribution: we estimate hazards from a kernel-smoothed empirical distribution from the survey and impose monotone hazards (Myerson regularity) by choosing the nondecreasing hazard sequence that minimizes the discrete sum of squared deviations in hazard rate. Columns 4 to 6 estimate welfare gains from uniform distributions with support $[1, 2]$. We estimate welfare changes and robust standard errors (in parenthesis), clustered by investigator, using the approach described in the text. Percent effects relative to the baselines are presented in brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Welfare Gains Under Alternative Proxies for Subsequent Maltreatment

| | (1) Baseline | (2) Inv. within 5 months | (3) Inv. within 4 months | (4) Inv. within 3 months | (5) Inv. within 2 months | (6) Inv. within 1 months | (7) Subst. Inv. within 6 months |
|---|---|---|---|---|---|---|---|
| Social Costs | -2,363.7*** | -2,014.7*** | -1,851.1*** | -1,659.7*** | -1,556.1*** | -1,477.8*** | -2,048.8*** |
| | (114.1) | (103.0) | (105.9) | (97.1) | (86.3) | (77.1) | (74.7) |
| | [-11.8%] | [-10.8%] | [-10.9%] | [-10.6%] | [-10.9%] | [-11.8%] | [-16.1%] |
| False Negatives | -1,516.1*** | -1,275.6*** | -1,170.4*** | -869.0*** | -788.4*** | -502.0*** | -924.2*** |
| | (83.5) | (69.7) | (80.2) | (77.0) | (54.6) | (43.6) | (54.5) |
| | [-3.0%] | [-2.9%] | [-3.2%] | [-3.0%] | [-3.9%] | [-4.6%] | [-6.1%] |
| False Positives | -2,028.8*** | -1,728.5*** | -1,591.0*** | -1,473.8*** | -1,388.6*** | -1,355.5*** | -1,814.3*** |
| | (96.9) | (88.3) | (96.6) | (89.4) | (70.3) | (68.3) | (71.7) |
| | [-27.7%] | [-22.8%] | [-20.5%] | [-17.6%] | [-15.2%] | [-13.7%] | [-20.3%] |
| Placements | -512.6*** | -452.9*** | -420.6*** | -604.8*** | -600.2*** | -853.5*** | -890.1*** |
| | (46.2) | (49.0) | (45.6) | (40.7) | (44.0) | (51.3) | (49.4) |
| | [-5.2%] | [-4.6%] | [-4.2%] | [-6.1%] | [-6.0%] | [-8.6%] | [-9.0%] |

**Notes.** This table reports the welfare gains derived from the LMS-TP mechanism for different definitions of maltreatment risk under the distributional assumption that investigator types are drawn from the regularized own-type empirical distribution. We estimate welfare changes and robust standard errors (in parenthesis), clustered by investigator, using the approach described in the text. Percent effects relative to the baselines are presented in brackets. For computational purposes, we compare the welfare gains across different proxies for subsequent maltreatment in the LMS-TP mechanism, for which the SMD-TP is an approximation. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Welfare Gains from Observed Counterfactual

| | (1) Own Type | (2) Coworker Type | (3) Known Type, Own Type |
|---|---|---|---|
| Social Costs | -1,277.0*** | -1,198.9*** | -3,439.5*** |
| | (298.5) | (271.1) | (441.7) |
| | [-6.3%] | [-5.9%] | [-16.9%] |
| False Negatives | -815.4*** | -773.8*** | -2,269.5*** |
| | (242.7) | (215.1) | (383.9) |
| | [-1.6%] | [-1.5%] | [-4.4%] |
| False Positives | -1,091.7*** | -1,024.7*** | -2,939.9*** |
| | (260.3) | (221.2) | (418.3) |
| | [-14.5%] | [-13.6%] | [-39.1%] |
| Placements | -276.3** | -250.8** | -670.4*** |
| | (108.9) | (105.9) | (203.5) |
| | [-2.8%] | [-2.5%] | [-6.7%] |

**Notes.** This table reports the welfare gains derived from the SMD-TP mechanism compared to a counterfactual approximating the observed assignment matrix that strictly restricts investigators to one county. This procedure creates some cases handled by investigators outside their modal counties—for such cases, we randomly reassign these cases to an investigator working in the focal county. Each column corresponds to a different distributional assumption for $p_j$. Column 1 presents gains from the regularized own-type empirical distribution: we estimate hazards from a kernel-smoothed empirical distribution from the survey and impose monotone hazards (Myerson regularity) by choosing the nondecreasing hazard sequence that minimizes the discrete sum of squared deviations in hazard rate (see Appendix E.2 for details). Column 2 performs the same procedure but using investigator reports of their coworkers' preferences. Column 3 uses the same type distribution as in Column 1, but assumes that $p_j$ is known to the designer. We estimate welfare changes and robust standard errors (in parenthesis), clustered by investigator, using the approach described in the text. Percent effects relative to the baselines are presented in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Welfare Gains for Five Largest Counties

| | (1)<br>Own Type | (2)<br>Coworker Type | (3)<br>Known Type,<br>Own Type |
|---|---|---|---|
| Social Costs | -846.5*** | -790.2*** | -2,005.7*** |
| | (211.7) | (172.4) | (340.3) |
| | [-9.7%] | [-9.1%] | [-23.0%] |
| False Negatives | -509.0*** | -476.5*** | -1,222.2*** |
| | (171.0) | (143.9) | (255.6) |
| | [-2.3%] | [-2.1%] | [-5.5%] |
| False Positives | -728.0*** | -681.0*** | -1,730.7*** |
| | (182.6) | (151.1) | (261.4) |
| | [-23.1%] | [-21.6%] | [-54.9%] |
| Placements | -219.0*** | -204.5*** | -508.5*** |
| | (84.8) | (76.3) | (143.9) |
| | [-5.4%] | [-5.0%] | [-12.5%] |

