DISCRIMINATION IN MULTI-PHASE SYSTEMS:
EVIDENCE FROM CHILD PROTECTION

E. Jason Baron
Joseph J. Doyle Jr.
Natalia Emanuel
Peter Hull
Joseph P. Ryan

ABSTRACT

We develop empirical tools for studying discrimination in multi-phase systems, and apply them to the setting of foster care placement by child protective services. Leveraging the quasi-random assignment of two sets of decision-makers—initial hotline call screeners and subsequent investigators—we study how unwarranted racial disparities arise and propagate through this system. Using a sample of over 200,000 maltreatment allegations, we find that calls involving Black children are 55% more likely to result in foster care placement than calls involving white children with the same potential for future maltreatment in the home. Call screeners account for up to 19% of this unwarranted disparity, with the remainder due to investigators. Unwarranted disparity is concentrated in cases with potential for future maltreatment, suggesting that white children may be harmed by "under-placement" in high-risk situations.

E. Jason Baron
Economics Department
Duke University
228B Social Sciences Building
Durham, NC 27708
and NBER
jason.baron@duke.edu

Joseph J. Doyle Jr.
MIT Sloan School of Management
100 Main Street, E62-516
Cambridge, MA 02142
and NBER
jjdoyle@mit.edu

Natalia Emanuel
Federal Reserve Bank of New York
33 Liberty Street
New York, NY 10045
natalia@nataliaemanuel.com

Peter Hull
Department of Economics
Box B, Brown University
Providence RI 02912
and NBER
peter_hull@brown.edu

Joseph P. Ryan
School of Social Work
University of Michigan
1080 South University Avenue
Ann Arbor, MI 48109
joryan@umich.edu

I. Introduction

Large racial disparities have been documented in many high-stakes settings—such as employment, healthcare, housing, and criminal justice—raising concerns of discrimination by individual decision-makers. At the same time, there is growing understanding that a focus on individual decisions can yield an incomplete view of discrimination. An extensive theoretical literature shows how discrimination can arise and compound across multiple decision-makers within interconnected systems (e.g., Loury, 1976; Pincus, 1996; Powell, 2008; Small and Pager, 2020). From this "systems-based" perspective, an analysis of individual discrimination by, for example, bail judges, may understate the true level of inequity in pretrial release decisions by failing to account for previous discrimination by police officers. Broader analyses of how discrimination arises and perpetuates across such multi-phase systems may be necessary to understand and form appropriate policy responses.

Measuring discrimination in multi-phase systems is challenging for several reasons, however. Raw disparities may either overstate or understate true levels of discrimination because of omitted variables bias (OVB), while conventional regression adjustment may add included variables bias (IVB) by controlling for channels of discrimination. Datasets linking multiple phases are often unavailable and may not include the kinds of exogenous variation that can help address such biases (e.g., Arnold, Dobbie, and Hull, 2022). Interpreting and integrating findings of discrimination across systems may further require new analytic tools (Bohren, Hull, and Imas, 2022).

We develop an empirical framework that overcomes these challenges and apply it to the setting of foster care placement by U.S. child protective services (CPS). CPS aims to prevent child maltreatment by investigating reported cases of abuse or neglect and placing children into foster care when deemed necessary to ensure their safety. In practice, CPS involvement is remarkably common: by age 18, 37% of children experience a maltreatment investigation and 5% spend time in foster care (Wildeman and Emanuel, 2014; Kim et al., 2017). CPS involvement is also racially disparate: the majority of Black children (53%) experience an investigation, compared to 28% of white children, and Black children are twice as likely to spend time in foster care (10%, compared to 5% of white children).

There is enormous interest in these racial disparities and the extent to which they reflect discrimination in the actions of CPS decision-makers. For example, both the United Nations and the American Bar Association recently released reports calling for the U.S. to take all appropriate measures to eliminate racial discrimination in child protection (Kelly, 2022; White and Persson, 2022). Such interest reflects the fact that CPS actions can have a

tremendous impact on the lives of children and parents and involve a difficult trade-off. On one hand, leaving children in high-risk situations may lead to subsequent maltreatment which is associated with decreased educational attainment and future earnings (Currie and Spatz Widom, 2010), and increased criminal activity (Currie and Tekin, 2012). On the other hand, foster care placement is among the most far-reaching government interventions with large potential effects on a child's educational attainment, earnings, and criminal activity (Bald et al., 2022). Discrimination in foster care placement thus stands to exacerbate inequities in many long-term outcomes.

Disparities in foster care placement can arise at two key phases within the CPS system. In an initial screening phase, incoming calls that allege child maltreatment are routed through a central state-level hotline. Call screeners at the center decide whether to "screen-in" the call: i.e., advance it to a formal investigation. In a subsequent investigation phase, screened-in cases are allocated to a regional investigator who visits the family and decides whether the child should be placed into foster care. Initial discrimination by screeners can thus be perpetuated, mitigated, or compounded by subsequent investigator decisions.[1]

We measure discrimination in this multi-phase system by unwarranted disparities (UDs): racial differences in screener and investigator decision rates, conditional on a child's potential for subsequent maltreatment in the home.[2] This measure builds on Arnold, Dobbie, and Hull (2021, 2022), who study discrimination in bail judge decisions via racial disparities in pretrial release rates conditional on a defendant's potential for pretrial misconduct. As in their context, our UD discrimination measures are natural given clear decision-maker objectives.[3] Unwarranted disparities capture classic drivers of inequity in economics such as racial bias (Becker, 1957) and statistical discrimination (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977), as well as indirect forms of discrimination arising from non-race characteristics.

The key identification challenge in measuring UDs is the selective observability of a child's potential for subsequent maltreatment in the home. Maltreatment potential is directly observed among children not removed from home, but unobserved among those placed into foster care. We address this challenge by leveraging the quasi-random assignment of both

---

[1]While there may also be discrimination among those reporting suspected child maltreatment to the state's hotline, most reporters are not CPS decision-makers. Consequently, this stage of potential discrimination falls outside the purview of CPS.

