Mechanism Reform: An Application to Child Welfare*

E. Jason Baron[†] Richard Lombardo[‡] Joseph Ryan[§]
Jeongsoo Suh[¶] Quitzé Valenzuela-Stookey[‖]

April 2024

Abstract

In many market-design applications, a new mechanism is introduced to reform an existing institution. Compared to the design of a mechanism in isolation, the presence of a status-quo system introduces both challenges and opportunities for the designer. We study this problem in the context of reforming the mechanism used to assign Child Protective Services (CPS) investigators to reported cases of child maltreatment in the U.S. CPS investigators make the consequential decision of whether to place a child in foster care when their safety at home is in question. We develop a design framework built on two sets of results: (*i*) an identification strategy that leverages the status-quo random assignment of investigators—along with administrative data on previous assignments and outcomes—to estimate investigator performance; and (*ii*) mechanism-design results allowing us to elicit investigators' preferences and efficiently allocate cases. Our proposed mechanism can be implemented by setting personalized non-linear rates at which each investigator can exchange various types of cases. In a policy simulation, we show that this mechanism reduces the number of investigators' false positives (children placed in foster care who would have been safe in their homes) by 10% while also decreasing false negatives (children left at home who are subsequently maltreated) and overall foster care placements. Importantly, the reforming mechanism is designed so that no investigator is made worse-off relative to the status quo. We show that a naive approach which ignores investigator preference heterogeneity would generate substantial welfare losses for investigators, with potential adverse effects on investigator recruitment and turnover.

*We thank Peter Arcidiacono, Pat Bayer, Sylvain Chassang, Michael Dinerstein, Laura Doval, Joseph Doyle, Federico Echenique, Natalia Emanuel, Maria Fitzpatrick, Ezra Goldstein, Peter Hull, Chris Mills, Matt Pecenco, Katherine Rittenhouse, Alex Teytelboym and seminar participants at the 2024 ASSA Annual Meeting, SAET 2024, UC Berkeley, and Duke University for helpful comments and suggestions.

[†]Duke University and NBER. E-mail: jason.baron@duke.edu.

[‡]Harvard University. Email: richardlombardo@fas.harvard.edu.

[§]University of Michigan. Email: joryan@umich.edu.

[¶]Duke University. Email: jeongsoo.suh@duke.edu.

[‖]UC Berkeley. Email: quitze@berkeley.edu.

I Introduction

The design of mechanisms to optimally allocate tasks among agents is a fundamental question in economics. These types of mechanisms arise in a wide range of high-stakes settings. Examples include the assignment of judges to cases, doctors to patients, and teachers to students. For a market designer interested in reforming such institutions, the presence of a status-quo system presents both opportunities and challenges, relative to the task of designing a mechanism in isolation. On the one hand, the designer may be able to leverage data generated by the existing mechanism, in addition to private information elicited from agents, as part of the design. On the other hand, the existence of the status-quo mechanism may impose political and institutional constraints on the design process. For example, "hold harmless" clauses may prevent the reform from making any agents worse-off relative to the status quo (Dinerstein and Smith, 2021), and agents themselves may resist the reform if they anticipate being made worse-off. This raises a question: how can a designer leverage the data generated by the current institution while also respecting the constraints that it imposes on the proposed alternative? We call this a problem of *mechanism reform*.

We study mechanism reform in the context of the U.S. child protective services (CPS) system. CPS investigators play a crucial role in preventing child maltreatment through the investigation of reported cases of abuse and neglect. 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.

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 (Kim et al., 2017; Wildeman and Waldfogel, 2014). Moreover, foster care placement is one of the most far-reaching government interventions. A growing literature leveraging the rotational assignment of investigators for identification has shown that placement has large effects on children's later-in-life outcomes including criminal justice contact and earnings (see Bald et al. (2022*b*) for a review).

The focus of this paper is on the investigator assignment mechanism itself. We show below that the current system ignores potentially valuable information: data produced by the rotational system allow us to estimate investigators' performance on cases with different observable characteristics. We therefore ask the question: how can this information be leveraged to improve the current system of assigning investigators to cases? The challenge comes from the fact that handling cases is costly for investigators. We must ensure, for reasons detailed below, that no investigators are made worse-off relative to the status-quo mechanism. This creates a non-trivial agency problem since investigators' preferences over caseloads are private information and unobservable to the designer.

The backbone of our mechanism reform approach consists of two sets of results: identification results which allow us to estimate investigator performance (the "output" side of the problem) and mechanism-design results allowing us to elicit investigators' preferences and efficiently allocate cases without making any investigator worse-off (the "input" side of the problem). We describe these two contributions in turn.[1]

The output side in this context concerns the objective function in the mechanism-design problem. Social preferences over assignments admit a utilitarian representation given by the sum of expected child and family welfare in each case, which is in turn a function of investigators' performance on cases as measured by their prediction mistakes (e.g., not placing a child in foster care who would experience subsequent maltreatment in their home). Identifying investigators' performance, however, is complicated by the selective observability of subsequent maltreatment: among children placed in foster care, we do not observe whether they would have experienced subsequent maltreatment in the home. Nonetheless, we provide novel identification results showing that the *relative* performance of any two investigators is identified when cases are quasi-randomly assigned. These relative performance parameters are sufficient to identify social preferences over assignments, and so the data are sufficient to deal with the output side of problem.

A fundamental aspect of this study is the recognition that the mechanism-design problem we consider does not exist in a vacuum: if our proposed mechanism is to replace the existing rotational system, we must contend with the political and institutional constraints that this system imposes. The input side of the problem deals with these constraints. Among CPS policymakers, perhaps the most cited barrier to reforming the current system is the concern that investigators may be made worse-off. Handling cases is costly for investigators—requiring time and energy as well as imposing emotional and psychological

[1] Both sets of results may be of broader interest. The mechanism-design results of Section III apply more generally to problems of allocating agents to tasks, which may or may not be decision problems as in the current context. The identification results, as discussed in Section IV, can be used to study a number of questions related to decision-maker performance and comparative advantage in other settings.

burdens—and this cost is likely to vary both across cases and investigators. If we wish to re-assign cases to investigators based in part on the cases' observable characteristics, we must consider the potential impacts on investigator welfare. Overburdened investigators are likely to quit, and an excessive workload may hinder recruitment in a system already suffering from staff shortages.[2] In response to these considerations, we impose the *status-quo constraint* that no investigator be made worse-off under the new mechanism relative to the current rotational system.

In the language of mechanism design, the problem we face is one of dynamic combinatorial allocation with a type-dependent participation constraint and without transfers, where an investigator's type represents their privately-observed preference over bundles of cases with different observable characteristics. This problem features several well-known technical challenges. To gain tractability, we divide cases into two categories, "high" and "low," and restrict attention to assignment mechanisms that are random conditional on case type. Even under this restriction, the optimal mechanism has not been identified in the literature. Our primary technical contribution is a solution to this mechanism-design problem.

We first solve a version of this problem which is relaxed along two dimensions. The first relaxation is to a static problem, 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. We refer to this relaxed version as the Large-Market Static (LMS) problem. The bulk of our work concerns the characterization of the optimal mechanism in the LMS problem (Theorem 1 and Theorem 2). The solution takes an intuitive and surprisingly simple form. We first endow each investigator with the assignment that they would have received under the status quo of rotational allocation. Investigator j is then presented with two "exchange rates" $p_j^1 < p_j^2$. Each investigator has the option of retaining their status-quo assignment. Alternatively, investigator j can trade their low-type cases for high-type cases at the rate of p_j^1 low-type cases for every high-type case, or they can trade their high-type cases for low-type cases at a rate of of p_j^2 low-type cases per high-type case. The fact that investigators have input into their assignment ensures that the mechanism is responsive to their preferences. Since they can always opt to retain the status quo we guarantee that they are not made worse-off. The two-part personalized

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

pricing scheme is surprisingly all the flexibility that is needed to steer investigators towards the optimal assignment. The prices assigned to each investigator are derived from the social preferences over assignments and the distribution of investigators' preferences.

We then take the optimal mechanism for the LMS problem and convert it into an approximately optimal mechanism for the applied problem of interest. 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 1). We then convert this mechanism into one that works in the dynamic setting, which we refer to as the Small-Market Dynamic (SMD) problem (Proposition 2). 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.

We also address a practical concern regarding the incentives investigators may have to degrade their performance on certain cases in order to receive a more favorable assignment. Fortunately, we show that under our proposed mechanism the scope for such manipulation is limited: at least locally, investigators have incentives to perform well on the cases to which they are assigned (Theorem 3). It turns out that 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 which can help explain to investigators why their caseloads are no longer identical, as under the rotational system.

Finally, in Appendix A.1 we ask whether the data requirements of the proposed mechanism can be further weakened: is it possible to improve upon the rotational system without any a priori knowledge of the distribution of preferences? The answer is negative: no mechanism improves upon the status quo for all preference distributions, and for some distributions the status quo is the unique feasible mechanism (Theorem 4). Nonetheless, solving the mechanism-design problem allows us to demand much less of the data than would be needed to directly estimate preferences. Moreover, we discuss in Appendix A.1 how the proposed mechanism can itself be used to learn about the distribution of preferences.

To quantify the welfare gains of our proposed mechanism, we use an administrative dataset from Michigan containing the universe of child maltreatment investigations in the state from 2008 to 2016. The data include worker assignments, case and child attributes, and the outcome of each investigation: whether the child was placed in foster care and, if not, whether the child experienced a subsequent maltreatment investigation in the home. Our analysis sample consists of 278,089 investigations involving 225,487 children assigned to 783 unique investigators; 3.0% of these investigations result in foster care placement.

We classify cases into high- and low-risk of future maltreatment in the home using a machine learning algorithm and a rich set of case and child characteristics. Leveraging the status-quo, as-if random assignment of investigators to cases, we demonstrate how to non-parametrically identify investigators' relative performance parameters. Our estimation strategy accounts for the fact that investigator assignment is quasi-random only within local CPS offices, and we use a split-sample strategy to mitigate over-fitting concerns by randomly dividing cases into a "training" set (50%) and an "evaluation" set (50%). That is, estimates of investigator performance in the training set are used to develop the assignment mechanism, while the evaluation set is used to simulate the welfare gains from this assignment.

We first present three empirical facts that motivate the use of our proposed mechanism in this setting. First, we document considerable variation in investigator comparative advantage in high-risk cases. Investigators with a comparative advantage score of one standard deviation above the mean achieve a 6% lower false negative rate (the share of children left at home who are subsequently maltreated) and a 21% lower false positive rate (the share of children placed in foster care who would have been safe in their homes) in high-risk cases. Interestingly, these investigators have similar false negative rates and a 44% *higher* false positive rate in low-risk cases, implying significant investigator specialization across cases. Second, we show that most of the variation in comparative advantage (85%) is driven by variation within, rather than across, counties—suggesting that re-assigning investigators only within counties could realistically achieve welfare gains. Third, we provide novel empirical evidence that high-risk cases are significantly more costly to investigators. Specifically, we leverage the fact that, although the composition of caseloads in expectation is equal across investigators within an office due to the status-quo rotational assignment, in practice there are time periods in which some investigators receive larger numbers of high- or low-risk cases by random chance. Using this variation, we show that a one standard deviation increase in the mean predicted risk of an investigator's caseload increases turnover risk by 104%.

Turning to our main policy simulation, we show that assigning investigators to cases according to our proposed mechanism can lower the number of investigators' false positives by 10%, false negatives by 1%, and total foster care placements by 1%. Importantly, the proposed mechanism involves reallocating only existing resources: we impose travel constraints by re-assigning cases to investigators only within counties and we ensure that no investigators are made worse-off. In fact, we show that the proposed mechanism improves investigator welfare relative to the status quo for 38% of investigators in our sample.

Finally, we demonstrate the importance of considering heterogeneity in investigator preferences over case types. Intuitively, an optimal mechanism that ignores investigator preferences would simply allocate high-risk cases to investigators with a comparative advantage in these cases without compensating them for the additional burden. We show that such a mechanism reduces investigator welfare for 30% of investigators relative to a counterfactual that splits cases equally within counties. The investigators with the greatest welfare losses are those with a comparative advantage in high-risk cases. Thus, failure to consider preferences results in large welfare losses for investigators, which could in turn harm recruitment and increase turnover rates in a system already suffering from staff shortages.

I.A Related literature

At a high level, this paper is part of the large market-design literature which combines theory and empirics. Applications include school choice (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu, Pathak and Roth, 2005; Pathak, 2011; Che and Tercieux, 2019; Abdulkadiroğlu et al., 2020); judicial scheduling (Bray et al., 2016); teacher assignment (Combe, Tercieux and Terrier, 2022); course allocation (Budish et al., 2017); organ transplants (Agarwal, Hodgson and Somaini, 2020); the U.S. Army branching process (Greenberg, Pathak and Sönmez, 2023); refugee resettlement (Delacrétaz, Kominers and Teytelboym, 2023); tax collection (Porto, Persico and Sahuguet, 2013; Kapon, Del Carpio and Chassang, 2022); and the allocation of food to food banks (Prendergast, 2022); among others. Similar to the current paper, much of this work exploits randomization inherent in the existing system to estimate structural parameters (e.g., Abdulkadiroğlu et al. (2017); Abdulkadiroğlu et al. (2022)). Within this literature we study a novel problem: the assignment of CPS investigators to cases. As detailed below, our problem presents a number of distinct challenges, both in terms of theoretical mechanism design and identification.

We first outline our relationship to the theoretical mechanism design literature. A key feature of the current paper is the focus on the private information of investigators and the constraints on the mechanism-design problem imposed by the status quo. 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). 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 no investigator is made worse-off relative to the status quo.

In the mechanism design problem, the status-quo constraint is equivalent to a type-dependent participation constraint, which makes this a problem of so-called countervailing incentives, as studied by Lewis and Sappington (1989), Maggi and Rodriguez-Clare (1995), Jullien (2000), among others. At a technical level we deal with this constraint using similar ironing techniques to Dworczak and Muir (2023). However the multi-item, multi-agent allocation problem that we study here has not been addressed in the countervailing incentives literature. In a two-sided matching context Combe, Tercieux and Terrier (2022) study the problem of assigning teachers to schools when teachers have a right to their status-quo assignment, which can be thought of as a form of type-dependent participation constraint. While our setting differs in many ways, one important conceptual difference is that in Combe, Tercieux and Terrier (2022) there is no systematic relationship between preferences and the value of the outside option, whereas here all agents have the same status-quo allocation, but differ in how they value this allocation.

We also relate to the combinatorial allocation literature, which studies the problem of allocating multiple indivisible items among multiple agents with unknown values. Seminal contributions include Gale and Shapley (1962), Shapley and Scarf (1974), and Varian (1973). This literature has focused on identifying mechanisms with various desirable properties, usually including some notion of Pareto efficiency or fairness among agents. In contrast, the current paper is concerned with maximizing a social welfare objective; the welfare of the agents (investigators) enters only as a constraint (see Section III.A for further discussion).

The problem that we ultimately want to solve (SMD) is inherently dynamic, in that cases must be assigned as they arrive, or "online." As with the static problem, we differ from the literature on dynamic combinatorial allocation (Combe, Nora and Tercieux, 2021; Nguyen, Teytelboym and Vardi, 2023) in terms of our objective. There is also a large literature in computer science studying so-called "online matching" problems (Karp, Vazirani and Vazirani, 1990; Mehta et al., 2007; Aggarwal et al., 2011) in which tasks arrive over time and need to be assigned immediately to a compatible agent. While we share the online-matching feature, our problem differs in terms of both the constraints and objective.

This paper is also related to a literature in personnel and labor economics examining ways to improve the performance of the public sector, where common tools to increase performance such as job-testing technologies, performance pay, and promotion incentives are typically unavailable (Hoffman, Kahn and Li, 2018; Bertrand et al., 2020). Similar to the current paper, this literature has studied the implications of mechanisms for allocating public-sector workers across tasks, teams, and employers. Bates et al. (2023), Biasi, Fu and Stromme (2021), and Laverde et al. (2023), for example, examine the allocation of teachers to schools in two-sided teacher labor markets. Bergeron et al. (2022) study the optimal assignment of property tax collectors to teams and neighborhoods. Ba et al. (2022) examine the implications of alternative mechanisms used to assign police officers to Chicago districts. As in many of these studies, the fact that we allow for idiosyncratic and unobservable investigator preferences creates an agency problem. What distinguishes our setting from much of this literature is that the status-quo random assignment of high- and low-type cases makes it difficult to estimate preferences from data, as we do not observe investigators' choices over case types. Instead, our mechanism-design approach elicits preferences directly from investigators and uses this information to inform assignments.

The essential features of the CPS setting are shared with a number of other high-stakes contexts involving examiners. As a result, our framework can be applied in a broad class of settings. Examples include bail judges determining whether to release or detain defendants before their court dates (Arnold, Dobbie and Yang, 2018; Dobbie, Goldin and Yang, 2018; Arnold, Dobbie and Hull, 2022) and radiologists working to diagnose pneumonia (Chan, Gentzkow and Yu, 2022). A key commonality between these settings is that examiners make a binary decision and the socially optimal action depends on the value of some underlying random variable—which realizes partially after one of the actions is taken. Our identification results apply in such settings, provided that assignments to cases are quasi-random. The mechanism-design results are applicable more broadly to any task allocation problem in which performance can be estimated and the designer's objective is linear in the assignment.

Finally, we contribute to the literature examining the efficacy of CPS systems. Most studies in this literature estimate the causal effects of foster care placement on children's and parents' outcomes (Doyle, 2007, 2008; Baron and Gross, 2022; Bald et al., 2022*a*,*b*; Gross and Baron, 2022; Helénsdotter, 2022; Grimon, 2020).[3] Our findings highlight the potential for alternative investigator assignment mechanisms to reduce child maltreatment rates. Moreover, our framework for measuring the performance of investigators could be used to examine other topics in the CPS context; for example, questions related to the recruitment, retention, and training of investigators.

The remainder of the paper is organized as follows. In Section II, we introduce the theoretical framework. Section III contains the central mechanism-design analysis, dealing with the input side of the problem. In Section IV, we prove the identification results which are needed for the output side. Section V discusses the data sources and estimation strategy. Section VI brings the mechanism to data and demonstrates the welfare gains. Section VII concludes, and Appendix A contains additional results and discussion.

[3] A related literature explores the impact of algorithmic decision tools within CPS (Grimon and Mills, 2022; Rittenhouse, Putnam-Hornstein and Vaithianathan, 2022; Fitzpatrick, Sadowski and Wildeman, 2022).

II Framework

We first clarify the assignment problem in a stylized model, and we then turn to the complications which need to be addressed to move towards a real-world mechanism.

II.A A stylized model

Consider a given set of cases $\mathcal{I} = \{1, \dots, I\}$, indexed by i, each of which must be assigned to an investigator from a set $\mathcal{J} = \{1, \dots, J\}$, indexed by j. Suppose that the designer's objective is to minimize social cost, C, given by

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

for some known function $c : \mathcal{I} \times \mathcal{J} \to \mathbb{R}$, where $Z_{ij} = 1$ indicates that investigator j is assigned to case i under assignment Z.

Each case i comes with a "price" $p(i)$, which in this stylized model we assume is symmetric across investigators and known to the designer. Then, the private cost to an investigator j from assignment Z is $\sum_{i \in \mathcal{I}} Z_{ij} p(i)$. Let $Z^{\bullet}$ be some status-quo assignment; e.g., the output of the rotational assignment mechanism. The set of admissible allocations are those which make no investigators worse-off relative to the status quo—i.e., the set of Z such that

$$\sum_{i \in \mathcal{I}} Z_{ij} p(i) \leq \sum_{i \in \mathcal{I}} Z^{\bullet}_{ij} p(i) \quad \forall\, j \in \mathcal{J}.$$

Assume that cases are divided into two types, call them h (high) and l (low). Let I^h and I^l be the set of high- and low-type cases respectively, and let n^h_j, n^l_j be the number of high- and low-type cases assigned to j under the status-quo.