**Notes.** This table reports the welfare gains derived from the SMD-TP mechanism, using a sample of the 5 largest counties in the data. Each column corresponds to a different distributional assumption for $p_j$. Column 1 presents gains from the regularized own-type empirical distribution. Column 2 performs the same procedure but using investigator reports of their coworkers' preferences. Column 3 uses the same type distribution as in Column 1, but assumes that $p_j$ is known to the designer. We estimate welfare changes and robust standard errors (in parenthesis), clustered by investigator, using the approach described in the text. Percent effects relative to the baselines are presented in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Welfare Gains from Binary versus Quartile Case Splits

| | (1) Binary Split | (2) Quartile Split |
|---|---|---|
| Social Costs | -1,506.1*** | -1,711.5*** |
| | (350.9) | (421.8) |
| | [-14.0%] | [-16.3%] |
| False Negatives | -922.2*** | -1,004.7*** |
| | (289.8) | (289.1) |
| | [-3.5%] | [-3.8%] |
| False Positives | -1,328.5*** | -1,477.0*** |
| | (326.2) | (303.8) |
| | [-32.0%] | [-37.7%] |
| Placements | -406.3*** | -472.3*** |
| | (152.5) | (160.8) |
| | [-7.6%] | [-9.1%] |

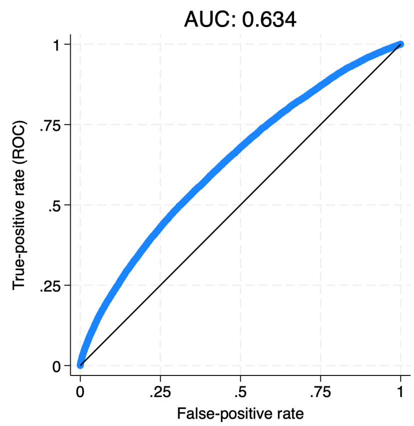
**Notes.** This table reports the welfare gains derived from the SMD-TP mechanism using a quartile split versus a binary split. To define performance for these new categories we re-estimate Equations 6 and 7, but allowing for interactions for 2nd risk quartile by worker, 3rd risk quartile by worker, and 4th risk quartile by worker. We estimate this separately in the training and test samples. This table is limited to workers assigned to 50+ cases of each risk quartile group for all analysis. In both columns, it is assumed that the designer knows worker preferences. In Column 1, preferences on high-risk (top quartile risk) cases are drawn according to regularized own-type empirical distribution: we estimate hazards from a kernel-smoothed empirical distribution from the survey and impose monotone hazards (Myerson regularity) by choosing the nondecreasing hazard sequence that minimizes the discrete sum of squared deviations in hazard rate (see Appendix E.2 for details). In Column 2, worker performance is estimated across risk quartiles. Worker preferences for the riskiest quartile, $p_{j4}$ are drawn from the regularized own-type empirical distribution as above. Worker preferences on the 3rd quartile, $p_{j3}$ are distributed uniformly between $[1.2, 1.5]$. Worker preferences on the 2nd quartile, $p_{j2}$ are distributed uniformly between $[0.9, 1.2]$. We then set $p_{j1} = 3 - p_{j2} - p_{j3}$ so that the mean of $p_{j1}, p_{j2}, p_{j3}$ is equal to 1 for comparability with binary results. We estimate welfare changes and robust standard errors (in parenthesis), clustered by investigator, using the approach described in the text. Percent effects relative to the baselines are presented in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

# E Empirical Appendix

## E.1 Maltreatment risk prediction

We estimate an algorithmic risk prediction that child $i$ will face subsequent maltreatment if left at home, $Pr(Y_i^\star = 1|X_i)$, where $X_i$ includes case and child attributes of case $i$ available to the investigator at the time of the placement decision. Following Kleinberg et al. (2018), we use a gradient boosted decision tree to predict $Pr(Y_i^\star = 1|X_i)$. We hypertune the algorithm to select for optimal parameters using a 5-fold cross-validation technique. Only children left at home are used to train the model since $Y_i^\star$ is unobserved for children placed in foster care. The features used to train the algorithm, $X_i$, are coded by the initial screener and include: the type of allegations in the investigation (physical abuse, medical neglect, physical neglect, domestic violence, substance abuse, improper supervision), the relationship of the alleged perpetrator to the child, prior child welfare investigation history, the gender and age of the child, and their residing county. The ROC of the algorithmic measure on left-at-home cases is displayed in Figure A4; the AUC on left-at-home cases on the 20% hold-out evaluation sample is 0.634.