[2]Specifically, we condition on a child's potential for a subsequent maltreatment investigation in the home within six months—a common proxy in the child welfare literature (Antle et al., 2009; Putnam-Hornstein, Prindle, and Hammond, 2021). We show robustness of our main findings to a wide range of alternative proxies.

[3]Our measure differs from those in earlier studies of disparities at the investigation phase which condition on observable traits rather than future maltreatment potential. Such studies include Drake et al. (2011); Font, Berger, and Slack (2012); Putnam-Hornstein et al. (2013).

screeners and investigators to cases, building on the "identification at infinity" approach in Arnold, Dobbie, and Hull (2022). This approach generates estimates of UD in screener and investigator decisions, which we then combine to estimate overall UD in eventual placement rates. We further decompose the placement UD into the shares attributed to screeners and investigators, via a decomposition that builds on Bohren, Hull, and Imas (2022).

To build intuition for our approach, consider estimating UDs in screening-in decisions and imagine a randomly assigned screener whose calls exhibit a placement rate of virtually zero (either because they screen-out virtually all calls or because virtually none of their screened-in calls result in placement). By virtue of random assignment and this low placement rate, the subsequent at-home maltreatment rates observed among Black and white children assigned to this screener are close to the average rates among all Black and white children. These race-specific maltreatment rates thus capture the correlation between maltreatment potential and race in the full population of calls, and they can be used to correct for OVB in raw screening-in disparities. Absent such a screener, we estimate these key parameters by extrapolating the at-home maltreatment rates of quasi-randomly assigned screeners with low placement rates. A similar approach can be used to estimate UDs at the investigation phase, focusing on investigators with low placement rates among screened-in children.

We implement this identification strategy using administrative data from Michigan CPS spanning 2008 to 2019 and present three key findings. First, we find significant evidence of unwarranted disparity by both screeners and investigators. Calls involving Black children are screened-in at a 5 percentage point (8%) higher rate than calls involving white children with identical potential for subsequent maltreatment. The estimated UD is roughly 35% larger than a controlled observational disparity of 3.7 percentage points, which points to the importance of accounting for the selectively observed maltreatment potential. At the investigation phase, we find that investigators amplify initial screening disparities—despite observing effectively everything seen by screeners and deliberating over a much longer time frame. Specifically, investigators place screened-in Black children at a 1.7 percentage point (50%) higher rate than screened-in white children with identical potential for future maltreatment. This estimated UD is nearly 70% larger than the corresponding controlled observational disparity.

Our second set of findings link the UDs in screener and investigator decisions via a decomposition of UD in eventual placement rates. Overall, calls involving Black children are 1.1 percentage points (55%) more likely to end up in foster care relative to calls involving white children with identical maltreatment potential. The decompositions show that screener decisions account for between 13% to 19% of overall placement UD, with investigators driving the remainder. The fact that call screeners drive a significant share of eventual UD in foster

3

care placement is somewhat surprising, since only a small share of screened-in investigations result in placement. This finding further illustrates the importance of a systems-based analysis of discrimination in high-stakes settings like CPS: our estimates show that eliminating UD in foster care placement rates may require intervention at both phases of CPS involvement.

In our third set of results, we document a striking form of heterogeneity in the UDs: the placement disparity is concentrated among children with subsequent maltreatment potential in the home, with calls involving Black children placed in foster care at twice the rate of calls involving white children in this subpopulation (8% versus 4%). In contrast, the placement disparity is small and statistically insignificant in the subpopulation of children without maltreatment potential. The finding that unwarranted disparity is concentrated among high-risk cases implies that a higher placement rate may offer relative protection to Black children. Indeed, prior work in our setting finds that both Black and white children at risk of subsequent maltreatment in the home have better outcomes when placed in foster care, including a lower likelihood of subsequent maltreatment and adult criminal justice contact along with better educational outcomes (Baron and Gross, 2022; Gross and Baron, 2022).

These findings add nuance to ongoing policy debates over the reform of CPS, which focus on the possibility that Black children are "over-placed" in foster care. While we do find that calls involving Black children with future maltreatment potential in Michigan are disproportionately placed in foster care, white children may be harmed by "under-placement" in these high-risk situations. A back-of-the-envelope calculation shows that lowering the placement rate of screened-in Black children to equalize placement rates across race would increase the Black-white adult conviction gap by 10%.

Further unpacking this third result, we show that investigators—the primary drivers of placement UD—exhibit a racial concordance effect in cases with maltreatment potential, being significantly less likely to place children of their own race than other children. Since the vast majority of investigators in Michigan are white, this concordance effect yields higher conditional placement rates for Black children. This finding suggests that the leniency afforded by white investigators to white parents may, perhaps counterintuitively, lead to worse outcomes for their children relative to Black children who are placed at higher rates.

Our findings withstand a battery of robustness checks and extensions, including qualitatively similar estimates based on non-parametric UD bounds that relax our baseline identifying assumptions. Importantly, we show that our findings are not unique to Michigan: national (though more limited) data allow us to construct non-parametric bounds on investigator UD for almost all states in the U.S. This analysis reveals that UD in low-risk cases tends to be small nationwide, while UD in high-risk cases is typically as large or larger than in Michigan.