We restrict attention to assignment mechanisms which are measurable with respect to case type. In other words, assignment mechanisms that specify only how many cases of each type go to each investigator, and are random conditional on case type. Then, in this stylized model, the designer solves the linear program:

$$\min_{(\hat{n}^h_j, \hat{n}^l_j)_{j \in \mathcal{J}}} \sum_{j \in \mathcal{J}} \hat{n}^h_j c^h(j) + \hat{n}^l_j c^l(j) \quad s.t. \quad \hat{n}^k_j \geq 0 \quad \forall\, j \in \mathcal{J}, \quad k \in \{h, l\} \tag{M2}$$

$$p\hat{n}^h_j + \hat{n}^l_j \leq p n^h_j + n^l_j \quad \forall\, j \in \mathcal{J}$$

$$\sum_{j \in \mathcal{J}} \hat{n}^k_j \geq |I^k| \quad \text{for } k \in \{h, l\}.$$

where $c^k(j) = \hat{\mathbb{E}}[c(i,j)|i \in I^k]$, $p = \frac{\hat{\mathbb{E}}[p(i)|i \in I^h]}{\hat{\mathbb{E}}[p(i)|i \in I^l]}$, and $\hat{\mathbb{E}}$ denotes the empirical expectation.[4] The solution to this program is simple: Define the *price-weighted relative advantage* of investigator j to be

$$\delta_j(p) = pc^l(j) - c^h(j).$$

The larger is $\delta_j(p)$, the better investigator j is at handling high- relative to low-type cases.

Lemma 1. Any solution to the social cost-minimization program (M2) is defined by a threshold in $\delta_j(p)$. More specifically, we can order investigators such that $\delta_j(p) - \delta_{j+1}(p) \geq 0$ and choose $j^* \in \mathcal{J}$ to be the smallest index such that $\sum_{j \leq j^*} n_j^h + \frac{1}{p} n_j^l \geq |I^h|$. Then, all $j < j^*$ are assigned only high-type cases ($\hat{n}_j^h = n_j^h + \frac{1}{p} n_j^l$), and all $j > j^*$ receive only low-type cases ($\hat{n}_j^l = p n_j^h + n_j^l$), with the remainder of the cases going to j^*.

Proof. Suppose that $\delta_j(p) > \delta_{j'}(p)$ and j has a strictly positive quantity of low-type cases while j' has a strictly positive quantity of high-type cases. Then, trade $\varepsilon > 0$ units of low-type cases from j to j' in exchange for $p\varepsilon$ units of high-type cases. The total reduction in the social cost is $\varepsilon\left(\delta_j(p) - \delta_{j'}(p)\right) > 0$. □

Remark 1. In this problem, there is a clear measure of an investigator's performance: the lower is $c^k(j)$, the better j performs on type-k cases. Moreover, Lemma 1 tells us that that in this problem the right way to measure the comparative advantage of investigator j for high-type cases, relative to j', is $\delta_j(p) - \delta_{j'}(p)$. Importantly, the appropriate comparative advantage metric depends on the constraint structure, as defined by p. We discuss the identification and estimation of performance and comparative advantage in detail in Section II.B.1 and Section V.

II.B Towards a real-world mechanism

To obtain a mechanism that could be brought to the field, we revisit two important simplifying assumptions made in the stylized model. First, we assumed that we could observe the social preferences over assignments. In particular, we assumed that the parameters $c^h(j)$ and $c^l(j)$ were known to the designer. Second, we assumed that each case i imposed the same burden, $p(i)$, on all investigators, and that this parameter was observed by the designer.[5]

[4] If p is an integer, the Birkhoff-von Neumann theorem implies that the extreme points of the set of feasible allocations are integral; i.e., they describe a deterministic assignment of cases to investigators, and so there is a solution to the linear program that is also integral. If p is not an integer, then the solution may require randomization (where we allow the investigators' cost constraints to hold in expectation). Since the stylized model is meant for illustration, we do not delve into these details here.

[5] The stylized model is also static, whereas in reality assignments need to be made dynamically (online). We address this dimension in Section III.E.

II.B.1 The "output" side: defining social preferences

Let Y_i^* be a latent variable equal to 1 if the child in case i would experience subsequent maltreatment if not placed in foster care (i.e., if left in the home), and 0 otherwise. Let D_{ij} be the random variable equal to 1 if investigator j would recommend foster care in case i, and 0 otherwise. There are four potential outcomes if case i were assigned to investigator j: true positives ($D_{ij} = 1$, $Y_i^* = 1$); false positives ($D_{ij} = 1$, $Y_i^\star = 0$); true negatives ($D_{ij} = 0$, $Y_i^\star = 0$); and false negatives ($D_{ij} = 0$, $Y_i^\star = 1$). The joint distribution of investigator j's decision and case i's outcome is described by $(\text{FN}_{ij}, \text{FP}_{ij}, \text{TN}_{ij}, \text{TP}_{ij})$, where $\text{TP}_{ij} = Pr(\{Y_i^\star = 1, D_{ij} = 1\})$, and similarly for the other outcomes. This is a well-known problem of statistical classification.[6]

The joint distribution over an investigator's decision and a case's outcome is the product of a number of factors, including the investigator's own preferences over outcomes, the signals they observe about Y_i^*, and their ability to process this information. We make no attempt to distinguish between these different factors since, as clarified below, what matters for evaluating a mechanism is just the joint distribution that these factors ultimately produce. Similarly, we will talk about measuring an individual investigator's *performance* in terms of outcomes, without attempting to separate whether this performance is produced, for example, by the investigator's information or their preferences.

To characterize social preferences over the set of assignments, denoted by $\mathcal{Z}$, we make three common assumptions generally believed to have normative appeal: *(i) Expected utility.* Social preferences over the uncertain outcomes of an individual case have an expected utility (EU) representation; *(ii) Utilitarianism.* Social preferences over the entire set of outcomes satisfy the axioms of Harsanyi (1955) with respect to the preferences over the outcomes of individual cases; *(iii) Symmetry.* Society is indifferent between switching the outcomes of cases i and i', holding all other outcomes fixed.

It is well known that under these assumptions, we can represent the social preferences over assignments with a utilitarian aggregation of the expected utility from each case. That is, social preferences are represented by minimization of the social cost

$$C(Z) := \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} Z_{ij} \underbrace{\left(\text{FN}_{ij} \cdot c_{FN} + \text{TN}_{ij} \cdot c_{TN} + \text{FP}_{ij} \cdot c_{FP} + \text{TP}_{ij} \cdot c_{TP}\right)}_{c(i,j)}$$

[6] We have assumed that the latent variable of interest, $Y_i^\star$, is binary, and we maintain this assumption throughout. However, this assumption is innocuous. Suppose $Y_i^\star$ takes values in some finite set $\mathcal{X}$. For $X \in \mathcal{X}$ define $PX_{ij} = Pr(\{Y_i^* = x, D_{ij} = 1)$ and $NX_{ij} = Pr(\{Y_i^* = x, D_{ij} = 0)$. A version of the identification lemma (Lemma 2 below) continues to hold, so we can still identify social preferences (see Appendix D). Once social preferences are identified, the mechanism-design exercise is unchanged.

for some parameters $c_{FN}, c_{TN}, c_{FP}, c_{TP}$. Without loss of generality, normalize c_{TN} to zero.[7]

The parameters c_{FN}, c_{FP}, c_{TP} represent the preferences of society over the uncertain outcomes of a given case. Ultimately, these cost parameters must be chosen by policymakers, and are an input into the mechanism-design exercise. We therefore treat c_{FN}, c_{FP}, c_{TP} as known for the purposes of designing the mechanism.[8] On the other hand, the parameters $\text{FN}_{ij}, \text{FP}_{ij}$, and TP_{ij} characterize the joint distribution of investigator decisions and outcomes, and are not directly observed. These parameters are needed to transform the (known) social preferences over the outcomes of each case into social preferences over the entire matrix of *assignments* of cases to investigators, and must be estimated using data on cases and investigators.

The expectations of the cost parameters $c(i, j)$ defining C cannot be non-parametrically identified. This is because the rates of false positives and true positives are fundamentally unobservable. However, we show in Section IV that under the quasi-random assignment of investigators to cases, the data are sufficient to identify $\mathbb{E}[c(i, j) - c(i, j')|i \in I]$ for any $I \subset \mathcal{I}$ such that both j and j' receive cases in I with positive probability (Lemma 2).[9] This is sufficient to identify social preferences over the class of assignment mechanisms that we consider, and so for the purposes of designing the mechanism we can proceed as if $c(i, j)$ is observed (see also Remark 5).

II.B.2 The "input side": investigators' preferences

The input side of the problem concerns the fact that handling cases imposes a burden on investigators. Importantly, not all cases impose the same burden, and investigators may differ in their preferences over cases. To model the input side, we assume that each case i comes with a price, $p_j(i)$, for investigator j. Given an assignment Z, the cost of investigator j's caseload is

$$\sum_{i \in \mathcal{I}} Z_{ij} p_j(i)$$

The cost of j's caseload represents their preferences over assignments. Note that while in reality cases are assigned over time, we treat preferences as static. That is, investigators care only about their total caseload over the specified time horizon. We discuss this assumption

[7]This means that the parameter c_{FN} should be interpreted as the difference in cost between a true negative and a false negative, and similarly for c_{FP} and c_{TP}. If one assumes a true negative is the best possible outcome, then we should have $c_{FN}, c_{FP}, c_{TP} \geq 0$; but we do not impose this assumption.

[8]In the empirical application below, we show robustness to a range of values for these parameters.

[9]Formally, we identify each case with a set of observable characteristics, so $I \subset \mathcal{I}$ refers to a subset of characteristics, and treat $c(i, j)$ as a random variable.

in more detail in Section III.E.[10]

We impose as a constraint on our mechanism that no investigator be made worse-off relative to the status-quo assignment. This constraint implies that the mechanism is revenue neutral, has no negative impact on recruitment and turnover, and should be politically feasible.[11] The difficulty is that investigators' preferences over caseloads are fundamentally unobservable. This is certainly true given the available data, in which cases are randomly assigned to investigators. Moreover, even with richer data it would be challenging to identify investigators' preferences if one wants to allow these to differ across investigators and cases. We see no grounds on which to rule out such heterogeneity ex-ante, and doing so incorrectly could lead to a mechanism which makes investigators worse-off. We therefore design a mechanism which elicits information about preferences directly from investigators.

III A mechanism-design approach to case allocation

The mechanism-design problem that we face can be described as one of dynamic combinatorial allocation with type-dependent participation constraints and no transfers.[12] To gain tractability, we focus primarily on mechanisms which divide the set of cases into two types, "high" and "low," and are random conditional on case type. Even under this assumption, the problem poses a number of difficulties that are well-known to be technically challenging. To address these, we solve three different versions of the problem, which lead us sequentially towards a practical mechanism to bring to the field.

Large Market, Static (LMS). We first study a static version of the assignment problem, in which the goal is to allocate a fixed set of cases among the investigators, and assume that there is a continuum of investigators.[13] In this setting, we derive a simple characterization

[10]We also maintain throughout the linear specification for the cost of an investigator's caseload. In other words, we assume constant marginal production costs. We find this restriction palatable, especially in light of the fact that 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 constant marginal 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. In reality, cases for a given investigator may overlap, so constant marginal costs is not precisely equivalent to time separability in our setting. For further discussion, see Section III.E.

[11]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.

[12]The results of this section are not limited to the assignment of investigators to cases; they apply more broadly to any task-allocation problem with these features.

[13]The latter assumption is equivalent to relaxing the problem so that that every case must be assigned to some investigator ex-ante, taking expectations over investigator types, rather than ex-post—i.e., for every

of the optimal mechanism. The problem and solution are novel in the mechanism-design literature, and may be of independent interest.

Small Market, Static (SMS). We then take the solution to the LMS problem and convert it into a mechanism which satisfies the ex-post feasibility constraint in the small market with a finite set of investigators. The mechanism remains incentive compatible and respects the status-quo constraint for each investigator. It is also approximately optimal, where the approximation is, unsurprisingly, improving in the size of the market and decreasing in the uncertainty about each agent's type.[14]

Small Market, Dynamic (SMD). Finally, we convert our mechanism for the SMS problem into one which accounts for the fact that cases arrive over time, and must be assigned as they come. In going from the static to the dynamic setting we lose exact incentive compatibility and respect for the status quo. These properties are obtained only approximately, where here the approximation improves as the time horizon grows. We validate in the empirical application that the approximation is close.

Our primary technical contribution comes from the design of the mechanism for the LMS problem, while the extensions to the SMS and SMD problems are relatively straightforward. A natural question is why, given that these extensions entail approximation, we do not simply solve the SMD problem directly. There are two reasons for this, discussed in greater detail in Section III.D. First, from the LMS problem we obtain a simple characterization of the optimal mechanism which can be easily explained to agents and implemented in practice. Second, both the SMD and SMS problems involve significant and well-known open technical challenges which, while interesting, are beyond the scope of the current paper to resolve. Before presenting the full mechanism-design analysis, we quickly preview the solution.

III.A Solution preview

A simple and widely-studied class of mechanisms for static assignment problems are those based on competitive equilibrium (CE) (Varian, 1973; Hylland and Zeckhauser, 1979; Budish, 2011; Nguyen and Vohra, 2021). 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

realized type profile.

[14] In the LMS and SMS problems we ignore integer constraints and allow for fractional assignments of cases. We could always induce a given fractional assignment in expectation by randomizing the assignment of cases. We do not even need to appeal to the Birkhoff-von Neumann theorem to show this, since the only hard constraint on the ex-post assignment is that each case be allocated to someone. If investigators are risk neutral, then the IC and status-quo constraints are unaffected. The mechanism that we ultimately propose to implement, the SMD mechanism, respects the integer constraints.

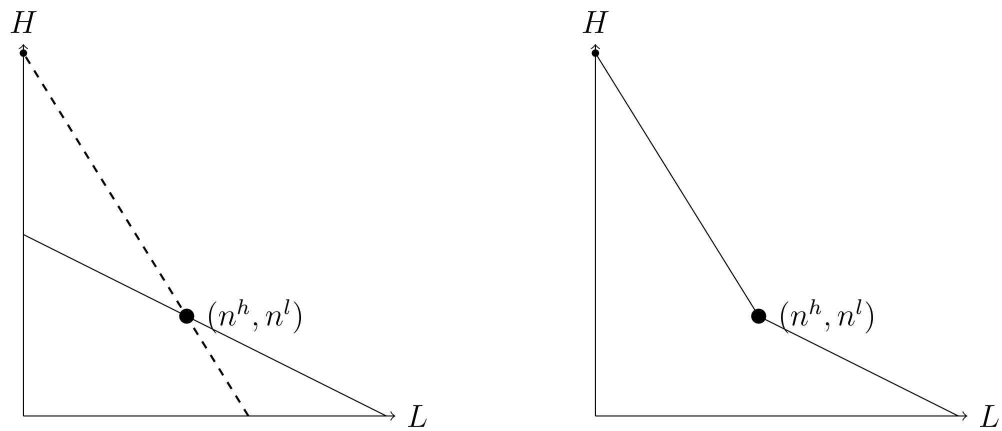
allow investigator j to choose their favorite bundle from the budget set $\{(\hat{n}^h_j, \hat{n}^l_j) : p\hat{n}^h_j + \hat{n}^l_j \geq pn^h + n^l\}$. The price p should be set so that the market clears, i.e. all cases are assigned. Assuming such a market-clearing price exists and that agents behave 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, the status quo is always affordable, so no investigator is worse-off.

The downside of the competitive market mechanism is that while it respects the preferences of investigators, it does not incorporate those of the designer; investigator performance measures do not enter into the construction. 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 should 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 $\{(\hat{n}^h_j, \hat{n}^l_j) : p^j\hat{n}^h_j + \hat{n}^l_j \geq p^j n^h + n^l\}$.

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 vertical axis. An increase in p^j corresponds to a rotation from the solid to the dotted budget set. The higher is p^j, the more attractive it is for j to take on additional high-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 on the one hand, and on the other hand ensuring that j handles their fair share of the overall caseload.

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 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^j_2 low-type cases, while if j wants to trade *for* a high-type case they can give up at most p^j_1 low-type cases, where $p^j_1 < p^j_2$. The induced budget set is depicted in Figure 1b. By increasing p^j_2 above p^j_1 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.

[15] Existence of market clearing price is not guaranteed with indivisible goods. However the mechanism can be well approximated in a way that preserves its desirable efficiency and incentive properties (Budish, 2011; Azevedo and Budish, 2019).



(a) Two one-price budget sets (b) One two-price budget set

Figure 1: Budget sets

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. The remainder of this section is devoted to showing that there exists an optimal mechanism for the LMS problem in which each investigator faces a personalized two-part pricing scheme; we then describe its approximate implementation in the SMS and SMD settings.

III.B Large Market Static

This problem is static in that there is a fixed set of cases, $\mathcal{I}$, to be allocated. It is "large-market" in that we can imagine that each investigator j is actually a unit-mass population of agents. Each agent has a non-negative two-dimensional type $(p_h, p_l) \in [\underline{p}_h, \bar{p}_h] \times [\underline{p}_l, \bar{p}_l]$, and the measure of agents in population j with types below (p_h, p_l) is $F_j(p_h, p_l)$. The designer chooses a mechanism consisting of functions $(H_j, L_j) : [\underline{p}_h, \bar{p}_h] \times [\underline{p}_l, \bar{p}_l] \to \mathbb{R}_+ \times \mathbb{R}_+$ for each $j \in \mathcal{J}$, where $(H_j(p_h, p_l), L_j(p_h, p_l))$ is the number of high- and low-type cases assigned to j when their type is (p_h, p_l). Let $c^k(j) = \hat{\mathbb{E}}[c(i,j)|\text{case } i \text{ is type } k]$ for $k \in \{h, l\}$. The objective is to minimize the expected social cost

$$\sum_{j \in \mathcal{J}} \int c^h(j) H_j(p_h, p_l) + c^l(j) L_j(p_h, p_l) dF_j(p_h, p_l)$$

subject to the incentive compatibility constraint that agents report their type truthfully (discussed below), the status-quo constraint, and feasibility. Let n^k be the number of type-k cases per investigator-population (recall that there are J investigators). Then, the status-quo constraint is that no agent be made worse-off relative to an equal split of cases:[16]

$$p_h H_j(p_h, p_l) + p_l L_j(p_h, p_l) \leq p_l n^l + p_h n^h \quad \forall\ j \in \mathcal{J},\ \text{and } (p_l, p_h) \in [\underline{p}_h, \bar{p}_h] \times [\underline{p}_l, \bar{p}_l]$$

and the market clearing constraints are:

$$\sum_{j \in \mathcal{J}} \int H_j(p_h, p_l) dF_j(p_h, p_l) = Jn^h \quad \text{and} \quad \sum_{j \in \mathcal{J}} \int L_j(p_h, p_l) dF_j(p_h, p_l) = Jn^l.$$

It is easy to see that this is equivalent to a model in which each investigator is just a single agent, but we only require the markets to clear in expectation when each agent's type is drawn according to F_j.[17]

We can divide this problem into two parts. First, we study an "inner" problem in which we characterize the set $\mathcal{F}_j$ of caseloads (high-type case count and low-type case count) that can be allocated to investigator-population j in expectation, using some mechanism. The second piece of the analysis is the "outer" problem in which we optimally spread the expected caseloads across investigator-populations by choosing a point in $\mathcal{F}_j$ for each j.

III.B.1 Step 1: LMS inner problem

In the inner problem, we study the design of a mechanism for a single investigator-population, and so drop the dependence on j in the notation. A mechanism (H, L) is incentive compatible—i.e., truthful reporting of their type is a best response for all agents—if and only if

$$\begin{aligned} p_h H(p_h, p_l) + p_l L(p_h, p_l) \leq p_h H(p'_h, p'_l) + p_l L(p'_h, p'_l) \\ \forall\ (p_h, p_l), (p'_h, p'_l) \in [\underline{p}_h, \bar{p}_h] \times [\underline{p}_l, \bar{p}_l]. \end{aligned} \tag{IC'}$$

First, observe that no incentive compatible mechanism will be able to distinguish between types $(p_h, p_l), (p'_h, p'_l)$ such that $\frac{p_h}{p_l} = \frac{p'_h}{p'_l}$. These agents have identical preferences over assignments, and so will make identical choices. Thus, it is without loss of generality to consider mechanisms which elicit only the relative price of high-type cases, $p := \frac{p_h}{p_l}$.[18] We henceforth refer to this ratio as the agent's type, and write mechanisms simply as a function

[16] Of course, we do not observe an exactly equal split of cases among investigators in the data. However, investigators have equal expected caseloads, and the differences are negligible over a long time horizon within CPS offices.