Figure A4: ROC of Subsequent Maltreatment Risk Prediction




**Notes.** This figure reports the ROC and AUC of the subsequent maltreatment risk algorithm. The algorithm is trained on an 80% random sample of data, and the ROC and AUC are computed on the remaining 20% of data. The figure is limited to referrals where a child was not placed so the outcome is observed.

## E.2 Investigator Survey

We partnered with the Michigan Department of Health and Human Services (MDHHS) to design and administer a statewide survey of CPS investigators. Our goal was to better understand how investigators perceive their work—how they evaluate different types of investigations, how they manage caseload tradeoffs, and whether our modeling assumptions reflect how they themselves think about assignment and workload.

The survey served four main purposes and yields several insights:

1. **Testing core modeling assumptions.** A central goal of the survey was to assess whether our model reflects how investigators perceive their work. Specifically, we sought to understand whether the nature of the cases—rather than just the number—shapes preferences over caseloads. To test this assumption, we asked respondents to compare high- and low-risk investigations along three dimensions—time demand, emotional toll, and satisfaction—using 0–10 scales. The results suggest that investigators perceive high-risk cases as substantially more burdensome, and only modestly more satisfying.

2. **Estimating heterogeneity in preferences.** A core motivation for our mechanism design is the hypothesis that investigators may have idiosyncratic preferences over high- and low-type cases. The survey allows us to test this. By estimating each respondent's MRS between high- and low-risk cases, we recover a distribution of $p$ across the workforce. The results show substantial variation—some investigators are highly averse to high-risk cases, while others are more willing to take them on. As we discuss in the main text, estimating the distribution of investigators' MRS values also inform our empirical simulations of the potential welfare gains from implementing the mechanism.

3. **Connecting case complexity to outcomes.** In our administrative data, a one standard deviation increase in caseload risk is associated with a large increase in turnover risk. The survey provides complementary, qualitative evidence for this finding. Investigators consistently report that high-risk cases are more stressful and emotionally draining. Many explicitly link complexity to burnout and turnover.

4. **Probing the focus on binary-classification mechanisms.** The survey allows us to assess whether our focus on binary classification mechanisms can broadly capture how investigators perceive variation in their work in the real world. Across both structured and open-ended responses, we observe consistent separation between the two case types as defined in the paper (high- and low-risk). Investigators broadly prefer low-risk cases and describe high-risk cases as more demanding. These patterns suggest that our partition captures meaningful differences in how cases are experienced and valued. Moreover, the relatively high MRS estimates—along with text-free responses that we summarize below—indicate that a binary classification can capture significant heterogeneity in preferences.

The survey was distributed to CPS investigators via an MDHHS-managed email. While the listserv included approximately 1,200 to 1,400 recipients, the exact number of investigators who received the email is uncertain due to turnover and staff on leave. We received 459 responses, and, after filtering out incomplete entries, retained approximately 380 responses for descriptive sections and 322 for the MRS-elicitation questions. This yields a response rate between 23–27%. The survey link is available here.

### E.2.1 Survey Design and Key Measures

The survey consisted of four sections: open-ended reasoning, preferences over current workloads, perceptions of case types, and MRS elicitation.

In order to derive a practical policy, the paper focuses on binary-classification mechanisms. From a theoretical perspective, this partition can be arbitrary; in the empirical application, we define it using the predicted risk of subsequent maltreatment and split cases into high-and low-risk. In this section we sometimes refer to them as "high-" and "low-complexity" cases, as this terminology is more commonly used within Michigan CPS.

**High-risk investigations** are those in the top quartile of predicted risk. These cases typically involve younger children, families with prior CPS involvement, and multiple or severe allegations (e.g., substance abuse, domestic violence, physical neglect).

**Low-risk investigations** fall into the bottom three quartiles and typically involve: older children, families with little or no CPS history, and less severe or isolated allegations (e.g., minor neglect or improper supervision).

These definitions, along with a summary table of case characteristics, were presented to respondents at the beginning of the survey.

### E.2.2 Qualitative Explanations for Tradeoff Choices

Before delving into responses to structured questions, we begin by presenting responses to open-ended questions about why investigators prefer certain bundles of cases over others. We analyzed these open-text responses to identify recurring themes:

**1. High-Complexity Cases Are More Demanding.** Respondents consistently described high-complexity cases as more time-consuming, emotionally taxing, and harder to manage. One wrote: "High complexity investigations are more time consuming, are far more taxing to manage, [and] involve multiple households." Another said: "Just one more high-complexity case can be the time equivalent of three or four low-complexity cases."

**2. Emotional Toll and Burnout Are Salient.** Burnout and emotional exhaustion were common themes. "Secondary trauma is exhausting," one respondent wrote. Another added: "CPS's turnover rate is not because it's 'not for everyone'... it's because once people go into full rotation, the workload is too high."