This study contributes to several related literatures. First and foremost, we build on Arnold, Dobbie, and Hull (2022) and Bohren, Hull, and Imas (2022) by developing a practical empirical framework for studying how discrimination perpetuates and compounds across multiple decision-makers in a high-stakes system. As mentioned above, a large theoretical literature emphasizes this possibility and its implications for policy.[4] But these insights, while potentially valuable in many areas within economics, are often hard to bring to data because of the non-random decision-making at either or multiple phases of a system. We provide a framework for conducting such analyses when the multiple decision-makers are quasi-randomly assigned.[5]

Second, we add to a large literature studying the equity and efficiency of CPS systems. Our analysis of unwarranted disparity leverages a new source of variation (quasi-random screener assignment) to study the decision to launch investigations. While there is a growing literature examining the causal effects of foster care on the outcomes of screened-in children (see Bald et al. (2022) for a review), much less is known about the broader effects of CPS systems.[6]

Third, we add to a recent literature using quasi-experimental variation to estimate various notions of bias and discrimination in high-stakes decisions, such as pretrial release (e.g., Arnold, Dobbie, and Yang (2018), Hull (2021), Arnold, Dobbie, and Hull (2022), Rambachan (2022), and Canay, Mogstad, and Mountjoy (2022)), traffic stops (e.g., Goncalves and Mello (2021) and Feigenberg and Miller (2022)) and lending (e.g., Dobbie et al. (2021)). Our analysis benefits from the fact that CPS systems feature many decision-makers with very low "treatment" rates, allowing for both precise non-parametric inferences on overall UDs and the statistical power to distinguish between disparities among high- and low-risk home situations. Our framework for linking such heterogeneity to welfare considerations and policy responses may be useful in future studies of unwarranted disparity.

The rest of the paper is organized as follows. Section II. describes the CPS system in more detail. Section III. develops our UD measures at each phase of the system, as well as the UD decomposition and our identification strategies. Section IV. describes our analysis samples

---

[4]See Bohren, Hull, and Imas (2022) for a review. Our decomposition of overall placement UD builds directly on their general framework for studying "systemic" discrimination, which can arise both over sequential decisions (as in this paper) as well as across interconnected contemporaneous decisions. Bohren, Hull, and Imas (2022) study both forms, and propose a general decomposition for measuring systemic discrimination with different kinds of (quasi-)experimental data and designs.

[5]In a related analysis, Harrington and Shaffer (2022) study how racial disparities change across police and prosecutor decision-making. The paper leverages quasi-experimental variation in prosecutor decision-making, though the setting does not include such variation for police officers.

[6]A related literature studies the impacts of algorithmic decision tools within CPS: see Chouldechova et al. (2018), Brown et al. (2019), Fitzpatrick, Sadowski, and Wildeman (2022), Grimon and Mills (2022) and Rittenhouse, Putnam-Hornstein, and Vaithianathan (2022).

and presents motivating results. Section V. presents our main findings on how UD propagates through CPS. Section VI. further unpacks the main source of UD at the investigation phase, examining heterogeneity by investigator characteristics, exploring potential drivers, and summarizing several extensions including the national UD analysis. Section VII. concludes.

II. Setting

The CPS system aims to protect children from maltreatment in their home environment. Figure I summarizes the process in Michigan and most states. CPS involvement begins when a call is made to the state's central hotline to report suspected child abuse (e.g., bruises or burns) or neglect (e.g., improper supervision due to parental substance abuse). Anyone can make a report to the hotline, though the most common reporters are educators and law enforcement personnel (Benson, Fitzpatrick, and Bondurant, 2022).

Calls to the Michigan CPS hotline are answered by screeners in two central offices—one in Grand Rapids and one in Detroit—which share a hotline number. Calls typically last about 15 minutes. Screeners have substantial discretion in whether to screen-in a call, though they follow general guidelines in screening-out calls that do not conform with state law and guidance from Michigan's Department of Health and Human Services (MDHHS). Screeners are instructed to screen-in calls to minimize the likelihood of subsequent maltreatment if the call is screened-out.[7] Screeners play no other role in the process: if a call is screened-in (roughly 60% of all calls), it is sent to the alleged victim's local child welfare office for formal investigation. A screened-out call concludes MDHHS involvement. Screeners do not systematically learn the eventual outcome of a given investigation or screened-out call.

Screened-in calls are assigned to an investigator who 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 then makes two primary decisions. First, they must decide whether there is enough evidence to substantiate the allegation. This determination is based on interviews with the child maltreatment reporter, family members, police, and potentially medical reports. Around 75% of all screened-in cases in our sample are unsubstantiated; an unsubstantiated finding concludes the investigation. If the investigation is substantiated, the investigator makes a judgement on whether to place the child in foster care.[8]

---

[7]For example, the MDHHS *Children's Protective Services Policy Manual* reads: "The intake process is focused on initial fact gathering and evaluation of information to determine the validity of the referral, whether it meets statutory criteria for investigation, and to assess the level of risk to the child" (MDHHS (2020), p.35).

[8]Investigators have immense discretion over the placement decision. While there is a standardized 22-question risk assessment form in Michigan that helps determine whether placement is appropriate, many of the questions are inherently subjective and ethnographic research suggests investigators often manipulate

Under CPS guidelines, the primary justification for placement is child safety: investigators are instructed to place the child in foster care if the child is in "imminent risk" of maltreatment in the home, but to otherwise keep the child with their family.[9] If the investigator determines that the potential for subsequent maltreatment is high, she requests that her supervisor submit a court petition to place the child in foster care. In practice, it is rare for either the supervisor or the judge to disagree with the investigator's recommendation.

Approximately 3.5% of all screened-in cases in our sample result in foster care placement. In these cases the child is placed with either an unrelated foster family, relatives, or (much less frequently) in a group home, while their custodial parents receive services to support reunification. On average, children spend approximately 17 months in foster care. Following this spell in foster care, roughly 50% of children are reunified with their birth parents, 34% are adopted or have legal guardianship transferred, 9% percent exit the system as independent adults upon turning 18; remaining children fall into less common exit categories such as informal guardianship with relatives (Gross and Baron, 2022).