[17] The large-market assumption is also what lets us write the problem in terms of the interim allocation rules $(H_j(p_h, p_l), L_j(p_h, p_l))$, as opposed to the allocation rule which maps type *profiles* to assignments.

[18] 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, Kominers and Akbarpour (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.

of the one-dimensional type. We maintain the following assumption on types.

Assumption. $p := \frac{p_h}{p_l}$ has a full-support distribution on a bounded interval $[\underline{p}, \bar{p}]$ with absolutely continuous CDF F and density f.

Say that a caseload $(\hat{n}^h, \hat{n}^l)$ is *incentive-feasible* if it is the expected caseload induced by an admissible mechanism; that is, if there exists a mechanism $(H, L) : [\underline{p}, \bar{p}] \to \mathbb{R}$ such that

$$-pH(p) - L(p) \geq -pH(p') - L(p') \quad \forall\, p, p' \in [\underline{p}, \bar{p}] \tag{IC}$$

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

$$\int H(p) dF(p) \geq \hat{n}^h \tag{h-capacity}$$

$$\int L(p) dF(p) \geq \hat{n}^l \tag{l-capacity}$$

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

Our goal in the inner problem is to characterize the set $\mathcal{F} \subset \mathbb{R}^2_+$ of incentive-feasible pairs. Observe that the set $\mathcal{F}$ is convex. This follows from the fact that the set of IC and IR mechanisms is convex. So if (H, L) implements $(\hat{n}^h_1, \hat{n}^l_1)$ and (H', L') implements $(\hat{n}^h_2, \hat{n}^l_2)$ then $(\alpha H + (1-\alpha)H', \alpha L + (1-\alpha)L')$ implements $(\alpha\hat{n}^h_1 + (1-\alpha)\hat{n}^h_2, \alpha\hat{n}^l_1 + (1-\alpha)\hat{n}^l_2)$. For the same reason $\mathcal{F}$ is also closed under downward scaling: if $(\hat{n}^h, \hat{n}^l) \in \mathcal{F}$ then so is $(\alpha\hat{n}^h, \alpha\hat{n}^l)$ for all $\alpha \in [0, 1]$.

There are a number of ways we could characterize the convex set $\mathcal{F}$. For solving the outer problem, it is convenient to do so by characterizing the support function of $\mathcal{F}$: the function $S : \mathbb{R}^2 \to \mathbb{R}_+$ defined by

$$S(a, b) = \max\left\{a\hat{n}^h + b\hat{n}^l \;:\; (\hat{n}^h, \hat{n}^l) \in \mathcal{F}\right\}.$$

Let $N^*(a, b) := \arg\max\{a\hat{n}^h + b\hat{n}^l : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}\}$. The support function is convex and continuous. Once we have identified S we can obtain the dual representation of $\mathcal{F}$:

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

To characterize the support function we need to maximize linear functions on $\mathcal{F}$. We do this by maximizing directly over IR and IC mechanisms: for arbitrary $(a, b) \in \mathbb{R}^2_+$, we solve

$$\max_{H,L} \quad a\int H(p)dF(p) + b\int L(p)dF(p) \tag{1}$$

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

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

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

Notice 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 a classical monopoly pricing problem as in Mussa and Rosen (1978), as well as recent papers which consider more general designer objectives, such as Akbarpour, Dworczak and Kominers (2023). Relative to this work, the two distinctive features of the LMS within program in eq. (1) are *(i)* a non-negativity constraint on L, and *(ii)* a type-dependent participation constraint determined by the need to respect the status quo.

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 to characterize the optimal mechanism. The main challenge lies in identifying for which types the IR constraint should bind. As in Jullien (2000), it turns out that the status-quo constraint binds for an intermediate interval of types.

Theorem 1. 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 B.1. □

Theorem 1 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 assumption that ϕ is increasing it is without loss to consider only two-part mechanisms.

Corollary 1. If the virtual value $\phi(p) = p - \frac{1-F(p)}{f(p)}$ is increasing, then for any (a, b) 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 ϕ is strictly increasing, then the optimal mechanism is unique (up to zero-measure perturbations).

Proof. Proof in Appendix B.1. □

Direct implementation

The direct implementation of this mechanism is easily interpretable. We can imagine that agents are granted an endowment (n^h, n^l) of cases. Agents can reduce their type-l assignment by exchanging type-h cases for type-l cases at a "rate" of p_3. Agents who want to reduce their type-l assignment can exchange type-l cases for type-h cases at a rate of p_2, as long as their type-h assignment remains below H_2. Those who want to further reduce their type l assignment can do so, but face a lower exchange rate p_1. If the virtual value ϕ is increasing, then the direct implementation is even simpler. We just need to choose a price p_1 at which agents can "buy" type-h cases by "selling" type-l cases, and a higher price p_2 at which they can sell type-h cases and buy type-l cases. The resulting budget set is depicted in Figure 1b.

Characterizing $\mathcal{F}$

Theorem 1 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 $N^*(a, b)$. Define the *efficient frontier* to be the set $\{(\hat{n}^h, \hat{n}^l) \in \mathbb{R}^2_+ : a\hat{n}^h + b\hat{n}^l =$

$S(a,b),\ (a,b) \in \mathbb{R}^2_+\}$. The set $\mathcal{F}$ is just the subset of the positive orthant that lies within the efficient frontier.

Abusing terminology, we say that $\mathcal{F}$ is *strictly convex* if the mixture of any two points on the efficient frontier lies in the interior of $\mathcal{F}$. Equivalently, the support function is strictly convex at any (a,b) such that $\arg\max\{a\hat{n}^h + b\hat{n}^l : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}\}$ is strictly positive.

Corollary 2. If F is strictly regular $\mathcal{F}$ is strictly convex.

Proof. Proof in Appendix B.2. □

III.B.2 Step 2: LMS outer problem

Theorem 1 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 now 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 investigator j is an element of $\mathcal{F}_j$.[19] That is, we want to solve

$$\min_{(\hat{n}^h_j, \hat{n}^l_j)_{j=1}^J} \sum_{j=1}^J c^h(j)\hat{n}^h_j + c^l(j)\hat{n}^l_j \tag{2}$$

$$s.t. \quad \sum_{j=1}^J \hat{n}^h_j \geq Jn^h \qquad (h\text{-feasible})$$

$$\sum_{j=1}^J \hat{n}^l_j \geq Jn^l \qquad (l\text{-feasible})$$

$$(\hat{n}^h_j, \hat{n}^l_j) \in \mathcal{F}_j \quad \forall\ 1 \leq j \leq J \qquad \text{(incentive feasible)}$$

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 distribution over $\mathcal{F}$ which averages to (n^h, n^l), as illustrated in Figure 2, where $\mathcal{F}$ is depicted as the shaded region.

[19] An interesting subtlety can arise when F is not regular and the efficient frontier has linear segments. Theorem 1 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 efficient 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 efficient frontier, it may in fact be necessary to use a mechanism that takes five distinct values. The details are omitted since in what follows we assume that regularity is satisfied.

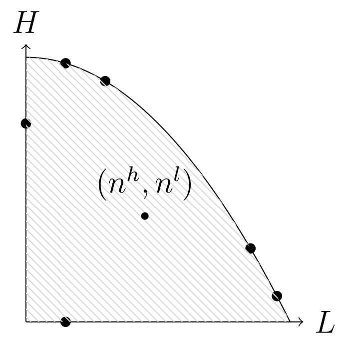


Figure 2: Outer problem with identical type distributions

We solve the outer problem by studying its dual. Let λ_h, λ_l be the dual variables corresponding to the h-feasibility and l-feasibility constraints. Then, the previous program is equivalent to

$$\min_{(\hat{n}_j^h, \hat{n}_j^l)_{j=1}^J} \max_{\lambda_h, \lambda_l \geq 0} \sum_{j=1}^{J} c^h(j)\hat{n}_j^h + c^l(j)\hat{n}_j^l + \lambda_h \left(Jn^h - \sum_{j=1}^{J} \hat{n}_j^h \right) + \lambda_l \left(Jn^l - \sum_{j=1}^{J} \hat{n}_j^l \right)$$

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

Strong duality holds, so the previous program is equivalent to

$$\max_{\lambda_h, \lambda_l \geq 0} \; \lambda_h J n^h + \lambda_l J n^l - \sum_{j=1}^{J} \max \left\{ \left(\lambda_h - c^h(j)\right) \hat{n}^h + \left(\lambda_l - c^l(j)\right) \hat{n}^l \quad : (\hat{n}^h, \hat{n}^l) \in \mathcal{F}_j \right\}$$

In words, the dual variables λ_h, λ_l are the average social costs among high- and low-type cases, respectively. Fix some choice of λ_h, λ_l. Then, the dual program says that each agent should be assigned cases so as to reduce the total social cost as much as possible, given that the "current" average social cost within each group is λ_h, λ_l. We can imagine this program as rewarding each agent a "prize" λ_k for each type-k case that they take on, while charging agent j a fee $c^k(j)$ in order to do so, and allowing agent j to choose from the budget set $\mathcal{F}_j$. We then need to find the values of λ_h, λ_l so that the market clears.

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

$$\max_{\lambda_h, \lambda_l \geq 0} \; \lambda_h J n^h + \lambda_l J n^l - \sum_{j=1}^{J} S^j \left(\left(\lambda_h - c^h(j)\right), \left(\lambda_l - c^l(j)\right) \right) \tag{3}$$

The support function for each j is convex. Thus, the objective in (3) is concave in (λ_h, λ_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 2. Let (λ_h, λ_l) solve the dual program in (3). Then, there exist selections from $N_j^*\left(\lambda_h - c^h(j), \lambda_l - c^l(j)\right)$ such that

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

and these constitute a solution to the social-cost minimization program. In particular, each investigator j receives a caseload on the boundary of $\mathcal{F}_j$. If there are no two investigators j, j' such that $c^k(j) = c^k(j')$ for some $k \in \{h, l\}$, then in any solution at most two investigators have non-zero allocations that are off of the efficient frontier. If in addition every investigator has a strictly regular type distribution, then there is a unique solution to the social-cost minimization problem. In particular, there is a unique solution, (λ^h, λ^l), to the dual, and $N_j^*\left(\lambda_h - c^h(j), \lambda_l - c^l(j)\right)$ is single-valued for all j whose assignment is on the efficient frontier.

Proof. Proof in Appendix B.3. □

As discussed above, if each F_j is strictly regular, then the optimal mechanism is implemented by giving each investigator a pair of prices (p_1^j, p_2^j) and allowing them to buy and sell given their induced kinked budget set. The dual variables λ_h, λ_l determine precisely which prices each investigator gets. We refer to the optimal mechanism under regularity as the *LMS two-price (LMS-TP) mechanism*, and focus primarily on this case.[20]

Investigators off their efficient frontier receive cases of at most one type. We refer to these as *remedial* agents. For such agents, it is as if we set $p_1^j = p_2^j = 1$ if j is to receive only low-type cases, and $p_1^j = p_2^j = \bar{p}$ if j is to receive only high-type cases. The only difference is that j does not need to exhaust their budget: they are allowed to choose $\hat{n}_j^h < n^h + n^l$ high-type cases in the case of $p_1^j = p_2^j = 1$, or $\hat{n}_j^l < \bar{p}n^h + n^l$ low-type cases in the case of $p_1^j = p_2^j = \bar{p}$.

Remark 2. 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 $\max\{y^*\}$. Having access to the derivative is useful for efficiently solving the dual program computationally.

[20]Note that if we are misspecified regarding the type distribution, for example because we impose regularity when the true type distribution is not regular, the solutions to the SMS and SMD problems, which we derive from the solution to the LMS problem as described below, may be sub-optimal but remain feasible and retain their incentive compatibility and status-quo-respecting properties.

From a computational perspective, Theorem 2 allows us to easily solve for the optimal mechanism. Moreover, this formulation of the outer problem allows us to perform comparative statics and answer questions related to the robustness properties of the mechanism, which we do below.

III.C Fairness and incentives for effort

Throughout we treat $c^h(j)$ 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 in the current mechanism if they expect their performance data to be used in the future to re-design the mechanism. We might therefore be concerned about the long-run implications of a mechanism which rewards low-performing workers with lower caseloads. Fortunately, we show that in the LMS-TP mechanism the scope for manipulation of this type is limited.

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

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 $(c^h(j), c^l(j))_{j\in J}$. In the second period, the outcomes from the first period are used to re-estimate the performance parameters, 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, which would be the full obedience condition for the mechanism. 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 gains from performance degradation.

A closely related condition concerns the fairness of a mechanism. For example, if j saw that j' was receiving more favorable exchange rates for high-type cases, one might worry that j would be discouraged about their own performance on high-type cases. However, these exchange rates could also be consistent with investigator j' performing poorly on low-type cases. In general, 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. However even if performance were observable and investigators' exchange rates were public information, it turns out that the mechanism satisfies an additional notion of fairness.

To see this, 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). Agent j's envy is *justified* if moreover (i) $F_j = F_{j'}$, and (ii) $H^j(p_j) > 0$ (resp. $L^j(p_j) > 0$) implies $c^h(j) < c^h(j')$ (resp. $c^l(j) < c^l(j')$); and $H^j(p_j) = 0$ (resp. $L^j(p_j) = 0$) implies $c^h(j) = c^h(j')$ (resp. $c^l(j) = c^l(j')$). That is, j is better than j' for all case-types that j is asked to do, and performs the same as j' on other cases.[21]

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 3. Assume F_j is regular and $pf_j(p) \geq \max\{F_j(p), 1 - F_j(p)\}$. Then, for each agent j, the optimal LMS-TP mechanism is locally fair for all types of j. Moreover, for all but at most two agents, the optimal LMS-TP mechanism is locally effort-inducing.

Proof. Proof in Appendix B.4. □

The only agents for whom the mechanism may not be locally effort-inducing are the remedial agents who are off the frontier. The proof of Theorem 3 is based on establishing comparative statics for the dual program in eq. (3) as a function of the performance parameters, which may be of independent interest. The result depends on restrictions on the type distributions.

[21] 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.

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, as described below, will be sub-optimal, but they remain feasible, IC, and IR.

The basic intuition behind Theorem 3 is the following. Consider 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. 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 receives more than the average number of high-type cases. There is an inescapable trade-off between these two options: they are both determined by the price p_1^j: as we try to assign more high-type cases (by lowering p_1^j) we induce some intermediate types to choose to retain the status quo.[22] It turns out that under the assumptions of Theorem 3, this trade-off is resolved in favor of increasing the probability that the investigator specializes in high-type cases, by giving them more favorable terms for doing so. Thus, by improving their performance on such cases the investigator gets even better terms. This is the comparative static underlying both the effort-inducing and fairness conclusions of Theorem 3.

Remark 3. Theorem 3 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. As Theorem 3 is derived from comparative statics results on these prices with respect to the cost terms $c^k(j)$, the proof tells us how changes in the cost terms will translate into changes in the prices used in the SMS and SMD mechanisms.

III.D Small market static

In the previous section, we solved a relaxed problem in which we only required that all cases be assigned in expectation. This allowed us to design an allocation rule for each agent which depends only on their type, and not the types of other agents. 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. To adopt the same approach of designing each agent's *interim* allocation rule, i.e. depending only on their own type, we would need

[22] The intuition here is incomplete, since we also need to consider changes in p_2^j. The proof in Appendix B.4 deals with this additional complication.

to guarantee that the collection of interim allocation rules for each agent can in fact be induced by some mechanism. This is a well-known problem in mechanism design, beginning with the work of Matthews (1984) and Border (1991). However, existing characterizations do not apply in the current multi-item allocation setting, for reasons explored in Gopalan, Nisan and Roughgarden (2018) and Valenzuela-Stookey (2023), among others. Even with a suitable characterization of the interim allocation rules which would allow us to obtain computational solutions for the optimal mechanism, it is unlikely that we would be able to describe the optimal mechanism in closed form. The ability to describe the mechanism in simple terms is desirable from a policy perspective.

Given the difficulties with finding the optimal mechanism in the SMS setting, our approach is to find a mechanism that is simple and tractable, but only approximately optimal. To do this we modify the LMS solution. Assume that all agents' type distributions satisfy Myerson regularity, and let $(p_1^j, p_2^j)_{j=1}^J$ be the prices defining the optimal LMS-TP mechanism.[23]

Let $P = (p_j)_{j=1}^J$ be a type profile. Fixing the mechanism, investigator j 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. Consider the following allocation.

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

- Assign (n^h, n^l) to all agents not in $\mathcal{B}$ or $\mathcal{S}$.
- There are $|\mathcal{B} \cup \mathcal{S}| n^k$ cases remaining of each type $k \in \{h, l\}$. To assign these cases, we solve the linear program:

$$\min_{\{b_j\}_{j\in\mathcal{B}}, \{s_j\}_{j\in\mathcal{S}}} \sum_{j\in\mathcal{B}} (n^h + b_j) c^h(j) + (n^l - p_1^j b_j) c^l(j) + \sum_{j\in\mathcal{S}} (n^h - s_j) c^h(j) + (n^l + p_2^j s_j) c^l(j)$$

$$s.t. \quad 0 \leq b_j \leq \frac{n^l}{p_1^j} \quad \forall\, j \in \mathcal{B}, \qquad 0 \leq s_j \leq n^h \quad \forall\, j \in \mathcal{S}$$

$$\sum_{j\in\mathcal{B}} b_j = \sum_{j\in\mathcal{S}} s_j \qquad (h\text{-capacity})$$

$$\sum_{j\in\mathcal{B}} p_1^j b_j = \sum_{j\in\mathcal{S}} p_2^j s_j \qquad (l\text{-capacity})$$

  Given solutions (b^*, s^*), the assignment of $j \in \mathcal{B}$ is $(n^h + b_j^*, n^l - b_j p_1^j)$ and of $j \in \mathcal{S}$ is $(n^h - s_j^*, n^l + p_2^j s_j^*)$.

[23] We can apply the same approach without regularity, but the resulting mechanism is more complicated.

Remark 4. The objective above can be replaced with $\sum_{j\in\mathcal{B}} b^j \left(c^h(j) - p_1^j c^l(j)\right) - \sum_{j\in\mathcal{S}} s^j \left(c^h(j) - p_2^j c^l(j)\right)$.

To understand the performance of the SMS-TP mechanism, we 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 1. Assume F_j satisfies strict regularity for all $j \in \mathcal{J}$. In the small-market-static setting, SMS-TP is strategy-proof and respects the status-quo. Moreover, $V_{SMS}(\{F\}_{j=1}^J|y)$ 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 B.6. □

III.E Small market dynamic

Ultimately, the setting we are interested in has a finite number of agents, and 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 is the time horizon.

The dynamic model is as follows. Time is continuous and runs from 0 to T.[24] High- and low-type cases arrive at Poisson rates ρ^h and ρ^l respectively. 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

[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.

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 letting $n^k = \frac{T}{J}\rho^k$ for $k \in \{h, l\}$. This is the expected number of type-k cases per investigator that will arrive by time T. 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 $\hat{n}_j^k(t) := \sum_{i=1}^t \mathbb{1}[z_t = j, \tau_t = k]$. Define the *score*

$$r_j(t,k) = \frac{\hat{n}_j^k(t)}{\dot{n}_j^k}$$

(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 $r_j(t, z_t)$ (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 static SMS-TP mechanism given the aggregate case counts from sequence A over time horizon T. Say that a mechanism is ε-IC (ε-IR) 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 $1 - \varepsilon$.

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 is able to do this depends on how far $N^h(T)$ and $N^l(T)$ are from their expected values $\rho^h T$ and $\rho^l T$. Unsurprisingly, 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 $T \to \infty$.[26]

Proposition 2. $\frac{V_{SD}((F_j)_{j=1}^J, A, T)}{V_{SMS}((F_j)_{j=1}^J, A, T)} \xrightarrow{a.s.} 1$ as $T \to \infty$. Moreover, for any $\varepsilon > 0$ there exists $\bar{T}$ such that the SMD-TP mechanism is ε-IC and ε-IR 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 of the simple dynamic algorithm, 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 ε-IC for large enough T. Note also that $\frac{1}{T} V_{SD}((F_j)_{j=1}^J, A, T)$ 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

[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] Taking $T \to \infty$ is isomorphic to increasing ρ^h and ρ^l.