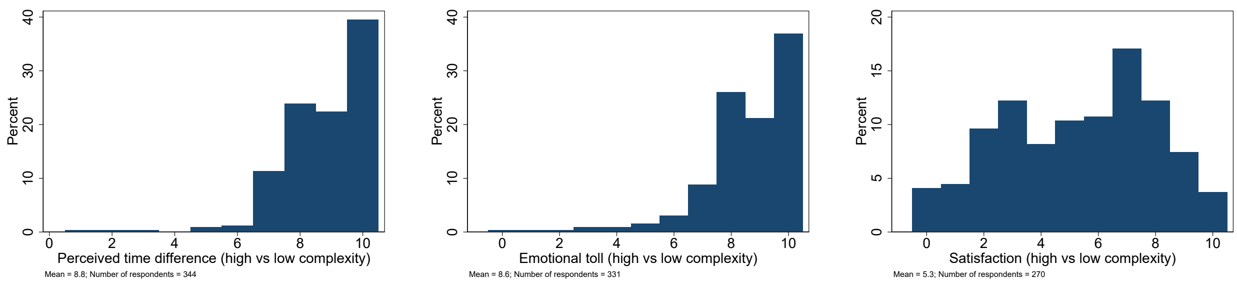
**3. Low-Complexity Cases Help Balance the Load.** Respondents emphasized that low-complexity cases allow time to complete other work and decompress. One wrote, "Low-complexity cases are a bit of a mental break. You need them to stay afloat."

**4. Tradeoff Logic Mirrors Our Model.** Many respondents articulated reasoning directly in line with the structure of our model. One wrote: "Two low complexity investigations normally equal one high complexity," while another noted: "I'd trade three low complexity cases for one high one, maybe."

### E.2.3 Testing Core Modeling Assumptions

To test whether high-risk cases are perceived as more burdensome, we asked respondents to evaluate them relative to low-risk cases across three dimensions: time demand, emotional toll, and satisfaction. Each was rated on a 0–10 scale, with higher values indicating more time, more stress, or more satisfaction associated with high-risk cases.

The average scores were 8.8 for time, 8.6 for emotional toll, and 5.3 for satisfaction—suggesting that investigators experience high-risk cases as significantly more demanding, but only modestly more rewarding. These responses offer support for our empirical finding that a one standard deviation increase in caseload risk is associated with a significant increase in turnover risk. Investigators consistently linked complexity to emotional burden, and, as we discussed above, several linked high-complexity cases to secondary trauma and turnover in open-ended responses.




(a) Time Demand (b) Emotional Toll (c) Satisfaction

Figure A5: Perceptions of High-Complexity Investigations (0 = Much Less, 10 = Much More)

### E.2.4 Eliciting the Marginal Rate of Substitution

To quantify tradeoffs between case types, we asked investigators to choose between hypothetical caseloads composed of different mixes of high- and low-complexity cases. Each respondent began by choosing between 12 high-complexity and 12 low-complexity cases. Based on their answer, we offered a sequence of comparisons in which one type was incrementally added and the other removed, until a preference switch was observed. These responses allowed us to estimate each investigator's MRS. Approximately 87% of respondents preferred 12 low-complexity cases to 12 high-complexity cases, implying an MRS of at least 1.

Figure A6 reports the CDF and density of the survey responses, along with the kernel density smoothed estimate. The average MRS in the sample was approximately 3.6—suggesting that, on average, investigators would be willing to trade one high-complexity case for 3.6 low-complexity ones. We also asked respondents to complete the same exercise from the perspective of a typical coworker. Results were similar, with a slightly higher average MRS of approximately 3.7.

The figure also shows the regularized own-type empirical distribution. As discussed in the main text, these survey responses provide a natural estimate of $F_j$ to use in our policy simulations. To apply the results in Corollary 1, we regularize the own-type distribution by estimating hazards from the kernel-smoothed distribution and imposing monotone hazards (Myerson regularity). Specifically, we select the nondecreasing hazard sequence that minimizes the discrete sum of squared deviations in hazard rates.