Regardless of the placement decision, investigators can formally open a CPS case and recommend "targeted services" to support the family in cases that have been substantiated. These services are typically preventative referrals and range from substance abuse to parenting classes, though parents are not typically compelled to use them. In 6.8% of all investigations in our sample, investigators opened a case and recommended targeted services without removing the child from their home. Anecdotally, takeup of such service recommendations is very low.[10]

Importantly for our analysis of this system, both screeners and investigators are quasi-randomly assigned. Incoming calls enter a queue, with the hotline system routing each call to the available screener who has been waiting the longest since her last call; this makes screener assignment as-good-as-random conditional on the exact day and shift. Once referred to a local office, screened-in calls are quasi-randomly assigned to the office's investigators. Every county in Michigan has at least one local office, with some larger and

---

responses to match their priors (Gillingham and Humphreys, 2010; Bosk, 2015). Moreover, the tool itself allows for discretionary adjustments to risk levels. As we show below, there is a strong first-stage relationship between investigator removal tendencies and foster care placement within offices, suggesting that even where investigators are guided by the risk assessment, they maintain significant discretion over placement.

[9]For example, the MDHHS *Children's Protective Services Policy Manual* reads: "placement of children out of their homes should occur only if their well-being cannot be safeguarded with their families" (MDHHS (2020), p.3). It further instructs 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).

[10]While our data do not contain information on the takeup of these services, we have reviewed data from MDHHS on the takeup rates of SafeCare, a parental education program commonly offered to families and a program that the state is currently evaluating. Takeup rates for this program typically range from 5% to 10%. Completion rates are significantly lower, on the order of 1% to 2%.

more urban counties containing multiple offices. Some offices further split investigators into geographic-based teams. Within teams, the assignment of most cases is rotational—reports cycle through investigators based on who is next up in the rotation and investigators are not assigned based on their specific characteristics or skill sets. There are two exceptions to this quasi-random assignment: cases of sexual abuse tend to be assigned to more experienced investigators, and repeat reports involving a child who was recently investigated are often re-assigned to the initial investigator. We exclude these cases from our analysis and isolate quasi-random assignment by conditioning on a child's ZIP code and investigation year.

III. Empirical Approach

III.A. Unwarranted Disparity Measures and Decompositions

We formalize our empirical approach by considering a population of cases referred to CPS. Each case i involves either a Black or white child, indicated by $R_i \in \{b, w\}$. Each child has a potential for at-home future maltreatment $Y_i^* \in \{0, 1\}$, with $Y_i^* = 1$ indicating future abuse or neglect when the child is not removed from the home. Cases are first handled by call screeners who decide whether to advance the case to investigation. Among screened-in cases, investigators then decide whether to place the child into foster care. If the child is either screened-out or screened-in but not placed, maltreatment potential is realized and observable via their maltreatment outcomes. Otherwise, Y_i^* is not realized and hence not observable.

To develop our discrimination measures and decompositions in this setting, we first imagine that cases are assigned to a single representative screener and, when screened-in, a single representative investigator. Decision-maker heterogeneity will play a central role in our identification strategy, developed below. Here we abstract away from heterogeneity to ease notation, letting $S_i \in \{0, 1\}$ indicate whether case i is screened-in by the representative screener and (when screened-in) $D_i \in \{0, 1\}$ indicate whether the representative investigator chooses foster care placement for the child. The product of these indicators, $P_i = S_i D_i$, then indicates whether incoming case i ultimately results in placement.

We measure discrimination in screening and investigation decisions as unwarranted disparities (UDs): racial differences in decision rates conditional on a child's potential for subsequent maltreatment in the home, Y_i^*. This measure builds on Arnold, Dobbie, and Hull (2021, 2022), who study discrimination in bail decisions via racial disparities in pretrial release rates conditional on a defendant's potential for pretrial misconduct. As in their context, our UD discrimination measure is natural given clear decision-maker objectives: under CPS guidelines, the primary justification for foster care placement is a potential for subsequent maltreatment

8

in the home.Arnold, Dobbie, and Hull (2021, 2022) show how such a measure aligns with the legal theory of disparate impact, economic notions of discrimination among equally productive workers, as well as more recent notions of algorithmic discrimination from the computer science literature.[11] Importantly, unwarranted disparity can arise from both "direct" discrimination on the basis of race itself and "indirect" discrimination through non-race characteristics that are correlated with race (such as poverty levels).

To build up to our UD measures, we first define a conditional disparity in screening-in rates among Black and white children without future maltreatment potential:

$$\Delta_0^S = E[S_i \mid R_i = b, Y_i^* = 0] - E[S_i \mid R_i = w, Y_i^* = 0], \tag{1}$$

along with the corresponding disparity among children with future maltreatment potential:

$$\Delta_1^S = E[S_i \mid R_i = b, Y_i^* = 1] - E[S_i \mid R_i = w, Y_i^* = 1]. \tag{2}$$

We measure the overall screener UD by averaging these two conditional disparities,

$$\Delta^S = \Delta_0^S(1 - \bar{\mu}) + \Delta_1^S\bar{\mu}, \tag{3}$$

with weights given by the average future maltreatment risk in the population of all cases, $\bar{\mu} = E[Y_i^*]$. Thus, Δ^S captures the expected level of UD when encountering a representative pool of children with unknown future maltreatment potential.

Next, among screened-in cases, we define the investigator's UDs:

$$\Delta_0^D = E[D_i \mid R_i = b, Y_i^* = 0, S_i = 1] - E[D_i \mid R_i = w, Y_i^* = 0, S_i = 1] \tag{4}$$
$$\Delta_1^D = E[D_i \mid R_i = b, Y_i^* = 1, S_i = 1] - E[D_i \mid R_i = w, Y_i^* = 1, S_i = 1]. \tag{5}$$

Here, Δ_0^D gives the investigator's placement rate disparity for screened-in cases without future maltreatment potential while Δ_1^D gives the screened-in placement rate disparity for cases with future maltreatment potential. We again measure the overall investigator UD by averaging these two conditional disparities: $\Delta^D = \Delta_0^D(1 - \bar{\mu}^{S=1}) + \Delta_1^D\bar{\mu}^{S=1}$, where now the weights $\bar{\mu}^{S=1} = E[Y_i^* \mid S_i = 1]$ correspond to the average maltreatment risk of screened-in cases.