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 a 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 E we discuss finite-sample adjustments to the assignment rule which move between these extremes.

III.F Extensions and additional properties

Appendix A contains additional results and further discussion of the proposed mechanism. We have assumed, as is standard in the Bayesian mechanism-design literature, that the designer knows the distribution of each agent's preference type. We first ask whether the data requirements of the proposed mechanism can be weakened: is it possible to improve upon the rotational system without assuming knowledge of the type distributions? The answer is no: for any IC mechanism that respects the status-quo welfare constraint for investigators, but is not the same as the status quo, there exists a type distribution under which the mechanism has the same expected social cost as the status quo. Moreover, there are type distributions for which the only feasible mechanism is the status quo. Thus, the only way to guarantee welfare gains is to begin with some information about the type distributions. We then discuss how the mechanism can be used to learn about the type distributions. Following an initial estimate of these distributions, informed by surveys of investigators, a trial run of the mechanism generates individual-level data about preferences which can be used to further refine our estimate of the preference distribution. We also discuss how the mechanism can be used to update the estimates of the cost parameters $c^k(j)$. Finally, we discuss extensions to mechanisms which condition on a richer partition of the set of cases, beyond the binary partition studied here.

IV Identifying social preferences

So far, we have assumed that the designer observes investigator cost parameters, $c^k(j)$. However, even under the random assignment of investigators to cases, $c^k(j)$ is not non-parametrically identified, as true positive and false positive rates are unobserved. This section discusses how we instead identify *differences* in social costs between any given investigator and a benchmark investigator whose cases are drawn from the same population. Since these differences are sufficient to identify social preferences over mechanisms, we are able to treat $c^k(j)$ as observed when designing the mechanism.

The following discussion assumes investigators are quasi-randomly assigned to cases. We discuss in Section V how we account for the fact that, in practice, investigators are conditionally randomly assigned to cases via a rotational system within offices. This assumption has been extensively probed in the Michigan CPS context. For example, Baron and Gross (2022); Gross and Baron (2022); Baron et al. (2024) show that, within zip code by years, a rich set of child and investigation characteristics do not jointly predict the placement tendencies of the investigator assigned to the case.[27] Our approach below also assumes an implicit exclusion restriction: that investigators can only influence children's potential outcomes via their placement decision. This assumption, too, has been probed extensively in our context in this previous work.

The data consist of an observed assignment of investigators to cases. Recall that $D_{ij} \in \{0, 1\}$ represents the *potential* decision of investigator j for case i, where $D_{ij} = 1$ if investigator j would recommend that the child involved in case i be placed in foster care. $Y_i^\star$ captures the child's future maltreatment potential, where $Y_i^\star = 1$ implies that the child would face subsequent maltreatment if left at home. Then, the *potential* outcome for subsequent maltreatment if case i were assigned to j is $Y_{ij} := (1 - D_{ij})Y_i^\star$.[28]

As is common in examiner settings, $Y_i^\star$ is selectively observed based on the assigned investigator and their potential decision. The core selection problem is that we observe $Y_i^\star$ if and only if case i is assigned to an investigator j satisfying $D_{ij} = 0$. If $D_{ij} = 1$ then $Y_{ij} = 0$ regardless of $Y_i^\star$, so that we cannot observe subsequent maltreatment potential.

Ideally we would like to identify $\text{FP}_{ij}, \text{TP}_{ij}, \text{FN}_{ij}$, and, consequently, the social cost of assigning investigator j to case i: $c(i, j) = \text{FN}_{ij} \cdot c_{FN} + \text{TN}_{ij} \cdot c_{FP} + \text{TP}_{ij} \cdot c_{TP}$ (recall that c_{TN} is normalized to zero). Since Y_i^* is observed when i is not treated, FN_{ij} is identified under random assignment. However, if a child is placed in foster care we cannot observe what would have happened had they been left at home, so FP_{ij} and TP_{ij} are not non-parametrically identified. Fortunately, while we cannot identify the social cost function $c(i, j)$ without further assumptions, we next provide identification results which demonstrate how one can

[27]See Appendix F for a detailed discussion of the CPS and foster care processes.

[28]Our benchmark estimates define $Y_i^\star$ as an indicator for whether the child would experience another CPS investigation within six months of the focal investigation if left at home. Because the average stay in foster care in our data is 17 months long and very few children return home within six months, it is usually impossible to directly observe whether a child who has been placed in foster care is maltreated at home within six months.

identify *social preferences*, i.e., the ranking over the set $\mathcal{Z}$ of possible assignments.

To see this, let $I \subset \mathcal{I}$ be a subset of cases. Let Z be the observed assignment, treated here as a random variable. We say that 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 j's assignment is *supported on* I if $Pr(\{i \in I, Z_{ij} = 1\}) \neq 0$. We wish to identify $\text{FP}_j^I := \mathbb{E}[\text{FP}_{ij} | i \in I]$ and $\text{TP}_j^I := \mathbb{E}[\text{TP}_{ij} | i \in I]$.

Lemma 2. 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 in false positive rates $\text{FP}_j^I - \text{FP}_{j'}^I$,
- the difference in true positive rates $\text{TP}_j^I - \text{TP}_{j'}^I$,
- the cost difference $\mathbb{E}[c(i, j) - c(i, j') | i \in I]$.

Proof. First, recall that $Y_{ij} = Y_i^\star(1 - D_{ij})$ 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 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, FN_j^I, TN_j^I, and P_j^I are identified as long as j's assignment is supported on I. We cannot non-parametrically identify TP_j^I or FP_j^I as Y_i^* is unobserved when $D_{ij} = 1$; we only know that $\text{TP}_j^I, \text{FP}_j^I \in [0, P_j^I]$ and $\text{TP}_j^I + \text{FP}_j^I = P_j^I$.

Define $S_j^I = \text{TP}_j^I + \text{FN}_j^I$. Under random assignment conditional on I, $S_j^I = S_{j'}^I = \mathbb{E}[Y_i^\star | i \in I]$ for all $j, j' \in \mathcal{J}$. In words, the share of any investigator's caseload with future maltreatment potential is equal across investigators since cases are randomly assigned. Then,

$$\begin{aligned}
\mathrm{FP}^I_j - \mathrm{FP}^I_{j'} &= \left(1 - \mathrm{TP}^I_j - \mathrm{FN}^I_j - \mathrm{TN}^I_j\right) - \left(1 - \mathrm{TP}^I_{j'} - \mathrm{FN}^I_{j'} - \mathrm{TN}^I_{j'}\right) \\
&= \left(1 - S^I_j - \mathrm{TN}^I_j\right) - \left(1 - S^I_{j'} - \mathrm{TN}^I_{j'}\right) \\
&= -\left(\mathrm{TN}^I_j - \mathrm{TN}^I_{j'}\right) \\
&= \left(P^I_j + \mathrm{FN}^I_j\right) - \left(P^I_{j'} + \mathrm{FN}^I_{j'}\right).
\end{aligned}$$

Similarly, $\mathrm{TP}^I_j - \mathrm{TP}^I_{j'} = \left(P^I_j - \mathrm{FP}^I_j\right) - \left(P^I_{j'} - \mathrm{FP}^I_{j'}\right) = -\left(\mathrm{FN}^I_j - \mathrm{FN}^I_{j'}\right)$. This is sufficient to identify the cost differences as well. □

Lemma 9, presented in Appendix D, generalizes Lemma 2 beyond binary Y_i^*.

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.

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.

In other words, under the conditions of Corollary 3, the cardinal ranking over $\mathcal{Z}$ is non-parametrically identified. Corollary 4 below gives a simple expression for the difference $\mathbb{E}[C(Z)] - \mathbb{E}[C(Z')]$. Note that from Lemma 2 we can also identify the expected difference in false negatives, false positives, and the placement rate across the two assignments.

Remark 5. One useful application of Lemma 2 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 second 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.

IV.A Performance and comparative advantage

The parameter $c^k(j) := \mathbb{E}[c(i,j)|i \in I_k]$ is a natural measure of the performance of investigator j on cases of type k, which may be of independent interest. Note that this performance measure is not non-parametrically identified, for the reasons discussed above. By Lemma 2, the difference $c^k(j) - c^k(j')$ is identified for any j, j', and is given by

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

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

Corollary 4. Let Z and Z' be assignments that are conditionally random given a partition $\{I_k\}_{k=1}^K$ or $\mathcal{I}$. Then

$$\mathbb{E}[C(Z) - C(Z')] = \sum_{k=1}^K \sum_{j\in\mathcal{J}} \gamma_j^k \sum_{i\in I_k} Z_{ij} - \sum_{k=1}^K \sum_{j\in\mathcal{J}} \gamma_j^k \sum_{i\in I_k} Z'_{ij}.$$

As described in Lemma 1, we are also interested in an investigator's *relative advantage* of investigator j for type-h versus type-l cases, given by $c^l(j) - c^h(j) = \delta_j(1)$. While relative advantage is not identified, differences in relative advantage, or the *comparative advantage* of j relative to j', is identified: $D(j,j') = \delta_j(1) - \delta_{j'}(1) = c^l(j) - c^l(j') - \left(c^h(j) - c^h(j')\right)$. When high- and low-type cases are equally costly for all investigators, comparative advantage is the sufficient statistic for the optimal assignment: if $D(j,j') > 0$ then j should be assigned to high-risk cases only if j' is.[29] 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 or comparative advantage.

More generally, define the *price-weighted relative advantage* of investigator j by

$$\delta_j(p) = pc^l(j) - c^h(j)$$

[29] In the trade literature, the standard measure of comparative advantage is the ratio of productivities. That would be the relevant measure for the "revenue maximization" problem in which price-weighted caseloads are maximized, subject to an upper-bound constraint on the social cost of each investigator's assignment. We study the dual problem of minimizing social cost subject to an upper bound on price-weighted caseloads, which is why the difference in costs is the relevant comparative advantage measure.

and let $d_j(p) := p\gamma_j^l - \gamma_j^h$ be j's *price-weighted comparative advantage score.* If all investigators were to place a relative weight of p on high-type cases then $d_j(p)$ would be the relevant sufficient statistic for the optimal allocation. In the general model with heterogeneous and unobserved preferences there is not a simple sufficient statistic for when one investigator receives more high-type cases than another. However we can use the dual formulation in eq. (3) to give a partial ranking. As one would expect, if $c^l(j) \geq c^l(j')$ and $c^h(j) \leq c^h(j')$ then j receives more high-type cases in expectation.

These identification results relate to a recent literature using the as-if random assignment of decision-makers to separately identify skills from preferences in examiner decisions (Angelova, Dobbie and Yang, 2023; Chan, Gentzkow and Yu, 2022; Rambachan, 2022; Arnold, Dobbie and Hull, 2022). Here, an investigator's performance may be a function of both skills (e.g., the quality of the signals that investigators observe about the potential outcome) and preferences (e.g., their relative distaste for false positives versus false negatives). Our approach makes no attempt at distinguishing between these different factors since, as discussed above, what matters for evaluating the mechanism is the outcome of each type of investigation when assigned to specific investigators. Our framework can therefore be used in a range of settings where a researcher may care about ranking the "performance" of decision-makers without attempting to separate the drivers of such performance. An appealing feature of our approach for measuring the relative (case-type-specific) performance of decision-makers is the relatively mild identifying assumptions. Our approach relies primarily on the conditionally quasi-random assignment of decision-makers to binary decision problems and, as we show in Appendix D, generalizes beyond binary case outcomes.

V Data and estimation strategy

In order to quantify the potential gains from our proposed mechanism, we next estimate differences in $c^k(j)$ using a rich administrative dataset from the State of Michigan.

V.A Data sources and analysis sample

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

We construct our analysis sample as follows. We begin with the set of child maltreatment investigations in Michigan between January 2008 and November 2016 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 ($N = 95,055$).

The resulting analysis sample consists of 278,089 investigations involving 225,487 children assigned to 783 unique investigators; 3.0% of these investigations result in foster care placement. Table A1 presents summary statistics for this analysis sample. Overall, 66% 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 (43%), and physical abuse (28%). 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 (Baron et al., 2024; Putnam-Hornstein and Needell, 2011; Putnam-Hornstein et al., 2015; Putnam-Hornstein, Prindle and Hammond, 2021). Nevertheless, it is clear that re-investigation is an imperfect proxy for actual child maltreatment, as it only accounts for cases that are re-reported to CPS.[30] 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

[30]While there are other potential proxies, such as a subsequent *substantiated* investigation, or a subsequent foster care placement, we prefer re-investigation because re-investigations within a few months may be assigned to the initial investigator who will again make substantiation and foster care placement 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 initial investigator's control.

is mechanically missing for children placed in foster care, which is the primary empirical challenge in this study. 17.1% of children experience subsequent maltreatment in the home within six months of the focal investigation.

V.B Estimation strategy

Lemma 2 provides the key results for using the observed data to identify investigators' performance across heterogeneous cases. Average differences in cost by case type between investigators, $\mathbb{E}[c(i,j) - c(i,j')|i \in \mathcal{I}^k]$, are identified under random assignment by the observed placement and false negative rates.

Because the focus of CPS is to prevent further child maltreatment, we partition cases based on the predicted risk of subsequent maltreatment.[31] We construct this measure by training a machine learning algorithm to predict the risk of subsequent maltreatment in the home among the set of children not placed in foster care.[32] We define high-risk cases as those in the top quartile of predicted algorithmic risk, and low-risk cases as all other cases.

The results of Section IV allow us to identify differences in social cost on low- and high-risk cases between investigators. Our strategy to estimate performance 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 any given office-by-year. 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.[33] 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 as in Dahlstrand (2022). Specifically, we randomize within the set of cases that each investigator was assigned into a "training" set (50%) and an "evaluation" set (50%).[34] We use the training set to derive the optimal

[31]Note that the methods in this section can be readily applied to other important case attributes such as cases involving abuse versus neglect, Black versus White children, or female versus male children.

[32]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.

[33]The results of Section IV 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.

[34]We randomize within investigators to maximize the number of cases per investigator across the sample.

investigator assignment mechanism, and then test its effectiveness on the evaluation set.

We first estimate investigator j's performance score across all cases, γ_j. 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 FN_i an indicator for whether the case is a false negative (the child was not placed in foster care but experienced a subsequent maltreatment investigation in the home).

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 a result, our non-parametric identification results apply separately to each zip code by year. To compare investigators across offices, we follow the literature and use a linear adjustment to estimate investigator placement and false negative rates (Arnold, Dobbie and Hull, 2022).[35] That is, we estimate regressions of the form:

$$D_i = \sum_j \phi_j^D Z_{ij} + \mathbf{X}_i' \alpha^D + u_i \tag{4}$$

$$\text{FN}_i = \sum_j \phi_j^{\text{FN}} Z_{ij} + \mathbf{X}_i' \alpha^{\text{FN}} + v_i \tag{5}$$

We estimate Equations 4 and 5 separately in the training and evaluation samples.[36] $\mathbf{X}_i$ is a vector of zip code-by-investigation year fixed effects to account for the level of randomization. In practice, $\mathbf{X}_i$ is de-meaned so that the ϕ_j are strata-adjusted investigator-specific estimates of each outcome. We denote the strata-adjusted investigator-specific estimates from Equations 4 and 5 as $\widehat{\phi_j^D}$ for placement rates and $\widehat{\phi_j^{\text{FN}}}$ for false negative rates. We use these to estimate γ_j, separately in the training and evaluation dataset, as:

$$\widehat{\gamma_j} = c_{FP} \widehat{\phi_j^D} + (c_{FN} + c_{FP} - c_{TP}) \widehat{\phi_j^{FN}} \tag{6}$$

Following Chan, Gentzkow and Yu (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

[35] As discussed in Arnold, Dobbie and Hull (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.

[36] Our results are robust to the inclusion of child and case controls in the investigator moment regressions. Figure A2 shows how average performance score γ_j and comparative advantage score d_j change if we adjust for the covariates in Table A1. We find that the ranking of investigators is robust to this decision. The rank correlation of the γ_j and d_j measures using the two approaches is 0.978 and 0.993, respectively.

choice of parameter values. To motivate this choice, note that CPS investigators in our context place 3.0% of children but 17.1% 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 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 $c_{FN} = 0.5$.[37] Finally, to reduce noise in the estimates of the investigator moments, we follow the literature (e.g., Arnold, Dobbie and Hull (2022)) and use empirical Bayes estimates of $\widehat{\gamma_j}$ that shrink the estimates using the posterior average effect approach of Bonhomme and Weidner (2022).

We next estimate performance scores across case types: γ_j^l for low-risk cases and γ_j^h for high-risk cases. Let Risky_i be an indicator equal to one if case i is a high-risk case. We estimate the following regressions separately for the training and evaluation set of cases:

$$D_i = \sum_j \beta_{j1}^D Z_{ij} + \beta_{j2}^D \text{ Risky}_i Z_{ij} + \mathbf{X}_i' \alpha^D + u_i \tag{7}$$

$$\text{FN}_i = \sum_j \beta_{j1}^{\text{FN}} Z_{ij} + \beta_{j2}^{\text{FN}} \text{ Risky}_i Z_{ij} + \mathbf{X}_i' \alpha^{\text{FN}} + u_i \tag{8}$$

Here, β_{j1} represent strata-adjusted investigator-specific estimates of each outcome for low-risk cases and $\beta_{j1} + \beta_{j2}$ are strata-adjusted investigator estimates of each outcome for high-risk cases. We denote the strata-adjusted investigator-specific estimates from Equations 7 and 8 as $\widehat{\phi_{j,l}^D}$ for placement rates of low-risk cases, $\widehat{\phi_{j,h}^D}$ for placement rates of high-risk cases, $\widehat{\phi_{j,l}^{\text{FN}}}$ for false negative rates of low-risk cases, and $\widehat{\phi_{j,h}^{\text{FN}}}$ for false negative rates of high-risk cases. Again, we use these to estimate performance scores on low-risk and high-risk cases, separately in the training and evaluation dataset, as:

$$\begin{aligned}\widehat{\gamma_j^l} &= c_{FP}\widehat{\phi_{j,l}^D} + (c_{FN} + c_{FP} - c_{TP})\,\widehat{\phi_{j,l}^{FN}}\\ \widehat{\gamma_j^h} &= c_{FP}\widehat{\phi_{j,h}^D} + (c_{FN} + c_{FP} - c_{TP})\,\widehat{\phi_{j,h}^{FN}}\end{aligned}$$

where we use the same social costs as in our baseline skill estimate, but show robustness to this decision in Figure A1. We again use empirical Bayes estimates of $\widehat{\gamma_j^l}, \widehat{\gamma_j^h}$ to adjust for finite sample error, and estimate $\tilde{c}^k(j)$ as $\widehat{\gamma_j^k} - \widehat{\gamma_{j^0}^k}$.

[37] We emphasize that we make assumptions on the value of these parameters simply for exposition—the empirical analysis seeks to simulate gains from investigator reallocation under reasonable social cost assumptions. However, our theoretical results hold without assumptions on the exact nature of these costs. In practice, the value of these parameters should be dictated by the relevant agency or society.

VI Main Empirical Results

VI.A Motivating empirical facts

We begin by presenting three empirical facts that motivate the use of our proposed mechanism in this setting.

Considerable variation in performance and comparative advantage: We first use the investigator moments to understand variation in performance across investigators and case types. Intuitively, gains from investigator reassignment in our proposed mechanism can only occur if there is sufficient heterogeneity in investigators' relative performance across cases. 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. Figure A3 plots the distributions of γ_j and d_j. The standard deviation of performance and comparative advantage scores is 5.1 and 10.6pp, respectively. To understand the significance of such variation, Table A2 estimates the relationship between performance metrics on the training dataset and prediction error rates in the evaluation dataset. Panel A shows that investigators with a one standard deviation γ_j below the mean achieve a 1.1pp [6.6%] reduction in false negatives and 1.6pp [67.1%] 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 1.5pp [5.9%] lower false negative rates and 0.8pp [21.7%] lower false positive rates in high-risk cases, but only 0.1pp [0.4%] lower false negative rates and 0.8pp [44.4%] *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. Importantly, these effects are estimated using a split-sample strategy, so that they are not mechanical.