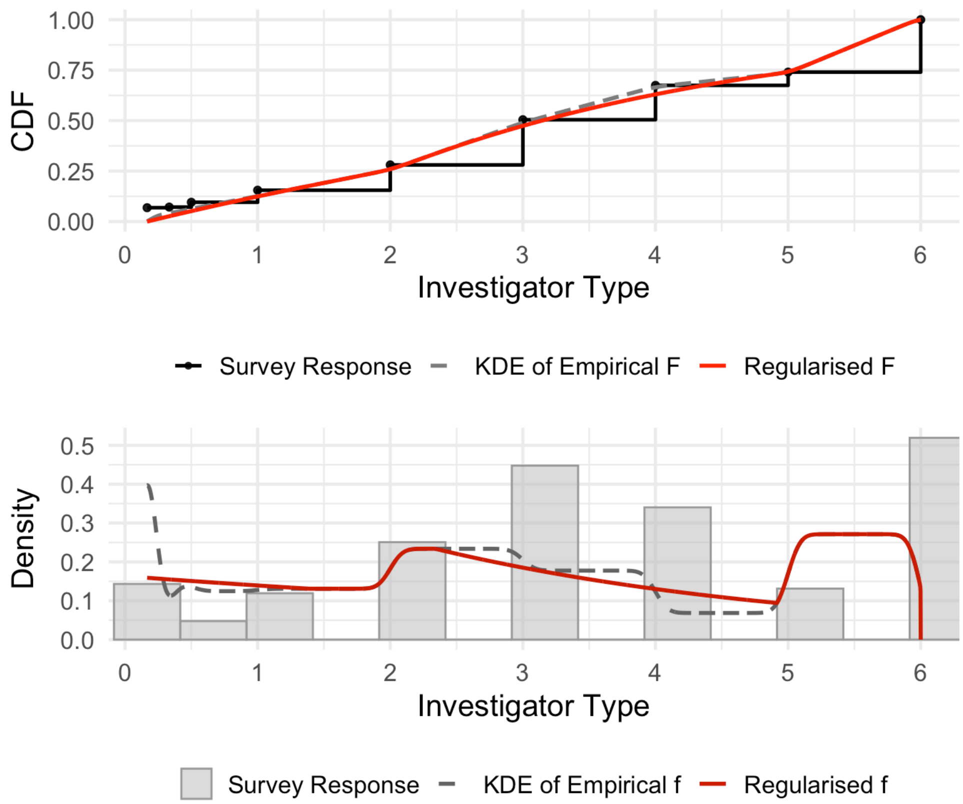



Figure A6: Investigator Own-Type Distribution

#### E.2.5 Field Validation of Complexity Definitions

The survey also allows us to asses whether the there is variation in the way that investigators experience and value case difficulty across the binary classification that we use. Across both structured and open-ended responses, we observe clear separation in preferences between case types. Most respondents preferred low-complexity bundles when asked to choose between hypothetical caseloads. When choosing bundles that reveal an MRS of greater than one, investigators tend to cite emotional toll and resource demands—validating that our partition can broadly capture how complexity is perceived. Taken together with the observed heterogeneity in MRS estimates—many of which are relatively high—the results suggest that our binary-classification captures meaningful variation in preferences.

### E.2.6 Summary

We believe that the survey broadly provides support for the assumptions and simplifications used in our model. The results also reinforce our empirical finding that complex caseloads are associated with increased turnover risk. Investigators clearly distinguish between high- and low-complexity cases, consistently describe high-complexity cases as more demanding, and exhibit substantial heterogeneity in preferences. Estimated MRS values vary meaningfully across the sample, and qualitative explanations confirm that investigators think about case assignments in terms of tradeoffs.

## E.3 Generalizing beyond binary outcomes

As discussed in the main text, the assumption that $Y_i^*$ is binary valued is innocuous. Suppose $Y_i^*$ takes values in a finite set $\mathcal{X}$. Maintain the assumption that when case $i$ is assigned to investigator $j$, $Y_i^*$ is observed if and only if $D_{ij} = 1$. For $X \in \mathcal{X}$ define $PX_{ij} = Pr(\{Y_i^* = x, D_{ij} = 1)$ and $NX_{ij} = Pr(\{Y_i^* = x, D_{ij} = 0)$. The joint distribution of $Y_i^*$ and $D_{ij}$ is described by the vector $(PX_{ij}, NX_{ij})_{X\in\mathcal{X}}$. The cost of assigning $i$ to $j$ is, as in the binary case, a linear function of the joint distribution, denoted by $c(i, j)$. Lemma 1 generalizes immediately to this setting.

**Lemma 6.** Assume that the observed assignment is random conditional on $I$. Then for any $j, j' \in \mathcal{J}$ whose assignments are supported on $I$, the following are identified:

- the difference $NX_j^I - PX_{j'}^I$,
- the cost difference $\mathbb{E}[c(i, j) - c(i, j')|i \in I]$.

*Proof.* As in the proof of Lemma 1, under the random assignment and full support assumptions we can identify $(NX_j^I)$ for all $X \in \mathcal{X}$. Let $S^I(X) = Pr(\{Y_i^* = X|i \in I\})$. Then $S^I(X) = PX_j^I + NX_j^I$, so

$$\begin{aligned} PX_j^I - PX_{j'}^I &= S^I(X) - NX_j^I - \left(S^I(X) - NX_{j'}^I\right) \\ &= -\left(NX_j^I - NX_{j'}^I\right). \end{aligned}$$

Given that we can identify the cost differences $\mathbb{E}[c(i, j) - c(i, j')|i \in I]$, the remainder of the mechanism-design analysis is unchanged. □

## E.4 Identification of false positive rates

Suppose we wish to identify $\mathbb{E}[FP_{ij}] = \mathbb{E}[D_{ij}] - \mathbb{E}[Y_i^\star] + \mathbb{E}[FN_{ij}]$. In that expression, $\mathbb{E}[FN_{ij}] = \mathbb{E}[FN_i|Z_{ij} = 1]$ and $\mathbb{E}[D_{ij}] = \mathbb{E}[D_i|Z_{ij} = 1]$ are identified under random assignment by the observed false negative rate and placement rate of each investigator (where $Z_{ij} = 1$

if investigator $j$ were assigned to case $i$).[45] However, $\mathbb{E}[Y_i^\star]$ is not identified as $Y_i^\star$ is not measured when $D_{ij} = 1$, or when the investigator places the child in foster care. Therefore, the identification challenge reduces to the challenge of identifying $\mathbb{E}[Y_i^\star]$.