---

[11]Disparate impact is one of two main legal doctrines of discrimination in U.S. case law. It concerns the discriminatory effects of a policy or practice rather than a decision-maker's intent. The disparate impact standard applies to programs and activities receiving federal financial assistance via Title VI of the 1964 Civil Rights Act, including the child protection systems we consider (DHHS, 2016; DOJ, 2016). Both screening and investigation are explicitly required to comply with this standard (DHHS, 2016). See Section I.A of Arnold, Dobbie, and Hull (2022) for more background and discussion of relevant case law.

Finally, we define placement rate disparities across all cases referred to CPS:

$$\Delta_0^P = E[P_i \mid R_i = b, Y_i^* = 0] - E[P_i \mid R_i = w, Y_i^* = 0] \tag{6}$$

$$\Delta_1^P = E[P_i \mid R_i = b, Y_i^* = 1] - E[P_i \mid R_i = w, Y_i^* = 1], \tag{7}$$

with the overall placement UD again given by an average: $\Delta^P = \Delta_0^P(1 - \bar{\mu}) + \Delta_1^P \bar{\mu}$.

We link the screener, investigator, and placement UDs via a decomposition of placement UDs into components due to the screener and investigator. Specifically, since $P_i = S_i D_i$, we have:

$$\underbrace{\Delta_y^P}_{\text{Placement UD}} = E[S_i \mid R_i = b, Y_i^* = y]E[D_i \mid S_i = 1, R_i = b, Y_i^* = y]$$

$$- E[S_i \mid R_i = w, Y_i^* = y]E[D_i \mid S_i = 1, R_i = w, Y_i^* = y]$$

$$= \underbrace{\Delta_y^S \omega_y^S}_{\text{Screener component}} + \underbrace{\Delta_y^D \omega_y^D}_{\text{Investigator component}}, \tag{8}$$

for $y \in \{0, 1\}$, where $\omega_y^S = E[D_i \mid R_i = w, Y_i^* = y, S_i = 1]$ and $\omega_y^D = E[S_i \mid R_i = b, Y_i^* = y]$.

Equation (8) decomposes placement UD into two components: one involving screener UD (Δ_y^S) and the other involving investigator UD (Δ_y^D). These UDs are weighted by ω_y^S (the placement rate of screened-in white children with $Y_i^* = y$) and ω_y^D (the screened-in rate of Black children with $Y_i^* = y$), respectively. These weights are non-negative, though they do not sum to one in general. The equation is derived similarly to classic Kitagawa–Oaxaca-Blinder (KOB) decompositions, by adding and subtracting (8) $E[D_i \mid R_i = w, Y_i^* = y, S_i = 1]E[S_i \mid R_i = b, Y_i^* = y]$ to the first line and rearranging terms. The decomposition of Δ_0^P and Δ_1^P can be further averaged with the $\bar{\mu}$ weights to write Δ^P in terms of Δ_y^S and Δ_y^D. These decompositions build on Bohren, Hull, and Imas (2022), who propose a general framework for studying how discrimination perpetuates across multiple connected decisions across time or domains.

As usual with KOB decompositions, an alternative version of Equation (8) comes from changing the "order" of decomposition. Namely, we can also write:

$$\Delta_y^P = \Delta_y^S \tilde{\omega}_y^S + \Delta_y^D \tilde{\omega}_y^D, \tag{9}$$

where $\tilde{\omega}_y^S = E[D_i \mid R_i = b, Y_i^* = y, S_i = 1]$ and $\tilde{\omega}_y^D = E[S_i \mid R_i = w, Y_i^* = y]$ by adding and subtracting $E[D_i \mid R_i = b, Y_i^* = y, S_i = 1]E[S_i \mid R_i = w, Y_i^* = y]$ to the first line of Equation (8) and again rearranging terms. Again, this alternative decomposition can be averaged with the $\bar{\mu}$ weights to write Δ^P in terms of Δ_y^S and Δ_y^D.

Either decomposition—Equation (8) or (9)—can be used to study how UD arises and propagates through the two-phase child protection system. Since placement rates tend to be low, the screener disparity weights ω_y^S and $\tilde{\omega}_y^S$ are likely to be small; the screener components $\Delta_y^S \omega_y^S$ and $\Delta_y^S \tilde{\omega}_y^S$ are therefore not likely to account for the majority of placement UD. Nevertheless, to the extent that meaningful unwarranted disparity is found at both the screener and investigator phase, Equations (8) and (9) show how much intervention may be required in each phase to eliminate unwarranted disparity in eventual foster care placement rates.

The fundamental challenge in bringing these UD measures and decompositions to data is the selective observability of maltreatment potential. Among children who are not placed into foster care, maltreatment potential is directly revealed by their future maltreatment outcomes. But since future maltreatment in the home is unobserved among children who are placed in foster care, we cannot directly estimate Equations (1)-(9). We next develop our quasi-experimental strategy for addressing this identification challenge.

III.B.   Identification Strategies

The first step of our empirical approach is to rewrite Equations (1)–(9) in terms of a set of directly estimable moments and a set of unknown parameters capturing the average at-home maltreatment risk of certain populations of Black and white children. The second is to estimate or bound these key parameters using the quasi-random assignment of screeners and investigators. Here we first develop this approach for estimating the screener UD measures. We then discuss how the approach extends to the investigator UD measures and decompositions of placement UDs; we also discuss an alternative bounding approach.

The screener UD measures (1)-(2) are based on terms that can be rewritten as:

$$E[S_i \mid R_i = r, Y_i^* = 1] = 1 - \frac{E[(1 - S_i)Y_i^* \mid R_i = r]}{E[Y_i^* \mid R_i = r]} \tag{10}$$

and

$$E[S_i \mid R_i = r, Y_i^* = 0] = 1 - \frac{E[(1 - S_i)(1 - Y_i^*) \mid R_i = r]}{1 - E[Y_i^* \mid R_i = r]}, \tag{11}$$

Since screening decisions S_i are directly observed, and since the future maltreatment outcomes of screened-out children (with $(1 - S_i) = 1$) directly reveal their at-home maltreatment potential, the numerators in these expressions do not suffer from the selective observability of Y_i^*. The challenge of estimating screener UDs therefore reduces to the challenge of estimating the parameters in the denominators, $\mu_b = E[Y_i^* \mid R_i = b]$ and $\mu_w = E[Y_i^* \mid R_i = w]$.