Most variation is within CPS offices: Intuitively, if variation in performance is mostly driven by across-office variation, then counterfactual exercises that re-assign investigators to cases only within offices would have limited promise. Figure A4 reports the results of this decomposition, following the method in Chetty, Friedman and Rockoff (2014).[38] 85% of variation in comparative advantage scores is within CPS offices.[39]

[38]This approach is similar to regressing γ_j on an exhaustive set of office fixed effects and interpreting the R^2 of the resulting regression as the variation in performance explained across offices.

[39]This result, while perhaps surprising, is consistent with that in Chetty, Friedman and Rockoff (2014), who find that 85% of variation in teacher value-added is within rather than between schools.

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. 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 104%. Column 2 reports that being assigned to an above-median share of high-risk cases increases turnover risk by 30%. Thus, exposure to a greater share of high-risk cases leads to large increases in investigator turnover, suggesting that these cases are in fact more costly to investigators.

Altogether, these patterns motivate the potential for the SMD-TP mechanism to achieve welfare gains in this context. There is evidence of significant comparative advantage across high- and low- risk cases in the data, and much of this variation is within offices. However, without compensating investigators accordingly, assignment to additional high-risk cases is costly to investigators and could result in substantial increases in turnover. We next quantify the potential welfare gains from the SMD-TP mechanism in a policy simulation.

VI.B Social welfare gains

Corollary 4 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 counterfactual which splits high- and low-risk cases equally across investigators within counties.[40]

Our strategy to estimate welfare gains accounts for (i) uncertainty in investigator types and (ii) over-fitting concerns. First, because investigator types are unknown and our approach requires some knowledge of the investigator-type distribution, we simulate the mechanism for a range of type distributions. Using a split-sample strategy, we combine 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

[40]The equal split of cases is the expected assignment under the rotational system. Table A4 shows that we obtain similar welfare gains when we benchmark the SMD-TP mechanism to the observed status quo. Note that the SMD-TP mechanism reassigns investigators within counties. While rare, in the status quo, investigators sometimes work across different counties over their tenure. This could be due, for example, to investigators working near county borders or moving to a different county. To avoid complicating comparisons of our assignment to the status quo, investigators are limited to serve cases in their modal county in the counterfactual. In the rare case where a county is not modal to any investigator, we combine that county with a neighboring county.

Table 1: Gains from Investigator Reallocation

	(1)	(2)	(3)	(4)	(5)	(6)
	Unif[1,2]	Unif[1,3]	$\mathcal{N}(2, 0.5^2)$	$\mathcal{N}(2, 1^2)$	$p_j = 2$	Known type, Unif[1,2]
Social Costs	-784.1***	-698.7***	-642.8***	-670.3***	-1,451.8***	-1,767.2***
	(203.9)	(194.0)	(217.7)	(215.1)	(158.8)	(209.7)
	[-4.3%]	[-3.9%]	[-3.6%]	[-3.7%]	[-8.1%]	[-9.8%]
False Negatives	-547.5***	-490.5***	-470.0***	-488.2***	-1,135.1***	-1,268.0***
	(151.5)	(170.9)	(175.1)	(165.8)	(121.7)	(161.0)
	[-1.2%]	[-1.1%]	[-1.0%]	[-1.1%]	[-2.5%]	[-2.7%]
False Positives	-662.5***	-587.8***	-538.2***	-560.2***	-1,196.2***	-1,485.9***
	(148.6)	(166.3)	(186.0)	(164.8)	(142.3)	(175.6)
	[-10.3%]	[-9.1%]	[-8.3%]	[-8.7%]	[-18.5%]	[-23.0%]
Placements	-115.0	-97.4	-68.2	-72.0	-61.1	-217.9**
	(79.8)	(80.4)	(80.0)	(88.5)	(74.9)	(92.0)
	[-1.4%]	[-1.2%]	[-0.9%]	[-0.9%]	[-0.8%]	[-2.7%]

Notes. This table reports the welfare gains derived from the SMD-TP mechanism. Each column corresponds to a different distributional assumption for p_j. Columns 1 and 2 present uniform distributions with supports $[1, 2]$ and $[1, 3]$, respectively. Columns 3 and 4 present truncated normal distributions (in $[1, 3]$) , both with a mean of 2 and standard deviations of 0.5 and 1, respectively. Column 5 presents a degenerate distribution where $p_j = 2$. Column 6 assumes that types are distributed uniformly with support $[1, 2]$, but 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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 the estimated welfare gains across a set of initial distributional assumptions. Each column presents changes in outcomes when investigators are reassigned to cases according to the SMD-TP mechanism within counties versus a counterfactual that equally splits high- and low-risk cases within counties. For example, when we assume $p_j \sim \text{Unif}[1, 2]$ in Column 1, we find declines in social costs of 784 (4.3%). This is due to a reduction of both 547 false negative cases (1.2%) and 663 false positive cases (10.3%).[41] The SMD-TP mechanism also reduces the number of total placements by 115 (1.4%), though this estimate is imprecise.[42]

For the remainder of the paper, we treat $p_j \sim \text{Unif}[1, 2]$ as our preferred type distribution. However, Columns 2–5 of Table 1 show that we find similar results across a range of

[41] 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 greater detail in Appendix G.

[42] Figure A5 presents the welfare changes from each of the 100 type draws when $p_j \sim \text{Unif}[1, 2]$.

distributional assumptions. The largest gains occur when p_j follows a degenerate distribution. As we show below, the welfare gains under this distribution are even larger than those under the LMS-TP mechanism (Column 1, Table A6). This gap highlights the fact that a naive analysis which ignores investigators' private information, and the attendant information rents, would significantly overstate welfare gains.

To estimate the importance of investigator private information for welfare gains, in Column 6 we again assume that $p_j \sim \text{Unif}[1, 2]$ but that the designer observes each investigator's type directly, and implements the first-best assignment for each realized type profile. In this simulation, the welfare gains increase dramatically—social costs decline by 9.8%, driven by a false negative decline of 2.7% and false positive decline of 23.0%. The comparison with Column 1 shows that information rents are significant when types are unobserved. As discussed in Appendix A.2, over time the designer will be able to use the data generated by the mechanism to reduce uncertainty about investigators' preferences. Column 6 represents an upper bound on welfare gains as the designer learns about individual investigators' types.

When comparing the gains across columns, one might expect more dispersed type distributions to induce lower welfare gains; we saw above that greater uncertainty over investigator types increases the information rents that the designer must pay (in the form of lower price-weighted caseloads) to induce truthful reporting. However, comparing Columns 3 and 4 in Table 1 shows that greater dispersion does not always reduce the welfare gains. To understand why, consider the stylized model in which all investigators have the same type, p, as in Column 5. In this model, the "level effect" of changing p on the welfare gains is ambiguous, as this parameter affects the value of the status-quo endowment as well as the value of alternative assignments.[43] With unobserved preference heterogeneity, increasing the variance of the type distribution has a negative "uncertainty effect" but an ambiguous "level effect," which can potentially offset the uncertainty effect. Since types follow the same distribution in Columns 1 and 6, comparing the welfare gains in these two columns isolates the uncertainty effect under that type distribution.

Altogether, the results in this section highlight that our SMD-TP mechanism could reduce both types of prediction mistakes, as well as overall foster care placement rates, by reallocating investigators within counties in a revenue-neutral way.

[43] The impact of this level effect on welfare gains depends on details of the distribution of comparative advantage and the relative prevalence of high- and low-risk cases.

VI.C Investigator preferences

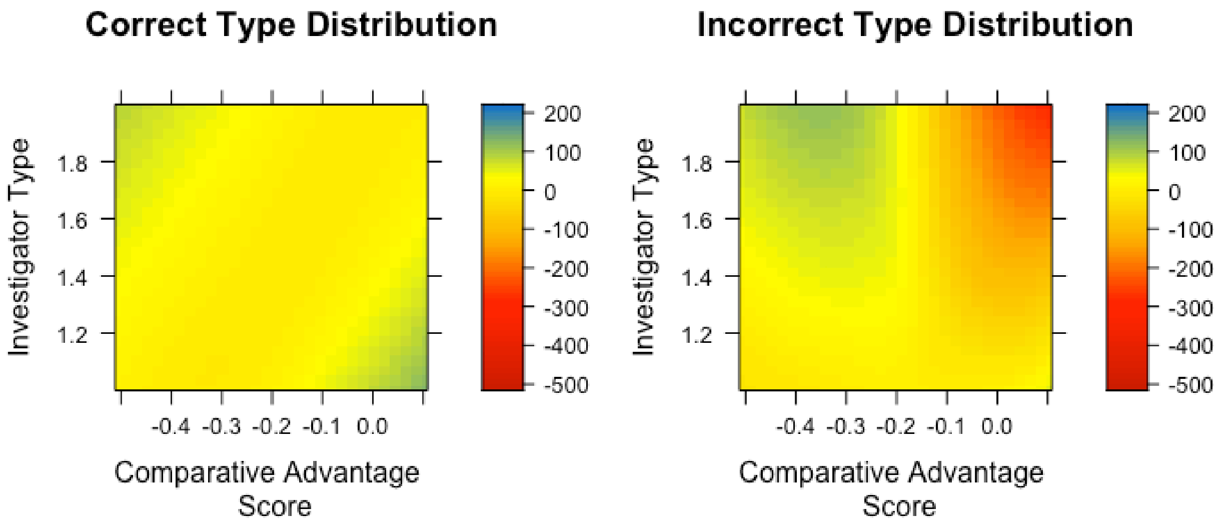
Figure 3 demonstrates the importance of considering investigators' heterogeneous preferences in the SMD-TP mechanism. We define investigator welfare for an investigator with type p_j and caseload (n^h, n^l) as $-(p_j n^h + n^l)$, the negative of their price-weighted caseload. In the left panel of Figure 3, we derive the optimal allocation of cases assuming that $p_j \sim \text{Unif}[1,2]$. 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 12 price-weighted cases, on average. Average investigator welfare is -405 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 5 price-weighted cases, and the 1st percentile of investigator welfare change is a loss of 1.5 cases.[44] In fact, 38% of investigators experience welfare gains under the correct SMD-TP mechanism of greater than five price-weighted cases–which could in turn help improve recruitment and retention in our context.[45]

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 are truly distributed as $p_j \sim \text{Unif}[1,2]$. Interestingly, under this scenario, the average investigator welfare change is small: a gain of 0.05 price-weighted cases relative to the average equal-split investigator welfare of -405. However, Figure 3 shows that this average obscures significant heterogeneity in investigator welfare loss by comparative advantage and type. 30% of investigators experience welfare losses greater than 5 price-weighted cases. The 10th percentile of welfare change is a loss of 137.3 cases (or 34% relative to the equal-split mean), while the 1st percentile is a loss of 271.0 (67%) cases. Importantly, 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 73.6 (18%), and those in the top quartile of both experience an average welfare loss of 157.3 (39%). On the other hand, investigators with low comparative advantage on high-risk cases and high p_j are made better-off under

[44] 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.

[45] The 90th percentile experiences welfare gains of 43 (11%) cases, and the 99th percentile experiences welfare gains of 85 cases (21%).

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 $p_j \sim \text{Unif}[1, 2]$. The right panel calculates changes in investigator welfare under an SMD-TP mechanism that assumes $p_j = 1 \ \forall j \in \mathcal{J}$, but where the true p_j is distributed according to $\text{Unif}[1, 2]$. We present results averaged across the 100 investigator-type draws.

the mechanism that ignores preference heterogeneity.[46]

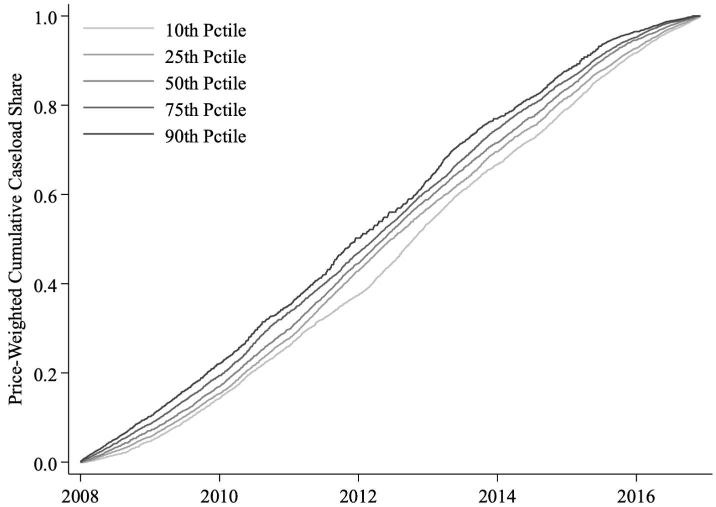
Figure 3 highlights why considering investigator preferences in the assignment problem is paramount. When the mechanism ignores types, investigators with large comparative advantage in high-risk cases receive more of these cases. But if assignment to high-risk cases is costly relative to low-risk cases, then these investigators are made substantially worse-off. 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.

VI.D 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. Smoothing caseloads over time does not directly enter the mechanism-design problem as a constraint, but the solution does attempt to spread caseloads over the period. It does this 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

[46] Figure A6 replicates this figure but adds a dot for each investigator, limiting to one randomly chosen investigator-type-draw for visual clarity. This figure highlights the number of investigators adversely impacted by the mechanism that does not consider preference heterogeneity.

Figure 4: 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 33 counties that appear in every sample year. Cases are assigned according to the SMD-TP mechanism for one investigator type draw where $p_j \sim \text{Unif}[1, 2]$. We compute cumulative price-weighted caseloads in day t for investigator j as $\frac{\hat{n}^l_j(t)+p_j\hat{n}^h_j(t)}{\hat{n}^l_j(T)+p_j\hat{n}^h_j(T)}$, where T is the last day of the sample period. We then report percentiles of this statistic for each day of our sample.

time, each investigator should have completed approximately the same percentage of their high- or low-risk cumulative caseload. Figure 4 describes how cumulative price-weighted workloads vary over time.[47] This figure shows that the SMD-TP mechanism is successful in spreading caseloads.

For reference, we also estimate the welfare gains from the SMS-TP and LMS-TP mechanisms, and compare these to outcomes under the SMD-TP mechanism in Table A6. Comparing Columns 2 and 3, we find that moving from LMS-TP to SMS-TP decreased welfare gains substantially. This difference represents the cost of aggregate uncertainty about investigators' types. However, differences between the SMS-TP and the SMD-TP are small, which shows that the need to assign cases without observing which cases will arrive in the future (the "online" nature of assignments) does not appear to be a first-order problem.

VI.E Allowing correlation between preferences and performance

Finally, we consider how our social welfare gains would change if we allowed for correlation between investigator preferences and their comparative advantage score, d_j. Let $\underline{d_j}, \bar{d}_j$ be the

[47]We limit this exercise to the set of 33 counties that appear in the sample each year. These counties make up roughly 90% of all cases in the sample. In Table A5, we re-estimate welfare gains under SMD-TP for this sample and find very similar results relative to those in Table 1.

minimum and maximum d_j within counties, respectively. Then, for this analysis, we assume that investigator types are draw from a uniform distribution with full support on $[g(d_j) + 1, g(d_j) + 2]$, where $g(d_j) = b\frac{d_j - \underline{d_j}}{\overline{d_j} - \underline{d_j}}$ for $b \geq 0$ and $g(d_j) = -b(1 - \frac{d_j - \underline{d_j}}{\overline{d_j} - \underline{d_j}})$ for $b < 0$.[48] Then, the associativity parameter b captures the strength and direction of the correlation between comparative advantages and type distributions, 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 A7 reports the results of this exercise. When investigators with high d_j tend to have lower type draws, we find that welfare gains are significantly larger: for $b = -2$, the welfare gains are $5,100$ 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 an almost five-fold increase in the welfare gains. This demonstrates the intuitive notion that when investigators who have a comparative advantage in high-risk cases also relatively prefer these cases, the mechanism can achieve larger welfare gains.[49] On the other hand, if investigators with a comparative advantage in high-risk cases tend to relatively dislike such cases, the welfare gains are mildly attenuated. The largest reduction in Figure A7 occurs when $b = 0.5$, in which case welfare gains are 900, a 15% decline compared to the $b = 0$ case. Thus, while a strong positive correlation between comparative advantage and p_j may reduce the potential welfare gains, there still exists a significant potential for welfare improvement even under this scenario.[50]

VII Conclusion

The ultimate objective of this work is a practical mechanism for assigning CPS investigators to reported cases of child maltreatment. This paper has sought to address what we view as the primary challenges that such a mechanism must overcome:

1. *Identifying social preferences over alternative mechanisms.* Using data from the status-quo random assignment mechanism, we showed in Section IV that we can identify the relevant moments of the joint distribution of investigator decisions and potential case

[48] 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 in the county under associativity parameter $-b$.

[49] On the other hand, there is potentially a countervailing force due to the fact that preferences also affect the value of the status quo. See the discussion of this "level effect" in Section VI.B.

[50] Strong positive correlation appears unlikely in the current context: When exploring the relationship between higher-risk caseloads and turnover (as in Table A3), 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.

outcomes, which is sufficient for evaluating a mechanism's performance. Moreover, we discuss in Appendix A.2 that it will be possible to continue to learn about this joint distribution under the new proposed mechanism.

2. *Unobservable investigator preferences and status-quo constraints.* In order to avoid negative impacts on the recruitment and turnover of investigators, and to facilitate the political task of convincing agencies to adopt the proposed mechanism, we restricted our attention to mechanisms that do not make any agents worse-off. Careful design of the mechanism is needed to deal with the fact that investigators' preferences are inherently unobservable.

3. *Effort incentives.* While we do not explicitly model the decision of investigators to exert effort, we showed that within our proposed mechanism investigators' payoffs are indeed improving in their measured performance, at least locally (Theorem 3). Thus, if the mechanism is implemented in successive periods (e.g., each year) and data from past performance is used to inform future assignments, the mechanism should provide self-interested investigators with motivation to perform well.

4. *Perceived fairness of the mechanism.* Within our mechanism, it is possible for investigators with the same preferences to receive different allocations. This is true even among investigators who exclusively handle the same type of case. One might be concerned with how investigators will react to this disparity (even if every investigator is better-off relative to the current system). Fortunately, we showed that disparate caseloads can be justified on the basis of performance: investigators who receive fewer type-k cases for the same type-report are precisely those who perform better on type-k cases (Theorem 3).

5. *Beyond binary case classifications.* We focused on mechanisms with conditional assignments on a binary partition of cases into high- and low-risk types. It is worth emphasizing 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 A.3). Moreover, our main identification results in Section IV do not depend on the binary partition assumption.

Before implementing the mechanism in the field, several practical considerations must be carefully addressed. One critical aspect involves effectively communicating the mechanism to investigators and establishing a clear protocol for reporting their preferences. Although communicating the structure of the mechanism itself is relatively straightforward—each investigator is provided with two exchange rates, one for gaining high-type cases and one for gaining low-type cases—it will be important to ensure that investigators fully comprehend the implications of their type reports.

We are currently working directly with CPS agencies to tackle these practical details and begin a pilot implementation of the mechanism. This pilot phase will yield valuable data on the distribution of investigator preferences which will be used to further refine the mechanism. Moreover, these data will shed light on key questions regarding investigator preferences, such as the determinants of these preferences, their correlation with performance and other observables, and strategies for improving investigator recruitment and retention. By applying our proposed mechanism, we hope to gain insights into these important questions, ultimately contributing to improving the quality of CPS responses.

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School choice: A mechanism design approach." *American economic review*, 93(3): 729–747.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. 2017. "Research design meets market design: Using centralized assignment for impact evaluation." *Econometrica*, 85(5): 1373–1432.

Abdulkadiroğlu, Atila, Parag A Pathak, and Alvin E Roth. 2005. "The new york city high school match." *American Economic Review*, 95(2): 364–367.

Abdulkadiroğlu, Atila, Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux. 2020. "Efficiency, justified envy, and incentives in priority-based matching." *American Economic Review: Insights*, 2(4): 425–442.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag Pathak. 2022. "Breaking ties: Regression discontinuity design meets market design." *Econometrica*, 90(1): 117–151.