To identify this parameter, we follow Arnold et al. (2022) and use an extrapolation-based identification strategy. To build intuition, suppose there exists an "infinitely lenient" investigator $j^\star$ with a placement rate of zero and that cases are randomly assigned to investigators. Then, $\mathbb{E}[Y_i^\star]$ of such an investigator would not suffer from selective observability concerns, since $D_{ij^\star}$ would equal zero for all $i$. Because cases are randomly assigned to investigators, the average subsequent maltreatment rate of cases assigned to this supremely lenient investigator would be close to the overall average: $\mathbb{E}\left[Y_i^\star | D_{ij^\star} = 0\right] \approx \mathbb{E}\left[Y_i^\star\right]$.

Without a supremely lenient investigator, this parameter can be estimated via extrapolation. Estimates of $\mathbb{E}[Y_i^\star]$ may come, for example, from the vertical intercept at zero of a linear, quadratic, or local linear regression of investigators' subsequent maltreatment rates (among children left at home) on their placement rates. As Arnold et al. (2022) discuss, this approach is similar to extrapolations of average potential outcomes near a treatment cutoff in a regression discontinuity design. Here, we extrapolate across randomly assigned investigators with very low placement rates. This method is related to "identification at infinity" approaches in sample selection models (Andrews and Schafgans, 1998; Chamberlain, 1986; Heckman, 1990) and has been used to identify selectively observed parameters in several recent studies (Arnold et al., 2021, 2022; Angelova et al., 2023; Baron et al., 2024). In practice, this approach works well whenever there are many decision-makers with low treatment rates. Because foster care placement rates are low (3% in our sample), the CPS setting is particularly well-suited to this approach, yielding limited extrapolation and precise estimates.

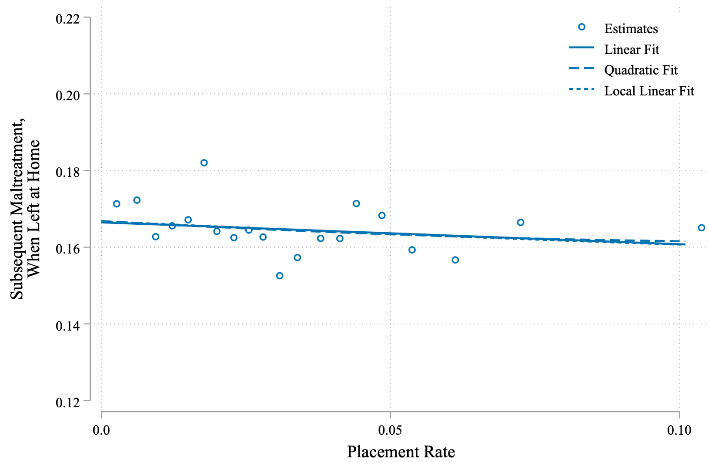
We use the strata-adjusted investigator-specific placement and subsequent maltreatment rates from Section IV to extrapolate toward the unselected first moment, $\mathbb{E}[Y_i^\star]$. Figure A7 reports the investigator-specific estimates that are used for the extrapolation, with a binned scatter plot of estimates of each investigator's placement and subsequent maltreatment rate (net of zip code by year fixed effects). The large mass of investigators with placement rates near zero suggests the extrapolation may be reliable in this context. We show extrapolations from linear, quadratic, and local linear regressions of each investigator's subsequent maltreatment rate among children left at home on their placement rate.

The vertical intercept at zero is the estimate of the unselected first moment of subsequent maltreatment. The most flexible local linear extrapolation yields an estimate of 0.167 (SE=0.001).

[45] We discuss how we handle conditional random assignment in Section IV.

Figure [A7](#) shows that alternative extrapolation specifications yield nearly identical point estimates.

Figure A7: Extrapolation Estimates of Average Subsequent Maltreatment Potential




**Notes.** This figure presents the results of the extrapolation strategy used to estimate $\mathbb{E}[Y_i^*]$. Binned scatter plot estimates of investigator-specific placement rates versus conditional subsequent maltreatment rates are displayed, with 20 bins. All estimates adjust for zipcode-by-year fixed effects, and are obtained from investigator-level regressions that inversely weight observations by variance of estimated subsequent maltreatment rate among children not placed in foster care. The local linear regression uses a Gaussian kernel with a rule-of-thumb bandwidth.