11

These parameters reflect the average at-home maltreatment risk of Black and white children in the full population of cases, and they are not directly estimable because of the selective observability of Y_i^*. Specifically, we cannot directly estimate each μ_r because Y_i^* is unobserved when child i is placed into foster care.

Our estimation strategy for the key μ_r parameters builds on Arnold, Dobbie, and Hull (2021, 2022) by leveraging variation across as-good-as-randomly assigned screeners in the observed maltreatment rates of children not placed into foster care. To build intuition for this approach, suppose that in addition to our representative screener we have identified a "supremely negligent" screener whose calls exhibit a placement rate of virtually zero (either because they screen-out virtually all calls or because virtually none of their screened-in calls result in placement). Suppose further that we randomly assign some subset of cases to this supremely negligent screener and then measure the foster care placement rates and subsequent maltreatment outcomes among cases not resulting in foster care placement. By virtue of randomization, the average Y_i^* among Black and white children in the cases handled by the supremely negligent screener would be the same as in the full population—that is, both would have the same race-specific subsequent maltreatment means, μ_r. Moreover, the observed maltreatment outcomes of the handled cases which do not result in foster care placement would be close to these means. This follows because the supremely negligent screener has a placement rate that is close to zero, making the assigned Black and white cases not resulting in foster care placement close to representative of the full sample of Black and white cases.

Absent such a supremely negligent screener, the key mean risk parameters μ_r can be estimated by extrapolating variation in observed at-home maltreatment rates across quasi-randomly assigned screeners with low placement rates. This approach is conceptually similar to how average potential outcomes at a treatment cutoff can be extrapolated from nearby observations in standard regression discontinuity designs. Here, potential maltreatment risk is extrapolated from quasi-randomly assigned screeners with low placement rates to the hypothetical supremely negligent screener whose placement rate is zero.[12]

Formalizing this approach to estimating screener UDs requires enriched notation to capture screener heterogeneity and an estimation strategy that allows for the fact that screeners are only conditionally as-good-as-randomly assigned. Consider two linear regressions:

$$S_i = \sum_j \phi_{jw}^S (1 - B_i) Z_{ij} + \sum_j \phi_{jb}^S B_i Z_{ij} + X_i' \gamma^S + \varepsilon_i^S, \tag{12}$$

---

[12]This approach builds on a long literature on "identification at infinity" in sample selection models (Chamberlain, 1986; Andrews and Schafgans, 1998; Heckman, 1990).

estimated in the full population of cases, with $S_i \in \{0, 1\}$ here denoting screen-in status, and:

$$Y_i^* = \sum_j \psi_{jw}^S (1 - B_i) Z_{ij} + \sum_j \psi_{jb}^S B_i Z_{ij} + X_i' \lambda^S + \nu_i^S \tag{13}$$

estimated only among screened-out cases (with $S_i = 0$). Here, $B_i = \mathbf{1}[R_i = b]$ indicates that case i involves a Black child, $Z_{ij} \in \{0, 1\}$ indicates that case i is assigned to one of J screeners j, and X_i is a vector of day-by-shift fixed effects, conditional on which screener assignment is as-good-as-random. When X_i is de-meaned, the ϕ_{jw}^S and ϕ_{jb}^S coefficients capture strata-adjusted screen-in rates of screener j for white and Black children, respectively. Likewise, the ψ_{jw}^S and ψ_{jb}^S coefficients capture strata-adjusted maltreatment rates among the white and Black children screened-out by screener j. Ordinary least square (OLS) estimates of these coefficients can be used to estimate screener-specific versions of the numerators in Equations (10) and (11).[13]

For estimating the remaining risk parameters μ_r, consider two analogous regressions:

$$P_i = \sum_j \phi_{jw}^P (1 - B_i) Z_{ij} + \sum_j \phi_{jb}^P B_i Z_{ij} + X_i' \gamma^P + \varepsilon_i^P, \tag{14}$$

estimated in the full population of cases, with $P_i \in \{0, 1\}$ here denoting placement status, and:

$$Y_i^* = \sum_j \psi_{jw}^P (1 - B_i) Z_{ij} + \sum_j \psi_{jb}^P B_i Z_{ij} + X_i' \lambda^P + \nu_i^P, \tag{15}$$

estimated only among cases not resulting in foster care placement (with $P_i = 0$). Here, the ϕ_{jr}^P coefficients can be used to identify screeners with low placement rates among children of race r, while the ψ_{jr}^P coefficients capture the subsequent maltreatment rates among children of race r who are not placed into foster care following assignment to screener j. Following the above intuition, μ_r can be estimated by extrapolating estimates of ψ_{jr}^P across screeners with low ϕ_{jr}^P estimates; for example, one could estimate μ_r by the vertical intercept of linear (or more flexible) regressions of the ψ_{jr}^P estimates on the ϕ_{jr}^P estimates across screeners j.