Agarwal, Nikhil, Charles Hodgson, and Paulo Somaini. 2020. "Choices and outcomes in assignment mechanisms: The allocation of deceased donor kidneys." National Bureau of Economic Research.

Aggarwal, Gagan, Gagan Goel, Chinmay Karande, and Aranyak Mehta. 2011. "Online vertex-weighted bipartite matching and single-bid budgeted allocations." 1253–1264, SIAM.

Akbarpour, Mohammad, Piotr Dworczak, and Scott Duke Kominers. 2023. "Redistributive allocation mechanisms." *Available at SSRN 3609182*.

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 Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

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.

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

Ba, Bocar, Patrick Bayer, Nayoung Rim, Roman Rivera, and Modibo Sidibé. 2022. "Police officer assignment and neighborhood crime." National Bureau of Economic Research Working Paper 29243.

Bald, Anthony, Eric Chyn, Justine S. Hastings, and Margarita Machelett. 2022*a*. "The Causal Impact of Removing Children from Abusive and Neglectful Homes." *Journal of Political Economy*, 130(7): 1919–1962.

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

Baron, E Jason, and Max Gross. 2022. "Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators." National Bureau of Economic Research Working Paper 29922.

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.

Bates, Michael, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin. 2023. "Teacher labor market policy and the theory of the second best." *Working Paper.*

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris. 2015. "The limits of price discrimination." *American Economic Review*, 105(3): 921–957.

Bergeron, Augustin, Pedro Bessone, John Kabeya Kabeya, Gabriel Z Tourek, and Jonathan L Weigel. 2022. "Optimal assignment of bureaucrats: Evidence from randomly assigned tax collectors in the DRC." *NBER Working Paper #30413.*

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. 2020. "The glittering prizes: Career incentives and bureaucrat performance." *The Review of Economic Studies*, 87(2): 626–655.

Biasi, Barbara, Chao Fu, and John Stromme. 2021. "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage." National Bureau of Economic Research Working Paper 28530.

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

Bonhomme, Stéphane, and Martin Weidner. 2022. "Minimizing sensitivity to model misspecification." *Quantitative Economics*, 13(3): 907–954.

Border, Kim C. 1991. "Implementation of reduced form auctions: A geometric approach." *Econometrica: Journal of the Econometric Society*, 1175–1187.

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

Bosk, Emily Adlin. 2015. "All Unhappy Families: Standardization and Child Welfare Decision-Making." PhD diss. University of Michigan.

Bray, Robert L, Decio Coviello, Andrea Ichino, and Nicola Persico. 2016. "Multitasking, multiarmed bandits, and the Italian judiciary." *Manufacturing & Service Operations Management*, 18(4): 545–558.

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

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.

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

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.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Che, Yeon-Koo, and Olivier Tercieux. 2019. "Efficiency and stability in large matching markets." *Journal of Political Economy*, 127(5): 2301–2342.

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.

Combe, Julien, Vladyslav Nora, and Olivier Tercieux. 2021. "Dynamic assignment without money: Optimality of spot mechanisms." *Available at SSRN 3894259*.

Condorelli, Daniele, and Balázs Szentes. 2020. "Information design in the holdup problem." *Journal of Political Economy*, 128(2): 681–709.

Dahlstrand, Amanda. 2022. "Defying Distance? The Provision of Services in the Digital Age." Working Paper.

Delacrétaz, David, Scott Duke Kominers, and Alexander Teytelboym. 2023. "Matching mechanisms for refugee resettlement." *American Economic Review*, 113(10): 2689–2717.

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

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108(2): 201–40.

Doyle, Joseph J. 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review*, 97(5): 1583–1610.

Doyle, Joseph J. 2008. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." *Journal of Political Economy*, 116(4): 746–770.

Dworczak, Piotr, and Ellen Muir. 2023. "A mechanism-design approach to property rights." working paper.

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

Fitzpatrick, Maria, Katharine Sadowski, and Christopher Wildeman. 2022. “Does Access to an Algorithmic Decision-making Tool Change Child Protective Service Caseworkers’ Investigation Decisions?” Unpublished.

Gale, David, and Lloyd S Shapley. 1962. “College admissions and the stability of marriage.” *The American Mathematical Monthly*, 69(1): 9–15.

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.

Gopalan, Parikshit, Noam Nisan, and Tim Roughgarden. 2018. “Public projects, boolean functions, and the borders of border’s theorem.” *ACM Transactions on Economics and Computation (TEAC)*, 6(3-4): 1–21.

Greenberg, Kyle, Parag A Pathak, and Tayfun Sönmez. 2023. “Redesigning the US Army’s Branching Progress: A Case Study in Minimalist Market Design.” *arXiv preprint arXiv:2303.06564*.

Grimon, Marie-Pascale. 2020. “Effects of the Child Protection System on Parents.” Working paper.

Grimon, Marie-Pascale, and Christopher Mills. 2022. “The Impact of Algorithmic Tools on Child Protection: Evidence from a Randomized Controlled Trial.” Job market paper.

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

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.

Harsanyi, John C. 1955. “Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility.” *Journal of political economy*, 63(4): 309–321.

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.

Helénsdotter, Ronja. 2022. “Health Effects of Removing a Child From Home.” Unpublished.

Hoffman, Mitchell, Lisa B Kahn, and Danielle Li. 2018. “Discretion in hiring.” *The Quarterly Journal of Economics*, 133(2): 765–800.

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: Journal of the Econometric Society*, 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.

Kapon, Samuel, Lucia Del Carpio, and Sylvain Chassang. 2022. “Using Divide-and-Conquer to Improve Tax Collection.” National Bureau of Economic Research.

Karp, Richard M, Umesh V Vazirani, and Vijay V Vazirani. 1990. "An optimal algorithm for on-line bipartite matching." 352–358.

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

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.

Laverde, Mariana, Elton Mykerezi, Aaron Sojourner, and Aradhya Sood. 2023. "Gains from Reassignment: Evidence from a Two-Sided Teacher Market." *Available at SSRN.*

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

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

Loertscher, Simon, and Ellen V Muir. 2021. "Wage dispersion, minimum wages and involuntary unemployment: A mechanism design perspective." Mimeo.

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

Matthews, Steven A. 1984. "On the implementability of reduced form auctions." *Econometrica: Journal of the Econometric Society*, 1519–1522.

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.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and product quality." *Journal of Economic theory*, 18(2): 301–317.

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." *arXiv preprint arXiv:2303.13967.*

Nguyen, Thành, and Rakesh Vohra. 2021. "Δ-Substitute Preferences and Equilibria with Indivisibilities." 735–736.

Pathak, Parag A. 2011. "The mechanism design approach to student assignment." *Annu. Rev. Econ.*, 3(1): 513–536.

Porto, Edoardo Di, Nicola Persico, and Nicolas Sahuguet. 2013. "Decentralized deterrence, with an application to labor tax auditing." *American Economic Journal: Microeconomics*, 5(1): 35–62.

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

Putnam-Hornstein, Emily, and Barbara Needell. 2011. "Predictors of Child Protective Service Contact Between Birth and Age Five: An Examination of California's 2002 Birth Cohort." *Children and Youth Services Review*, 33(8): 1337–1344.

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.

Putnam-Hornstein, Emily, Julie A Cederbaum, Bryn King, Andrea L Eastman, and Penelope K Trickett. 2015. "A Population-Level and Longitudinal Study of Adolescent Mothers and Intergenerational Maltreatment." *American Journal of Epidemiology*, 181(7): 496–503.

Rambachan, Ashesh. 2022. "Identifying Prediction Mistakes in Observational Data." Harvard University Working Paper.

Rittenhouse, Katherine, Emily Putnam-Hornstein, and Rhema Vaithianathan. 2022. "Algorithms, Humans, and Racial Disparities in Child Protective Services: Evidence from the Allegheny Family Screening Tool." Working paper.

Roberts, Donald John, and Andrew Postlewaite. 1976. "The incentives for price-taking behavior in large exchange economies." *Econometrica: journal of the Econometric Society*, 115–127.

Shapley, Lloyd, and Herbert Scarf. 1974. "On cores and indivisibility." *Journal of mathematical economics*, 1(1): 23–37.

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.*

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

Wildeman, Christopher, and Jane Waldfogel. 2014. "Somebody's children or nobody's children? How the sociological perspective could enliven research on foster care." *Annual review of sociology*, 40: 599–618.

Appendix

Mechanism Reform: An Application to Child Welfare

E. Jason Baron, Richard Lombardo, Joseph Ryan, Jeongsoo Suh, Quitzé Valenzuela-Stookey

A Extensions and additional properties

A.1 An impossibility result for robust design

We have assumed, as is standard in the Bayesian mechanism design literature, that the designer knows the distribution of each agent's preference type. An alternative would be to design a mechanism which is robust to uncertainty about these distributions.

Suppose that we do not know the type distributions $\{F_j\}_{j=1}^J$, and we want to design an assignment mechanism that is robust to this uncertainty. Given a profile of distributions, $\{F_j\}_{j=1}^J$, and a mechanism M, let $V(\{F_j\}_{j=1}^J, M)$ be the induced expected cost. The robust mechanism design problem is to choose a mechanism to solve

$$\min_{M} \max_{\{F_j\}_{j=1}^J} V(\{F_j\}_{j=1}^J, M)$$

and a solution is called a *robustly optimal* mechanism. Let M_{SQ} be the status-quo assignment. This is a type-independent assignment, and so the cost is independent of $\{F\}_{j=1}^J$. Denote this cost by V_{SQ}. If $\max_{\{F_j\}_{j=1}^J} V(\{F_j\}_{j=1}^J, M) < V_{SQ}$ we say that M *robustly improves upon the status quo*.

We might also be interested in understanding which distributions are worst for the principal, in the sense that they make it hardest to minimize cost. That is, we would like to know which F_j (and M) solve

$$\max_{\{F_j\}_{j=1}^J} \min_{M} V(\{F_j\}_{j=1}^J, M).$$

Obviously weak duality holds, i.e.

$$\max_{\{F_j\}_{j=1}^J} \min_{M} V(\{F_j\}_{j=1}^J, M) \leq \min_{M} \max_{\{F_j\}_{j=1}^J} V(\{F_j\}_{j=1}^J, M).$$

Lemma 3. Strong duality holds. In particular

$$\max_{\{F_j\}_{j=1}^J} \min_{M} V(\{F_j\}_{j=1}^J, M) = \min_{M} \max_{\{F_j\}_{j=1}^J} V(\{F_j\}_{j=1}^J, M) = V_{SQ}$$

Lemma 3 is proved along with Theorem 4 below. Call a mechanism *simple* if each agent's assignment depends only on their own type.

Theorem 4. There is no BIC and IR mechanism that robustly improves on the status quo. Moreover, the status quo is the unique simple and robustly optimal mechanism.

Remark 6. The impossibility result obtains even if we assume that the type distributions are regular, i.e. ϕ is weakly increasing, and symmetric.

The full proof of Theorem 4 is found in Appendix B.5. The general idea is as follows. Suppose that all agents have the same type distribution, F, with incentive-feasible set $\mathcal{F}$. Then, we can think of the assignment problem as choosing J points in $\mathcal{F}$ which average to (n^h, n^l). If (n^h, n^l) is an extreme point of $\mathcal{F}$ then by definition the only such distribution is to set all agents assignments identically equal to (n^h, n^l). We show that indeed there exist distributions for which (n^h, n^l) is an extreme point of $\mathcal{F}$; in particular, this is true if the virtual value is constant.[51]

Since the minimizing player, i.e., the mechanism designer, has the option to choose the status-quo assignment, it must be that the value of the min-max problem is less than V_{SQ}. Conversely, the maximizing player can choose a symmetric type distribution with a constant virtual value, which by the previous discussion implies that the value of the max-min problem is at least V_{SQ}. This proves Lemma 3.

Since the impossibility result holds in the relaxed LMS problem, it holds *a fortiori* in the SMS and SMD problems. One might be concerned that there are no gains to be had under any type distribution. The empirical application illustrates that this is not the case. More explicitly, it is easy to show that there exist type distributions such that (n^h, n^l) is in the interior of $\mathcal{F}$.

Example 1. Assume that F is uniformly distributed on $[1, 2]$ and $n^l = \frac{4}{3}n^h$ under the status quo. Then, the status quo, (n^h, n^l), is in the interior of $\mathcal{F}$. To see this, consider a two-price mechanism with $p_1 = \frac{4}{3}$ and $p_2 = \frac{5}{3}$. These induce $\hat{n}^h = n^h$, and $\hat{n}^l = \frac{39}{36}n^l$.

Theorem 4 implies that we need to make use of information about investigators' type distributions in order to improve upon the status quo. It is worth emphasizing that despite this impossibility result, solving the mechanism-design problem still allows us to demand less of the data: we do not need the data to tell us each agent's type, just the distribution of this type. Fortunately, this information can be realistically obtained, unlike knowledge of each investigator's realized type, which as discussed above is fundamentally unobservable. To learn about the type distributions, we need to first come up with a practical mechanism which can be implemented in the field, and which induces agents to truthfully report their types, using our best a-priori guess of the type distributions. By running this mechanism, we will then gain insights into the type distribution which can be used for further refinement.

[51] This holds if types follow a Pareto distribution. Pareto distributions appear in relation to robust design problems in other contexts, e.g. Bergemann, Brooks and Morris (2015) and Condorelli and Szentes (2020).

An alternative approach is to use a mechanism that is not simple. That is, to obtain information about the type distribution by using a mechanism in which each agent's allocation depends not only on their own type report, but also on the profile of type reports in the population (which is exactly the type distribution under truthful reporting with a continuum of investigators). Theorem 4 tells us that we cannot get a robust improvement in this way, but it may be possible to strictly improve upon the status quo for some type distributions, while doing no worse in any instance.

For example, consider the direct implementation of the LMS-TP mechanism from Theorem 2, in which each investigator is presented with a pair of prices (p_1^j, p_2^j). When the profile of type distributions was fixed at $\{F^j\}_{j=1}^J$, and known to the designer, we allowed agents to exchange their case-types freely at these prices, and chose prices to ensure that the market cleared. Suppose that instead the designer present agents with the same prices, but asks them to report whether they would like to be "buyers" or "sellers" of high-type cases given these prices, or retain the status quo. The designer can then choose which trades to execute so that markets clear. If the realized type distributions are in fact $\{F^j\}_{j=1}^J$ then all trades are executed. However, for some distributions the designer can simply choose to implement the status quo. Thus, the allocations respond to the realized type distribution, and so the impossibility result is circumvented. We use precisely such a mechanism in Section III.D to adapt the LMS-TP mechanism to the small market, in which there is aggregate uncertainty about the type profile.

A.2 Learning in the mechanism

There are two dimensions along which the designer would like to learn while running the mechanism.

Learning about the type distribution

As demonstrated by Theorem 4, we need to place some restrictions on the type distribution in order to get improvements over simple random assignment. While it is a common assumption in the mechanism-design literature that agents' type distributions are known to the designer, this assumption becomes challenging when the mechanism needs to be applied in practice.

One simple solution to the learning problem is simply to run the mechanism for a trial period using our best guess of the type distribution for each agent. This initial guess of the distribution can be informed, for example, by a preliminary survey of investigators. Since agents' choices in the mechanism reveal information about their types, we can use the observed choices to learn more about the type distribution. This information will also allow us to better predict individual investigators' types based on observables, and so reduce the size of information rents and improve the mechanism's performance.[52]

This approach raises the question of how the trial period should be chosen. On the one hand, we learn more with a longer trial period. Moreover, the performance of the SMD-TP mechanism is improving in the length of the assignment period. On the other hand, we want to use the information about type distributions to optimize prices as soon as possible. A full solution to this problem, which involves carefully weighing this trade-off, is beyond the scope of the current paper. The recent work of Nguyen, Teytelboym and Vardi (2023) provides a model for how this problem could be approached.

Learning about $c^h(j), c^l(j)$

In Section IV, we leverage the quasi-random nature of the observed assignment to identify the cost parameters $c^k(j)$. A natural concern is that if we implement the proposed SMD-TP mechanism we will 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 2. 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. 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).

A.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

[52] 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.

Lemma 2 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 robustness (Theorem 4) and effort incentives (Theorem 3). Nonetheless, generalizations beyond binary partitions, particularly by pursuing the second approach above, are an interesting direction for future work.

B Omitted proofs

B.1 Proof of Theorem 1 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 IR constraint. Write the IR constraint as

$$n^h + n^l - \underline{p}H(\underline{p}) - L(\underline{p}) - \int_{\underline{p}}^{p} (H(z) - n^h)dz \geq 0 \quad \forall p \in [\underline{p}, \bar{p}]$$

or equivalently

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

Let $\phi(p) := p - \frac{1-F(p)}{f(p)}$ be the *virtual type* of p. As in Myerson (1981), we say that F is (strictly) regular if ϕ is (strictly) increasing. Putting together our previous observations, the incentive feasibility of $\hat{n}^h, \hat{n}^l$ boils down to finding a mechanism such that

$$H \text{ is non-increasing} \tag{IC'}$$

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

$$\int_{\underline{p}}^{\bar{p}} H(p)dF(p) \geq \hat{n}^h \tag{h-capacity}$$

$$\underline{p}H(\underline{p}) + L(\underline{p}) - \int_{\underline{p}}^{\bar{p}} H(p)\phi(p)dF(p) \geq \hat{n}^l \tag{l-capacity'}$$

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

and the program defining the support function $S(a, b)$ becomes

$$S(a,b) = \max_{H,L}\ b\left(\underline{p}H(\underline{p}) + L(\underline{p})\right) + \int_{\underline{p}}^{\bar{p}} H(p)\left(a - b\phi(p)\right)dF(p) \tag{9}$$

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

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

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

By inspection of the program in eq. (9), it is optimal to choose $L(\underline{p})$ so that the (IR') constraint binds. Then the program becomes

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

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

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

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

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

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

$$n^h + n^l - \underline{p}H(\underline{p}) - \int_{\underline{p}}^{p^*} (H(z) - n^h)dz \geq 0$$

$$\text{(non-negative } L)$$

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

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

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. 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 $x \in (0, n^h)$.

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

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

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

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

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

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 $x \in (n^h, H(\underline{p}))$. That is, 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 ϕ is strictly increasing.

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. This implies that (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, we conclude that 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 1. When ϕ 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. □

B.2 Proof of Corollary 2

Proof. If F is strictly regular, Theorem 1 tells us that for any point $(\hat{n}^h, \hat{n}^l)$ on the efficient 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 efficient 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 α mixture of the two-price mechanisms that induce $(\hat{n}_1^h, \hat{n}_1^l)$ and $(\hat{n}_2^h, \hat{n}_2^l)$.[53] 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 efficient frontier. □

B.3 Proof of Theorem 2

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 implied by the strong duality relationship observed above.

[53] By a mixture of two mechanisms, we mean the mixture of the corresponding (H, L) function pairs.

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 efficient 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 efficient frontier if and only if either $a \leq b = 0$ or $b \leq a = 0$. If agents are not identical, for any λ_h, λ_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 strict 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. Recall that under strict regularity, S^j is strictly convex over the set of (a, b) that such that $N_j^*(a, b)$ is on the interior of the efficient frontier. Thus this solution must be unique.

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 clearly 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 $j \in A \cap B'$. 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.

B.4 Proof of Theorem 3

Proof. We begin with some preliminary comparative statics observations.

Lemma 4. 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. (2) has increasing differences in $c^k(j)$ and $\hat{n}_j^k$. The second case is immediate from Equation (3) and the definition of S^j. □

Given Lemma 4, 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. 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 $\lambda_h - c^h(j) > \lambda_h - c^h(j')$ and $\lambda_l - c^l(j) = \lambda_l - c^l(j')$ for all λ_h, λ_l. Let $(\lambda_h^*, \lambda_l^*)$ be the solution to the dual in eq. (3). The prices p_1^j, p_2^j defining the optimal LMS-TP mechanism solve

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

Then then solutions (p_1^j, p_2^j) and $(p_1^{j'}, p_2^{j'})$ satisfy $p_1^j > p_1^{j'}$ if and only if $p \mapsto \frac{F_j(p)}{p}$ is increasing, which is equivalent to the condition $p f_j(p) \geq F_j(p)$. The payoff of agent j is

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

By the envelope theorem (Milgrom and Segal, 2002), if $p_j \leq p_1^j$ then the right derivative of the workload with respect to p_1^j is $(p_1^j)^{-2} p_j n^l > 0$. This proves local fairness for j.