## E.5 Correlation between Preferences and Performance

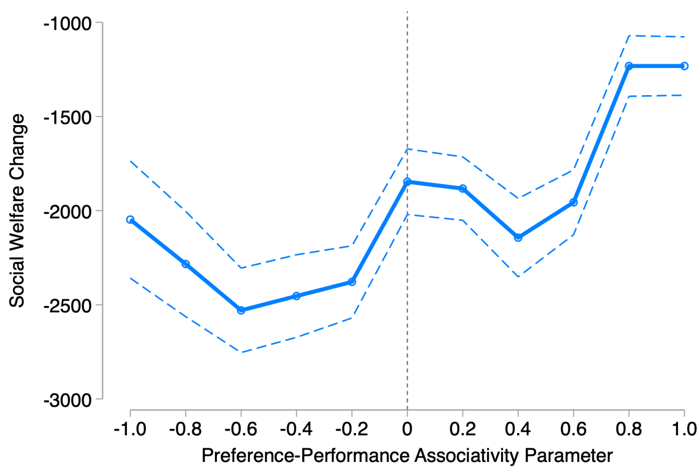
Finally, we consider how social welfare gains change if investigator preferences are correlated with their comparative advantage score, $d_j$. Let $\underline{d_j}, \bar{d}_j$ be the minimum and maximum $d_j$ within counties, respectively. Assume that investigator types are drawn from a uniform distribution with full support on $[g(d_j) + 1, g(d_j) + 2]$, where $g(d_j) = b\frac{d_j - \underline{d_j}}{\bar{d}_j - \underline{d_j}}$ for $b \geq 0$ and $g(d_j) = -b(1 - \frac{d_j - \underline{d_j}}{\bar{d}_j - \underline{d_j}})$ for $b < 0$.[46] Then the associativity parameter $b$ captures the strength and direction of the correlation between comparative advantage and preferences, where $b > 0$ indicates that investigators who are relatively good at high-risk cases tend to find such cases more costly.

For computational purposes, we compare the welfare gains for different values of $b$ in the LMS-TP mechanism, of which SMD-TP is an approximation. Figure [A8](#) reports the results

[46]Note that if $b = 0$, this reduces to the Unif$[1, 2]$ setting. This construction of $g(.)$ is also symmetric: for any $b \geq 0$, the investigator with maximal $d_j$ in the county has the same distribution under associativity parameter $b$ as the investigator with minimal $d_j$ under parameter $-b$.

of this exercise. When investigators with high $d_j$ tend to have lower type draws, we find that welfare gains are significantly larger: for $b = -1$, the welfare gains are $2,047$ relative to the expected social cost of the status quo. Compared to the $b = 0$ case where there is no association between preferences and performance, this is a 11% increase in the welfare gains. This confirms the intuition that when investigators with a comparative advantage in high-risk cases also relatively prefer these cases, the mechanism can achieve larger welfare gains. On the other hand, if investigators with a comparative advantage in high-risk cases tend to dislike such cases, the welfare gains are attenuated. The largest reduction in Figure A8 occurs when $b = 1$, in which case welfare gains are $1,232$, a 33% decline compared to the $b = 0$ case. Thus, while a strong positive correlation between $d_j$ and $p_j$ may reduce the potential welfare gains, there still exists a significant potential for welfare improvement even under this scenario.[47]

Figure A8: LMS Welfare Changes, Correlation Between Preference and Performance




**Notes.** This figure presents the welfare gains from the LMS-TP mechanism under distributional assumptions that allow for correlation between investigator type distributions, $F(p)$, and their comparative advantage score, $d_j$. Investigator types are drawn from a uniform distribution $[g(d_j)+1, g(d_j)+2]$, where $g(d_j) = b\frac{d_j - \underline{d_j}}{\bar{d_j} - \underline{d_j}}$ for $b \geq 0$ and $g(d_j) = -b(1 - \frac{d_j - \underline{d_j}}{\bar{d_j} - \underline{d_j}})$ for $b < 0$. 95% confidence intervals are reported.

# F Description of the CPS and Foster Care Systems

This section describes the CPS and foster care systems in Michigan, which work similarly to other states. The process begins when someone calls the state's child abuse hotline to report

[47] Strong positive correlation appears unlikely in the current context: We find no evidence that investigators with an above-median comparative advantage in high-risk cases are differentially likely to quit when their caseload includes an above-median share of high-risk cases.

an allegation of child abuse (e.g., bruises or burns) or neglect (e.g., inadequate supervision due to substance abuse). While anyone can call the hotline, the most frequent reporters are those legally required to do so, such as educational personnel (Benson et al., 2022). There are two central hotline call centers in Michigan, one in Detroit and one in Grand Rapids, but they share the same hotline number. When a new call comes in, it is quasi-randomly routed to the screener who has been on queue the longest, with no exceptions. Screeners have discretion on whether to "screen-in" the call: about 60% of all initial calls are screened-in, which launches a formal CPS investigation. A screened-out call concludes CPS involvement.

Once a call is screened-in, the screener transfers all relevant paperwork to the alleged victim's local child welfare office, including the alleged maltreatment type (e.g., physical abuse versus physical neglect), and basic demographics of the child such as age, gender, and race. Each county in Michigan has its own local office and some larger counties can have multiple offices. When the local office receives the report, it assigns the case to a CPS investigator based on a rotational assignment system rather than their particular skill set or characteristics. There are two exceptions to the rotational assignment of investigators, both of which we exclude from the analysis. First, given their sensitivity, reports of sexual abuse tend to be assigned to more experienced investigators. Second, new reports involving a child for whom there was a very recent prior investigation are usually assigned to the original investigator given the investigator's familiarity with the case. Accordingly, we exclude cases involving sexual abuse and those involving children who had been the subject of an investigation in the year before the report.