With estimates of μ_r and the screener-specific numerators in Equations (10)-(11),

---

[13]Formally, $E[(1 - S_{ij})Y_i^* \mid R_i = r] = (1 - \phi_{jr}^S)\psi_{jr}^S$ and $E[(1 - S_{ij})(1 - Y_i^*) \mid R_i = r] = (1 - \phi_{jr}^S)(1 - \psi_{jr}^S)$, where S_{ij} denotes the potential screening decision of screener j when assigned to case i. These expressions follow when screeners are quasi-randomly assigned (so the Z_{ij} are independent of $(R_i, Y_i^*, (S_{ij})_{j=1}^J)$ given X_i) and when regression-adjustment for X_i is sufficient (see Arnold, Dobbie, and Hull (2022)). We show below that our results are nearly identical without regression-adjustment, however. This is because variation in the relevant outcome variables is largely driven by variation within, rather than across, the strata in X_i (Baron and Gross, 2022).

screener-specific UD measures Δ_{j0}^S and Δ_{j1}^S can be estimated following Equations (1) and (2). These can then be aggregated across screeners, weighting by caseloads, to estimate average screener UDs Δ_0^S and Δ_1^S. These can then be averaged, weighting by the natural estimate of $\bar{\mu} = \mu_b Pr(R_i = b) + \mu_w Pr(R_i = w)$, to obtain an estimate of overall average screener UD.

The same approach can be used to estimate investigator UDs, placement UDs in the full sample of cases, and the two UD decompositions. For investigator UDs, which condition on screened-in cases, the key parameters $\mu_r^{S=1} = E[Y_i^* \mid R_i = r, S_i = 1]$ capture the race-specific maltreatment risk of screened-in cases; this follows analogously to Equations (10)-(11). These $\mu_r^{S=1}$ can again be estimated by extrapolating estimates from versions of Equations (14)-(15) that condition on $S_i = 1$ (such that $P_i = D_i$, the investigator decision) and which replace the screener assignment dummies Z_{ij} and controls X_i with corresponding investigator assignment dummies and rotation (ZIP code by year) fixed effects. These regressions also yield estimates of the remaining moments in the analogous equations to Equations (10)-(11), giving estimates of investigator-specific UD measures Δ_{j0}^D and Δ_{j1}^D (following Equations (4) and (5)), average investigator UD estimates that aggregate these across investigators (weighting by caseloads), and an overall measure of average investigator UD that weights by the natural estimate of $\bar{\mu}^{S=1} = \mu_b^{S=1} Pr(R_i = b \mid S_i = 1) + \mu_w^{S=1} Pr(R_i = w \mid S_i = 1)$. Finally, average placement UDs in the full population of calls (and their decompositions) can be estimated by combining estimates of average screener and investigator UDs with resulting estimates of the decomposition weights: either (ω_y^S, ω_y^D) or $(\tilde{\omega}_y^S, \tilde{\omega}_y^D)$.

An alternative strategy of bounding unwarranted disparity—which does not rely on any extrapolation—uses the fact that the selectively observed Y_i^* is binary. Specifically, $\mu_r \in [\mu_{rL}, \mu_{rU}]$ where $\mu_{rL} = E[(1 - P_i)Y_i^* \mid R_i = r]$ and $\mu_{rU} = 1 - E[(1 - P_i)(1 - Y_i^*) \mid R_i = r]$. Similarly, $\mu_r^{S=1} \in [\mu_{rL}^{S=1}, \mu_{rU}^{S=1}]$ where $\mu_{rL}^{S=1} = E[(1 - P_i)Y_i^* \mid R_i = r, S_i = 1]$ and $\mu_{rU}^{S=1} = 1 - E[(1 - P_i)(1 - Y_i^*) \mid R_i = r, S_i = 1]$. The lower bounds μ_{rL} and $\mu_{rL}^{S=1}$ come from assuming all children placed into foster care, for whom future at-home maltreatment is unobserved, have $Y_i^* = 0$. The upper bounds μ_{rU} and $\mu_{rU}^{S=1}$ come from assuming all such children have $Y_i^* = 1$. These bounds are directly estimable from observed placement and maltreatment rates and can be tightened by focusing on screeners or investigators with lower placement rates (Arnold, Dobbie, and Hull, 2022). Bounds on the different UD measures and decompositions can then be formed by combining these μ_r and $\mu_r^{S=1}$ bounds with the directly estimable moments. Note that the width of these bounds will be equal to the either the placement rate in the full population of calls (when bounding μ_r), or the placement rate in the population of screened-in calls (when bounding $\mu_r^{S=1}$). Therefore, these bounds are likely to be informative in our context, given relatively low placement rates. We explore this alternative approach below.

14

III.C.   *Identifying Assumptions*

Our empirical strategy relies on two primary assumptions. First, we require the as-good-as-random assignment of screeners and investigators who vary in their tendency to place children into foster care. This assumption is consistent with our understanding of the assignment process in Michigan CPS, detailed in Section II., as well as previous studies using such variation to estimate the causal effects of foster care placement. Below, we show a variety of balance tests that indirectly validate as-good-as-random assignment in our data.

The second key assumption is an exclusion restriction: that screener and investigator assignment only systematically impact subsequent maltreatment potential through the decision to place a child into foster care. If agents differentially affect maltreatment potential in other ways, then our estimates of average maltreatment rates may reflect these additional impacts and may not be representative of the full population. This assumption is consistent with the relatively narrow mandates of screeners and investigators described in Section II.. Nevertheless, we develop below a series of extensions that test and relax the exclusion restriction by allowing for different possible direct effects of investigation, such as those from targeted services or unobserved contact effects.

Importantly, although our approach can be understood as leveraging quasi-random screener and investigator assignment as instruments for placement—similar to conventional instrumental variables (IV) studies of foster care effects—it does not require the conventional IV assumption of first-stage monotonicity: i.e., that decision-makers have a common ranking of cases by their appropriateness for foster care placement.[14] Intuitively, the extrapolated estimates of μ_r and $\mu_r^{S=1}$ are valid as long as the *average* relationship between subsequent maltreatment rates and placement rates across screeners and investigators can be reliably estimated (at least for low placement rates).[15] In practice, the large number of screeners and investigators with low placement rates in our setting help make these extrapolations reliable.

---

[14]See Mueller-Smith (2015); Norris (2019); Mogstad, Torgovitsky, and Walters (2021) and Frandsen, Lefgren, and Leslie (2023) for critiques of conventional first-stage monotonicity in similar settings.