Consider now the case of $p_j \geq p_2^j$. We first conclude from the assumption that $p f(p) \geq (1 - F(p))$ that p_2^j is decreasing in $c^l(j)$. The remainder of the proof is symmetric to the case of $p_j \leq p_1^j$.

Finally, if $p_j \in (p_1^j, p_2^j)$ then the agent's welfare is invariant to local perturbations of $c^h(j), c^l(j)$.

The claim regarding the local incentive compatibility of the mechanism follows from the same comparative statics. The only caveat is that it does not apply to remedial investigators. □

B.5 Proof of Theorem 4

Theorem 4 follows from two intermediate results.

Lemma 5. Suppose (n^h, n^l) is an extreme point of $\mathcal{F}$ given type distribution F. Then, if every agent has type distribution F, the uniquely optimal mechanism is the status-quo, and $\min_M V(F, M) = V_{SQ}$.

Proof. An assignment mechanism is a finite-support distribution on $\mathcal{F}$ which averages to (n^h, n^l). If (n^h, n^l) is an extreme point of $\mathcal{F}$, then by definition the only such distribution is the degenerate distribution on (n^h, n^l). □

Distributions F for which the virtual value is constant play an important role. Such distributions exist: the Pareto distribution on $[1, \infty)$ is an example. Note that any such distribution must have a decreasing density.

Lemma 6. If F has a constant virtual value equal to $\bar{\phi}$ then for any $a, b \geq 0$ such that $a - b < b\bar{\phi} < a$, the unique optimal mechanism which sets $H = n^h$ and $L = n^l$ for all p.

Proof. Suppose that the virtual value is constant, equal to $\bar{\phi}$. In this case for any given p^*, the optimal mechanism on $[p^*, \bar{p}]$ is the solution to

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

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

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

The unique solution is to set $H = 0$ if $a - b\bar{\phi} < 0$, and n^h if $a - b\bar{\phi} > 0$.

Consider now the choice of H on $(\underline{p}, p^*]$. The program is

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

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

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

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

Fix $H(\underline{p}) > n^h$. Ignoring the non-negative L constraint, if $a - b - b\bar{\phi} < 0$ the solution is to set $H = n^h$. We can guarantee that the non-negative L constraint is satisfied by choosing $H(\underline{p}) = n^h$.

If $a - b - b\bar{\phi} < 0$ and $a - b\bar{\phi} > 0$ then the above implies that the unique (up to zero-measure perturbations) optimal mechanism is to set $H = n^h$ for all p. □

Lemma 6 implies that (n^h, n^l) is an extreme point of $\mathcal{F}$ (in fact, it is an exposed extreme point). Then Lemma 5 implies the Theorem 4.

B.6 Proof of Proposition 1

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. (3). 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

$$
\begin{aligned}
\min_{(b_j,s_j)_{j=1}^J} \quad & \sum_{j=1}^J \hat{F}_j(p_1^j|P,y) b_j \left(C^h(j) - p_1^j C^l(j)\right) \\
& - \sum_{j=1}^J \left(1-\hat{F}_j(p_2^j|P,y)\right) s^j \left(C^h(j) - p_2^j C^l(j)\right) \\
s.t. \quad & 0 \le b_j \le \frac{n^l}{p_1^j} \quad \forall\, j \\
& 0 \le s_j \le n^h \quad \forall\, j \\
& \sum_{j=1}^J b_j \hat{F}_j(p_1^j|P,y) = \sum_{j=1}^J s_j \left(1-\hat{F}_j(p_2^j|P,y)\right) \qquad (h\text{-capacity}) \\
& \sum_{j\in\mathcal{B}} p_1^j b_j \hat{F}_j(p_1^j|P,y) = \sum_{j\in\mathcal{S}} p_2^j s_j \left(1-\hat{F}_j(p_2^j|P,y)\right) \qquad (l\text{-capacity})
\end{aligned}
$$

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 $(b_j,s_j)_{j=1}^J$ that are feasible in the above program given parameters $(\hat{F}_j^1,\hat{F}_j^2)_{j=1}^J$.

Lemma 7. R is upper and lower hemicontinuous.

Proof. Define

$$
\begin{aligned}
\varphi\left((\hat{F}_j^1,\hat{F}_j^2)_{j=1}^J\right) := \Big\{ & (b_j,s_j)_{j=1}^J : 0 \le b_j \le \frac{n^l}{p_1^j} \quad \forall\, j, \quad 0 \le s_j \le n^h \quad \forall\, j, \\
& \sum_{j=1}^J b_j \hat{F}_j(p_1^j|P,y) = \sum_{j=1}^J s_j \left(1-\hat{F}_j(p_2^j|P,y)\right) \Big\}
\end{aligned}
$$

and

$$\eta\left((\hat{F}_j^1,\hat{F}_j^2)_{j=1}^J\right) := \Big\{(b_j,s_j)_{j=1}^J : 0 \le b_j \le \frac{n^l}{p_1^j} \quad \forall\ j,\quad 0 \le s_j \le n^h \quad \forall\ j,$$
$$\sum_{j\in\mathcal{B}} p_1^j b_j \hat{F}_j(p_1^j|P,y) = \sum_{j\in\mathcal{S}} p_2^j s_j \left(1-\hat{F}_j(p_2^j|P,y)\right)\Big\}$$

so that $R=\varphi\cap\eta$. Both ϕ and η are given by the intersection of a hyperplane in $\mathbb{R}^{2J}$ with the hypercube $\{b_j,s_j)_{j=1}^J : 0 \le b_j \le \frac{n^l}{p_1^j},\ 0 \le s_j \le n^h \quad \forall\ j\}$, where the normal vector to the hyperplane is a linear function of $(\hat{F}_j^1,\hat{F}_j^2)_{j=1}^J$. Thus both φ and η are upper and lower hemicontinuous. Since both are also convex and compact valued, upper and lower hemicontinuity of $R=\varphi\cap\eta$ follows.[54] □

Given Lemma 7, 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 8. $(F_j)_{j=1}^J \mapsto V_{OPT}((F_j)_{j=1}^J|y)$ and $(F_j)_{j=1}^J \mapsto (p_1^j,p_2^j)_{j=1}^J$ are continuous.

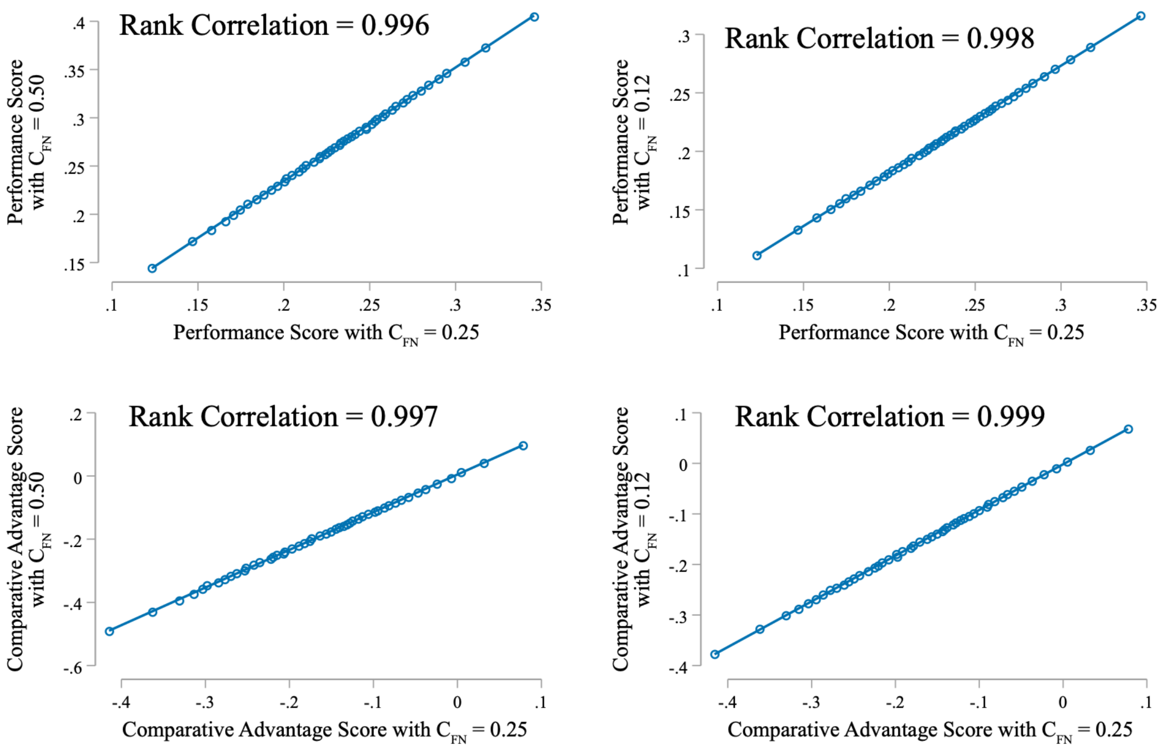
Proof. Recall that $(p_1^j,p_2^j)_{j=1}^J$ are defined from the solutions to eq. (3). S^j is continuous in F_j. Moreover, if F^j satisfies strict regularity for all j then the objective in eq. (3) is unique. The lemma follows from Berge's maximum theorem. □

Moreover, by essentially the same argument as that of Lemma 7, we can show that V_{SMS} is continuous in $(p_1^j,p_2^j)_{j=1}^J$. Combined with Lemma 8, this implies that V_{SMS} converges to V_{OPT} along any sequence of strictly regular $(F_j^n)_{j=1}^J$, as desired.

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

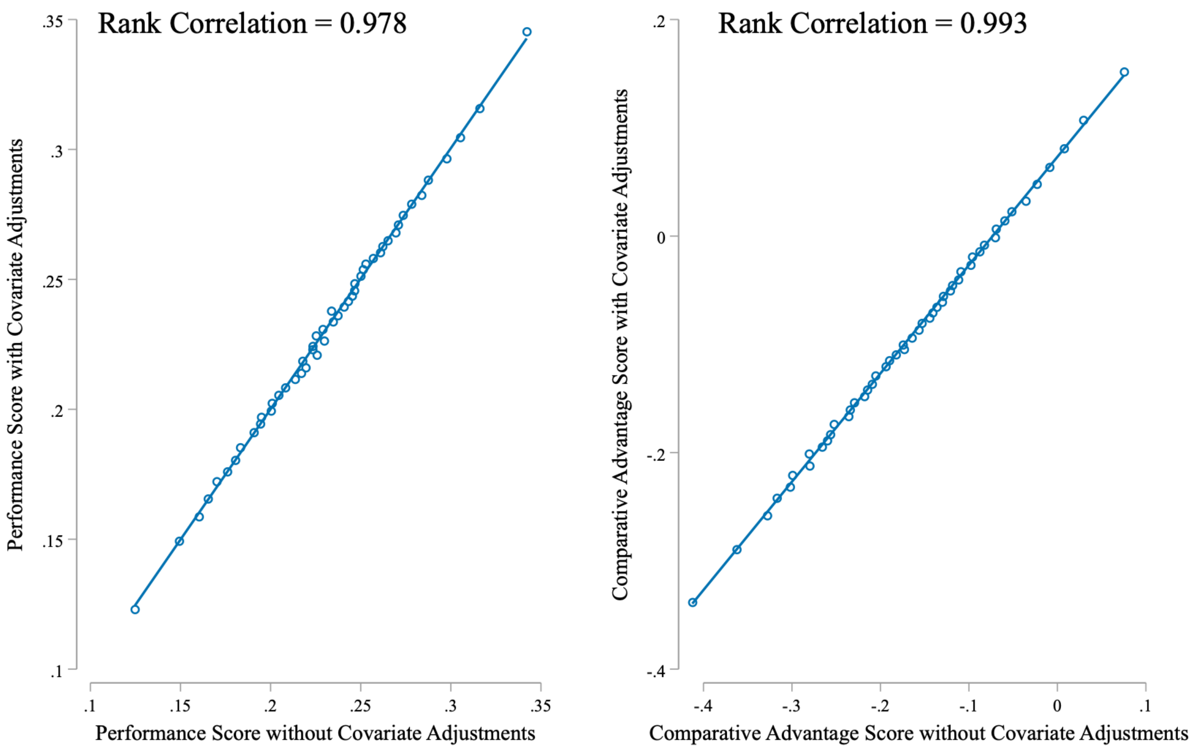
C Supplemental Figures and Tables

Figure A1: Robustness to Different Choices of Social Costs



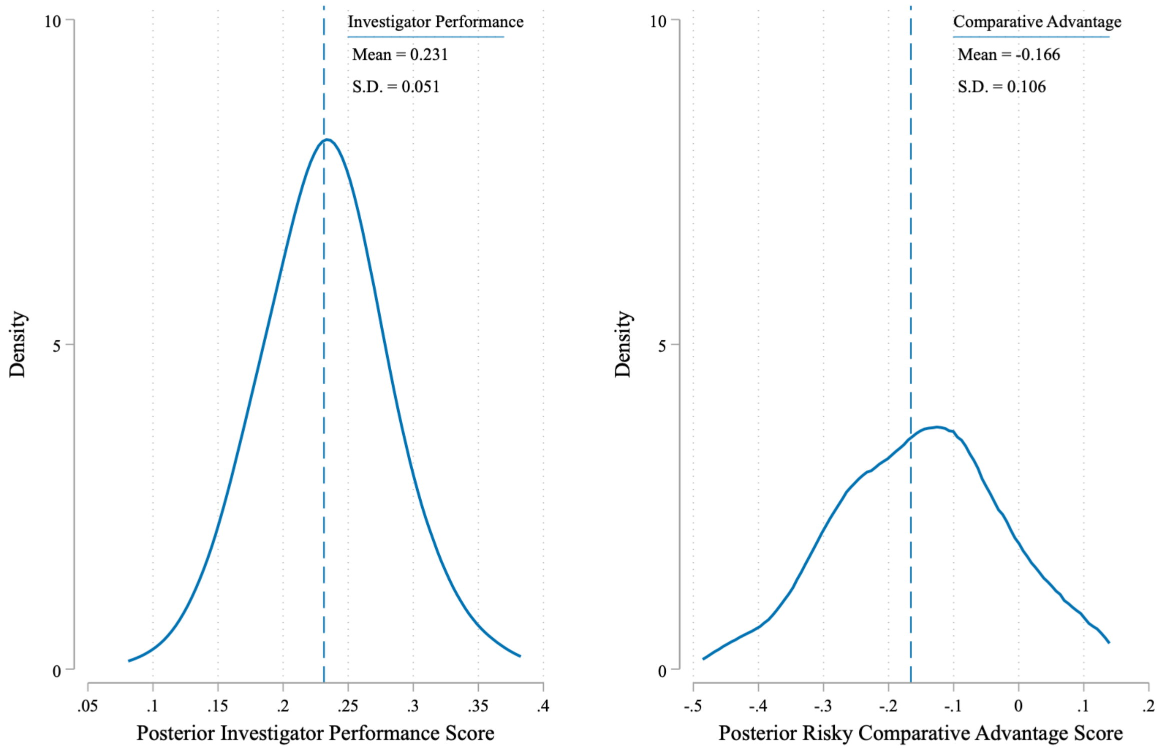
Notes. This figure plots the relationship between performance scores, γ_j, and comparative advantage scores, d_j, as we vary the choice of social costs. Benchmark estimates of γ_j and d_j are reported in the x-axis of each subfigure, with $(c_{TP}, c_{FN}, c_{FP}) = (0, 0.25, 1)$. In the left subfigures, we re-estimate γ_j and d_j with $c_{FN} = 0.50$. In the right subfigures, we re-estimate γ_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. Empirical Bayes posteriors are computed using the shrinkage procedure of Bonhomme and Weidner (2022).

Figure A2: Robustness to Covariate Adjustment



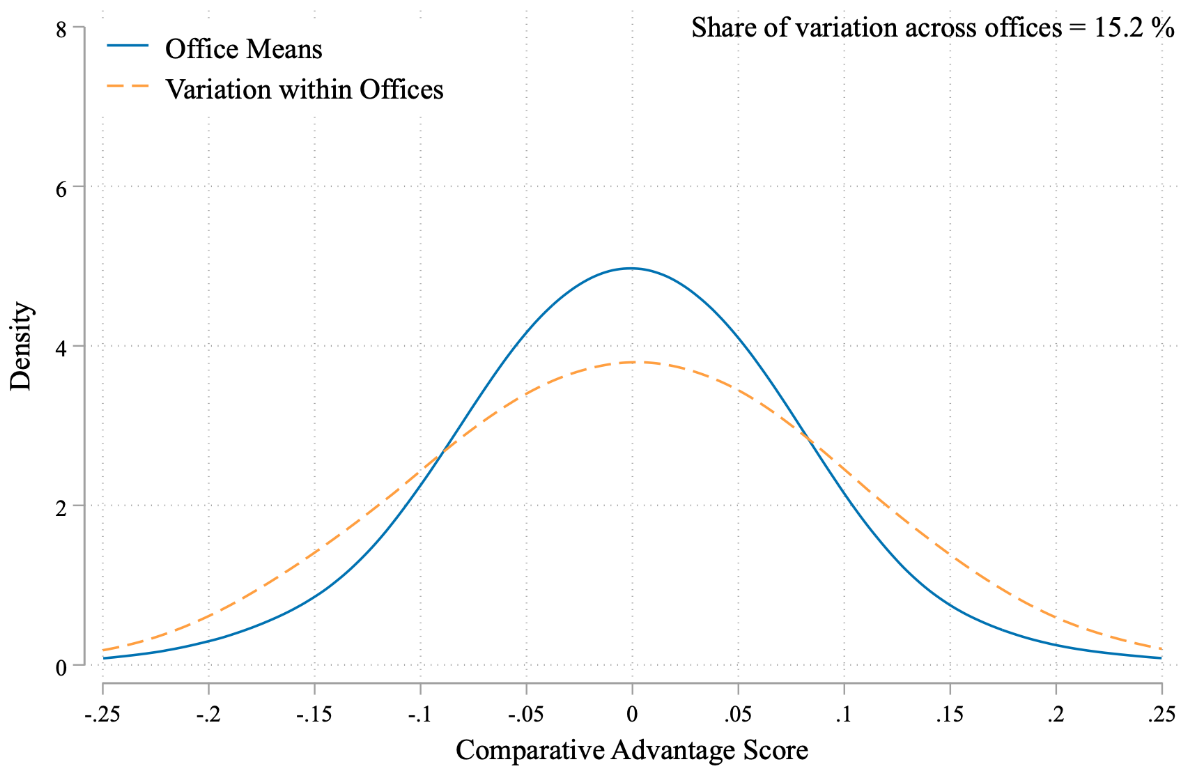
Notes. This figure plots the relationship between performance scores, γ_j, and comparative advantage scores, d_j, based on whether we include child and investigation controls. Benchmark estimates of γ_j and d_j are reported in the horizontal axis of each subfigure, which do not include child and investigation controls. We then re-estimate the regressions of placement and false negative outcomes on investigator effects, controlling for the child and investigation traits in Table A1. 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. 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. All investigator-specific estimates adjust for zipcode-by-year fixed effects. Empirical Bayes posteriors are computed using the shrinkage procedure of Bonhomme and Weidner (2022).