The investigator has 24 hours to begin an investigation, 72 hours to establish face-to-face contact with the alleged child victim, and 30 days to complete the investigation. The investigator makes two sequential decisions that determine the outcome of the investigation. First, the investigator interviews the people involved, reviews any relevant police or medical reports, and decides whether there is enough evidence to "substantiate" the allegation. In Michigan, 74 percent of investigations were unsubstantiated during our sample period. In these cases, CPS concludes the investigation and there is no further contact with the family.

Conditional on a substantiated investigation, the investigator must also decide whether to temporarily place the child in foster care. Under CPS investigator guidelines in Michigan, the primary justification for foster care placement is a potential for subsequent maltreatment in the home: Investigators are instructed to recommend placement if the child is in imminent danger of maltreatment in the home, but to keep the child with their family otherwise.[48]

[48] As an example, Michigan's Department of Health and Human Services' *Children's Protective Services Policy Manuals* reads: "placement of children out of their homes should occur only if their well-being cannot

While there is technically a standardized 22-question risk assessment that helps the investigator determine whether foster care placement is appropriate, in practice investigators have immense discretion over placement. Many of the questions in the assessment are inherently subjective and previous research suggests that investigators tend to manipulate responses in order to match their priors (Gillingham and Humphreys, 2010; Bosk, 2015).

If the investigator believes there is a potential for subsequent maltreatment in the home, they request to the office's supervisor to file a petition with the local court to temporarily place the child in foster care. In practice, it is rare for either the supervisor or the judge asked to sign the petition to disagree with investigators' recommendations. Regardless of the placement recommendation, investigators can also recommend prevention-focused services. These services range from referrals to food pantries or support groups to substance abuse or parenting classes. Nevertheless, families are usually not mandated by the courts to engage in these services. Previous research conducted in our setting has indicated that the preventive services' impact on subsequent maltreatment within the home and other outcomes is generally small (Baron et al., 2024; Gross and Baron, 2022; Baron and Gross, 2022).

The foster care system in Michigan is similar to the rest of the country. Children are temporarily placed with either an unrelated foster family, relatives, or (in about 10% of cases) in a group home or residential setting. During our sample period, children spend roughly 17 months in foster care on average; most children are reunified with their birth parents once the court decides that the parents have made the necessary changes in their lives to get their children back.

# References

Andrews, Donald and Marcia M. A. Schafgans (1998) "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65 (3), 497–517.

Angelova, Victoria, Will Dobbie, and Cyrstal Yang (2023) "Algorithmic Recommendations and Human Discretion," Working paper.

Arnold, David, Will Dobbie, and Peter Hull (2021) "Measuring Racial Discrimination in Algorithms," *AEA Papers and Proceedings*, 111, 49–54.

Arnold, David, Will S Dobbie, and Peter Hull (2022) "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 112 (9), 2992–3038.

Baron, E. Jason, Joseph J. Doyle, Natalia Emanuel, Peter Hull, and Joseph Ryan (2024) "Discrimination in Multi-Phase Systems: Evidence from Child Protection," *The Quarterly Journal of Economics*, Forthcoming.

be safeguarded with their families" (p.3). It also directs investigators to recommend placement "in situations where the child is unsafe, or when there is resistance to, or failure to benefit from, CPS intervention and that resistance/failure is causing an imminent risk of harm to the child" (p.5).

Baron, E Jason and Max Gross (2022) "Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators," NBER Working Paper 29922.

Benson, Cassandra, Maria D Fitzpatrick, and Samuel Bondurant (2022) "Beyond Reading, Writing, and Arithmetic: The Role of Teachers and Schools in Reporting Child Maltreatment," *Journal of Human Resources*, forthcoming.

Bolton, Patrick and Christopher Harris (1999) "Strategic experimentation," *Econometrica*, 67 (2), 349–374.

Border, Kim C (2013) "Introduction to correspondences."

Bosk, Emily Adlin (2015) *All Unhappy Families: Standardization and Child Welfare Decision-Making* Ph.D. dissertation, University of Michigan.

Chamberlain, Gary (1986) "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32 (2), 189–218.

Gillingham, Philip and Cathy Humphreys (2010) "Child Protection Practitioners and Decision-Making Tools: Observations and Reflections from the Front Line," *British Journal of Social Work*, 40 (8), 2598–2616.

Gross, Max and E. Jason Baron (2022) "Temporary Stays and Persistent Gains: The Causal Effects of Foster Care," *American Economic Journal: Applied Economics*, 14 (2), 170–99.

Hart, Sergiu and Philip J Reny (2015) "Maximal revenue with multiple goods: Nonmonotonicity and other observations," *Theoretical Economics*, 10 (3), 893–922.

Heckman, James (1990) "Varieties of Selection Bias," *American Economic Review*, 80 (2), 313–18.

Kasy, Maximilian and Alexander Teytelboym (2023) "Matching with semi-bandits," *The Econometrics Journal*, 26 (1), 45–66.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018) "Human Decisions and Machine Predictions," *Quarterly Journal of Economics*, 133 (1), 237–293.

Lechicki, Alojzy and Andrzej Spakowski (1985) "A note on intersection of lower semicontinuous multifunctions," *Proceedings of the American Mathematical Society*, 95 (1), 119–122.

Weitzman, Martin (1978) *Optimal search for the best alternative*, 78: Department of Energy.