Figure A3: Investigator Performance and Comparative Advantage Score Distributions



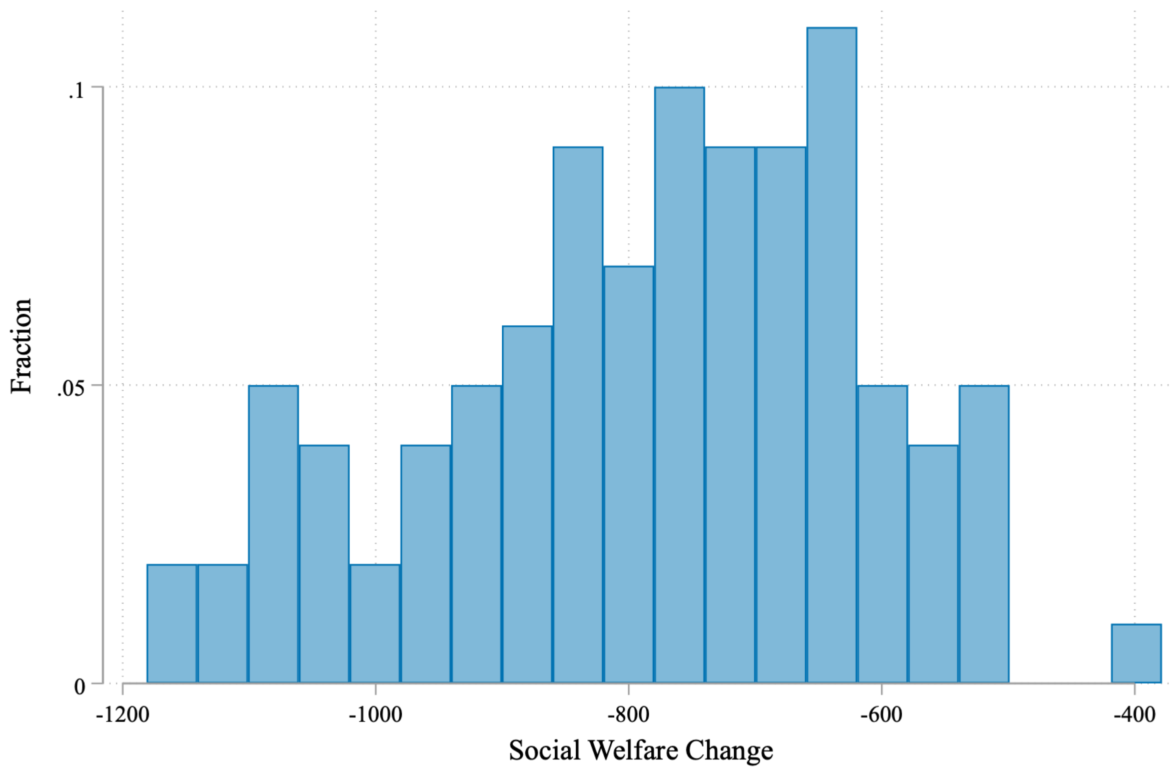
Notes. This figure plots the distribution of performance scores, γ_j, and comparative advantage scores, d_j, on high-risk cases. High-risk cases are those in the top quartile of the predicted risk distribution. The sample for the comparative advantage score distribution is limited to investigators that 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. Empirical Bayes posteriors are computed using the shrinkage procedure of Bonhomme and Weidner (2022). Means and standard deviations refer to the estimated prior distribution.

Figure A4: Within- and Across-Office Variation in Comparative Advantage



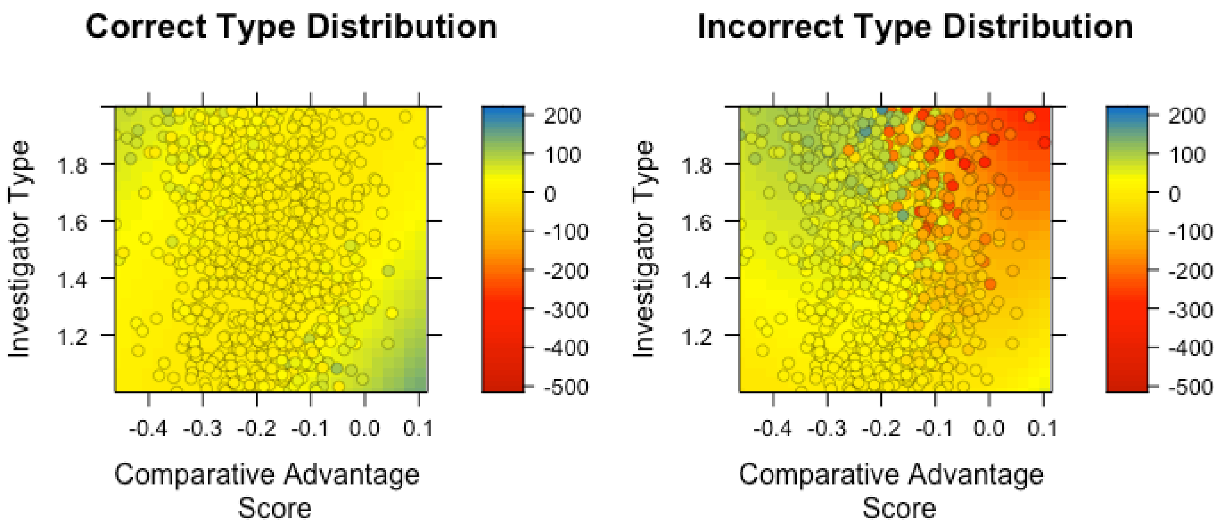
Notes. This figure decomposes the variation of the posterior distribution in the benchmark comparative advantage scores, d_j, into the share explained across or within offices (zipcode-by-year pairs). Investigator-specific estimates of subsequent maltreatment and placement rates are estimated via a regression adjustment for zipcode-by-year fixed effects. Empirical Bayes posteriors are computed using the shrinkage procedure of Bonhomme and Weidner (2022). The distribution of d_j is computed from these estimates as posterior average effects, and we compute means within each office to estimate the variation explained within or across offices.

Figure A5: Distribution of Welfare Change Estimates



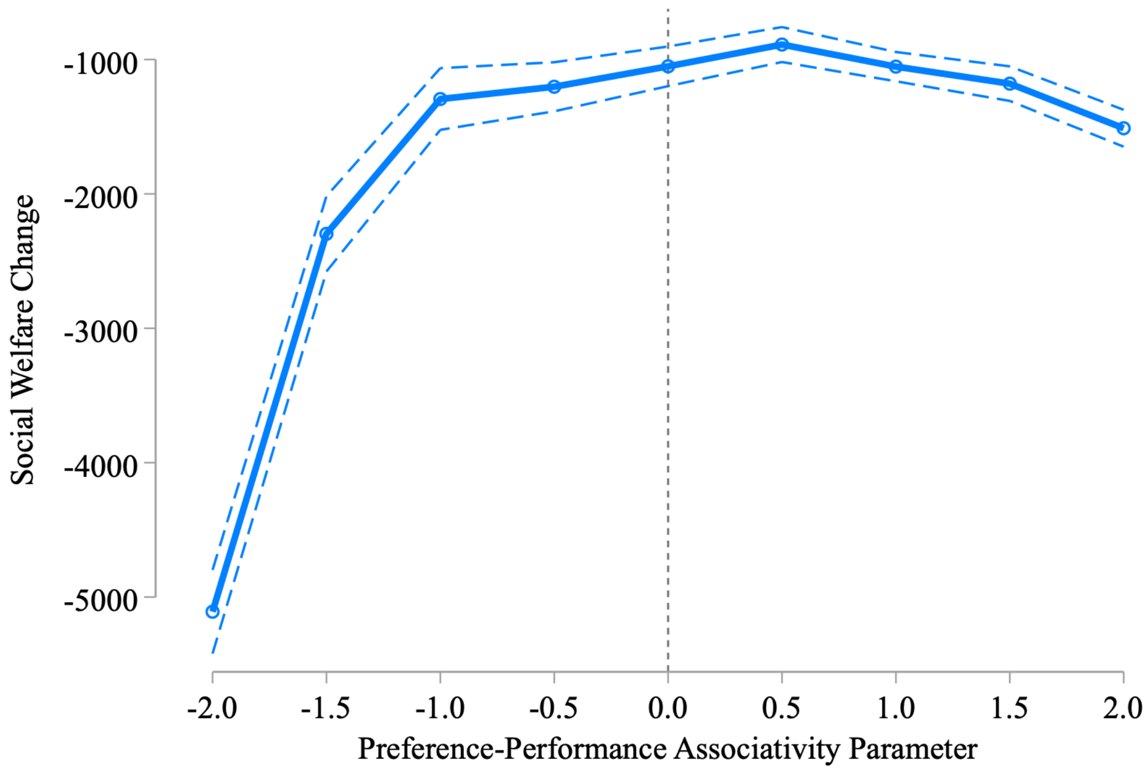
Notes. This figure presents the frequency of welfare gains for 100 draws from the type distribution. The average of this distribution is summarized in Table 1. All welfare changes are derived by applying SMD-TP mechanism, under the assumption that $p_j \sim \text{Unif}[1, 2]$.

Figure A6: The Importance of Accounting for Investigator Preferences (Single-Draw)



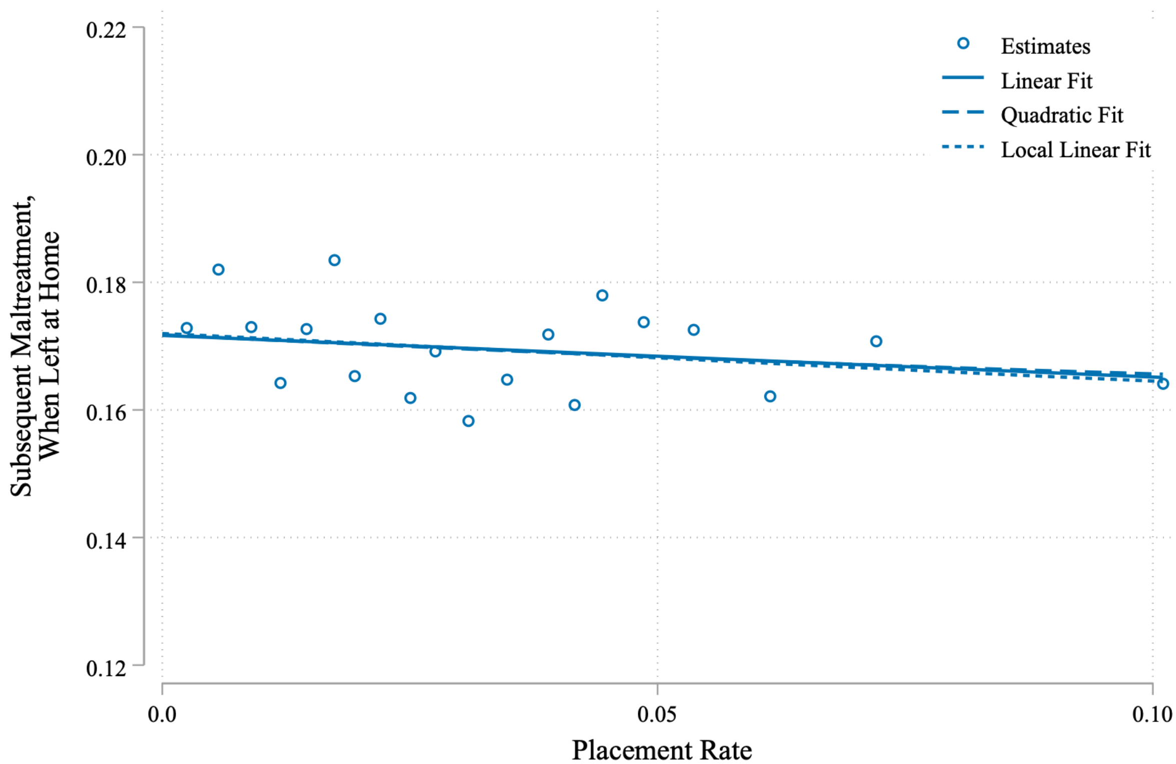
Notes. This figure plots the effect of reassignment according to the SMD-TP mechanism on investigator welfare by their comparative advantage score, d_j and their type, p_j. Investigator welfare for an investigator type p_j and assigned to caseload (n^h, n^l) is $-(p_j n^h + n^l)$. We report the difference between investigator welfare under the SMD-TP mechanism and a counterfactual in which cases are equally-split within offices. The left panel assumes that the true distribution of investigator types is $p_j \sim \text{Unif}[1, 2]$. The right panel calculates changes in investigator welfare under an SMD-TP mechanism that assumes $p_j = 1\ \forall j \in \mathcal{J}$, but where the true p_j is distributed according to $\text{Unif}[1, 2]$. We present results for one investigator-type draw, where each dot corresponds to a single investigator in the sample.

Figure A7: LMS Welfare Changes Under 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 $g(d_j)$ is defined as in the text. 95% confidence intervals are reported.

Figure A8: 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.

Table A1: Summary Statistics

Panel A: Child Socio-Demographics	
White	0.660
Black	0.199
Female	0.481
Child had a previous investigation	0.452
Number of previous investigations	1.061
Age at investigation	6.759
Panel B: Investigation Traits	
Alleged perpetrator is the mother/stepmother	0.768
Alleged perpetrator is the father/stepfather	0.346
Alleged perpetrator is a non-parent relative	0.052
Investigation included a domestic violence allegation	0.108
Investigation included an improper supervision allegation	0.534
Investigation included a medical neglect allegation	0.043
Investigation included a physical abuse allegation	0.280
Investigation included a physical neglect allegation	0.433
Investigation included a substance abuse allegation	0.174
Panel C: Outcome, if left at home	
Re-investigated for child maltreatment within 6 months	0.171
Foster care rate	0.030
Number of investigations	278,089
Number of children	225,487
Number of investigators	783

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 278,089 unique investigations of 225,487 children assigned to 783 investigators. Investigations can include multiple allegations and perpetrators, so these categories are not mutually exclusive.

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

	(1) False Negative	(2) Foster Care Placement	(3) False Positive
Panel A: Across all Cases			
Standardized	1.10***	0.51***	1.61***
Performance Score	(0.13)	(0.08)	(0.14)
Panel B: Across High-Risk Cases			
Standardized Comparative	-1.46***	0.70	-0.76
Advantage Score	(0.45)	(0.88)	(1.23)
Panel C: Across Low-Risk Cases			
Standardized Comparative	-0.06	0.91	0.84
Advantage Score	(0.25)	(0.78)	(0.62)

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, γ_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 2. 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 16.6% and 2.4% over all cases, 24.6% and 3.5% over high-risk cases, and 13.5% and 1.8% over low-risk cases. False positive rates are identified via an identification-at-infinity strategy, described in Appendix G. 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	2.036***	
(Normalized)	(0.274)	
Above Median		1.301**
High-Risk Share		(0.144)
Investigator Count	783	783

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: Welfare Gains from Observed Counterfactual

	(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) $p_j = 1$
Social Costs	-793.9***	-708.5***	-652.5***	-680.1***	-1,461.6***	-1,883.3***
	(215.6)	(204.0)	(222.6)	(205.2)	(166.2)	(150.2)
	[-4.4%]	[-3.9%]	[-3.6%]	[-3.7%]	[-8.0%]	[-10.4%]
False Negatives	-504.4***	-447.4**	-426.9***	-445.1***	-1,092.0***	-1,305.7***
	(151.8)	(178.1)	(164.4)	(161.7)	(130.5)	(121.5)
	[-1.1%]	[-1.0%]	[-0.9%]	[-1.0%]	[-2.4%]	[-2.8%]
False Positives	-690.1***	-615.4***	-565.8***	-587.7***	-1,223.7***	-1,601.5***
	(158.5)	(172.8)	(168.2)	(162.8)	(150.3)	(131.6)
	[-10.5%]	[-9.3%]	[-8.6%]	[-8.9%]	[-18.6%]	[-24.3%]
Placements	-185.7**	-168.1**	-138.9*	-142.7	-131.8*	-295.9***
	(82.1)	(81.1)	(77.8)	(101.6)	(77.2)	(76.7)
	[-2.3%]	[-2.1%]	[-1.7%]	[-1.8%]	[-1.6%]	[-3.7%]

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. Columns 1 and 2 present uniform distributions with supports $[1, 2]$ and $[1, 3]$, respectively. Columns 3 and 4 present truncated normal distributions (in $[1, 3]$) , both with a mean of 2 and standard deviations of 0.5 and 1, respectively. Column 5 presents a degenerate distribution where $p_j = 2$, and Column 6 with $p_j = 1$. 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: Welfare Gains for Counties in Balanced Sample

	(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) $p_j = 1$
Social Costs	-737.1***	-651.5***	-601.8***	-627.4***	-1,308.5***	-1,720.9***
	(176.5)	(188.6)	(195.9)	(192.3)	(148.6)	(140.3)
	[-4.7%]	[-4.1%]	[-3.8%]	[-4.0%]	[-8.3%]	[-11.0%]
False Negatives	-514.0***	-457.7***	-442.2***	-458.7***	-1,021.5***	-1,241.5***
	(170.2)	(168.7)	(164.5)	(155.2)	(93.3)	(98.3)
	[-1.3%]	[-1.1%]	[-1.1%]	[-1.1%]	[-2.5%]	[-3.1%]
False Positives	-620.8***	-546.5***	-502.2***	-522.2***	-1,080.2***	-1,442.7***
	(160.3)	(166.2)	(169.0)	(160.0)	(123.2)	(118.0)
	[-11.0%]	[-9.7%]	[-8.9%]	[-9.2%]	[-19.1%]	[-25.6%]
Placements	-106.8	-88.8	-60.1	-63.5	-58.7	-201.3***
	(78.4)	(77.3)	(76.2)	(87.0)	(63.1)	(58.8)
	[-1.5%]	[-1.2%]	[-0.8%]	[-0.9%]	[-0.8%]	[-2.8%]

Notes. This table reports the welfare gains derived from the SMD-TP mechanism, using a sample of 33 counties that appear in every sample year. Each column corresponds to a different distributional assumption for p_j. Columns 1 and 2 present uniform distributions with supports $[1, 2]$ and $[1, 3]$, respectively. Columns 3 and 4 present truncated normal distributions (in $[1, 3]$) , both with a mean of 2 and standard deviations of 0.5 and 1, respectively. Column 5 presents a degenerate distribution where $p_j = 2$, and Column 6 with $p_j = 1$. 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: Importance of Small Market and Dynamic Considerations

	(1) LMS-TP	(2) SMS-TP	(3) SMD-TP
Social Costs	-1,036.6***	-760.5***	-784.1***’
	(75.2)	(204.2)	(191.9)
	[-5.8%]	[-4.2%]	[-4.3%]
False Negatives	-723.1***	-534.3***	-547.5***
	(56.7)	(163.8)	(167.2)
	[-1.6%]	[-1.2%]	[-1.2%]
False Positives	-874.6***	-643.1***	-662.5***
	(64.1)	(162.3)	(159.0)
	[-13.6%]	[-10.0%]	[-10.3%]
Placements	-151.5***	-108.8	-115.0
	(32.2)	(79.8)	(89.0)
	[-1.9%]	[-1.4%]	[-1.4%]

Notes. This table reports the welfare gains derived from the LMS-TP, SMS-TP, and SMD-TP mechanisms, under the distribution assumption that $p_j \sim \text{Unif}[1,2]$. 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.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D Generalizing beyond binary uncertainty

As discussed in footnote 6, 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 $D_{ij} = 1$. Define PX_{ij} and NX_{ij} as in footnote 6. 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 case i to investigator j is, as in the binary case, a linear function of the joint distribution, denoted by $c(i,j)$. Lemma 2 generalizes immediately to this setting.

Lemma 9. 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 2, 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 Finite-sample adjustments to SMD-TP mechanism

To move between these 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 ε 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 10. 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 ε large yields the result. □

Thus, by adjusting ε 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.

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 an allegation of child abuse (e.g., bruises or burns) or neglect (e.g., inadequate supervision due to substance abuse). Anyone can call the hotline but the most common reporters are those who are mandated by law to do so (e.g., teachers, medical personnel, or police officers).

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.[55] 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

[55]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 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).

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.

G 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).[56] 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, Dobbie and Hull (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:

[56] We discuss how we handle conditional random assignment in Section V.

$$\mathbb{E}\left[Y_i^{\star} | D_{ij^{\star}} = 0\right] \approx \mathbb{E}\left[Y_i^{\star}\right] \tag{10}$$

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, Dobbie and Hull (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, Dobbie and Hull, 2021, 2022; Angelova, Dobbie and Yang, 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 V to extrapolate toward the unselected first moment, $\mathbb{E}[Y_i^{\star}]$. Figure A8 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 0.168 (SE=0.001). Figure A8 shows that alternative extrapolation specifications yield nearly identical point estimates.