[15]To see how reliable extrapolation is possible when first-stage monotonicity fails, consider a simple model of investigators' placement decisions: $D_{ij} = \mathbb{1}[\kappa_j \geq \nu_{ij}]$ where $\nu_{ij}|\kappa_j, \lambda_j \sim U(0,1)$ and (κ_j, λ_j) are random investigator-specific parameters (here we implicitly condition on $S_i = 1$ and a single race R_i throughout). Further assume $E[Y_i^*|\nu_{ij}, \kappa_j, \lambda_j] = \mu^{S=1} + \lambda_j(\nu_{ij} - \frac{1}{2})$. This model can violate conventional first-stage monotonicity, since investigators can differ both in their ordering of individuals by the appropriateness of placement (ν_{ij}) and their relative skill at predicting subsequent maltreatment (λ_j). Nevertheless, when $E[\lambda_j|\kappa_j]$ is constant (linear) in κ_j, average future maltreatment rates are linear (quadratic) in placement rates, such that simple parametric extrapolations identify mean risk μ.

IV. DATA

IV.A. Data Sources

Most of our analysis is based on data from MDHHS. We observe all child maltreatment investigations (screened-in calls) in Michigan from January 2008 to December 2019. Additionally, we observe data on screened-out calls from January 2017 to December 2019. From these data we construct two analysis samples: a "screener sample" to estimate screener UDs from 2017 to 2019 and an "investigator sample" to estimate investigator UDs from 2008 to 2019. We combine estimates from these samples to derive and decompose placement UDs.

For each hotline call, we observe child and screener identifiers, the child's age and gender, the child's race and ethnicity (discussed more below), the call's exact date and time, and the allegation and reporter types as coded by the screener (e.g., physical abuse versus neglect, and educational personnel versus law enforcement). For screened-in calls, we also observe the relationship of the child to the alleged perpetrator (as coded by the investigator), the child's ZIP code, and indicators for substantiation and foster care placement. Screened-in calls also include the identifier of the investigator who handled the case. For a subset of these cases (from January 2008 to June 2017), we further see the name of the investigator, which we use to predict the investigator's race, ethnicity, and gender.

Two variables in our analysis merit special focus. The first is child race, R_i. Generally, hotline screeners in Michigan do not directly ask callers for the race of the alleged victim.[16] Rather, they observe racial information from a centralized database. If a child has had prior interactions with a CPS investigator—either as the alleged victim or as another involved child (e.g., siblings or family members)—this database contains the race as previously coded by the CPS investigator who visited the home. Otherwise, the database contains the self-reported race from a state-wide database (called MIBridges) with detailed information on families receiving various state benefits.[17] In practice, the vast majority of children reported to the hotline have prior interactions with either CPS or MIBridges such that the database contains

---

[16]Screeners do, however, ask whether the report pertains to a member of a Native American tribe. In such cases, for jurisdiction concerns, the allegation may be redirected to the relevant tribe.

[17]These benefits include programs such as healthcare coverage, the food assistance program, cash assistance (including the Family Independence Program, Refugee Cash Assistance, and TANF), child development and care, and state emergency relief.

a child's race in most cases.[18] We use this database to construct R_i.[19]

The second key variable for our analysis is subsequent maltreatment potential Y_i^*. Our primary maltreatment measure considers whether a child was re-investigated within six months of the focal investigation. Re-investigation is a common measure of subsequent maltreatment in the child welfare literature and in quality measurement among policy makers (Antle et al., 2009; Putnam-Hornstein and Needell, 2011; Casanueva et al., 2015; Putnam-Hornstein et al., 2015; Putnam-Hornstein, Prindle, and Hammond, 2021). While it serves as an imperfect proxy for actual child maltreatment, as it only accounts for cases reported to CPS, re-investigation is a substantive interaction with authorities that entails a report to CPS, a decision by a hotline screener to screen-in the case (as detailed above, only 60% of all hotline calls are assigned for investigation), and an investigation of the family for up to 30 days. Below we show robustness of our main findings to a large number of other proxies, such as re-investigation over other time horizons, substantiation for subsequent maltreatment, and subsequent placement into foster care. A subsequent investigation is also a natural starting point for the analysis because it is not impacted by decisions of the initial investigator: while subsequent investigations within a few months may be re-assigned to the initial investigator who will again make the substantiation and foster care placement decision, neither the decision to report nor to screen-in a case (the two steps required for re-investigation) involve the initial investigator.[20]

Importantly, our analysis is not premised on the view that differences in re-investigation rates are unaffected by discrimination in society more broadly. Differences in re-investigation risk could, for example, be driven by the over-reporting of Black children to CPS. Indeed, prior research suggests that Black children may be disproportionately likely to be reported by medical personnel conditional on case severity (Lane et al., 2002). This scenario could cause us to understate UDs, since they would inflate measured maltreatment risk for Black children. Nevertheless, our goal in conditioning on re-investigation risk is to isolate a particular form of UD from the CPS system which may be reliably targeted by policy, holding fixed other forms

---

[18]Race information is missing in fewer than 10% of all calls. The state believes that the majority of these instances involve Native American children, for whom the screener does not input further race details once their Native American status is established. Nevertheless, to check whether these missing data affect our results, we have confirmed we obtain virtually identical UD estimates when inverse-weighting by an estimate of the probability that the race field is non-missing as a function of observable characteristics of the call (such as the child's age, the reporter type, and the nature of the allegation).

[19]One concern with race data in child welfare may be the intentional miss-coding of race on the part of investigators, particularly in light of recent evidence that such miss-coding is present in the criminal justice system (Luh, 2022). We believe that typical models of miss-coding in our context would attenuate the racial gaps that we document; that is, if investigators tried to hide racial gaps in foster care placement, they may code a Black child as white, which would attenuate our estimates of unwarranted disparity.

[20]Note that a subsequent investigation for child maltreatment in the home within six months is generally not observable for children placed into foster care. The average stay in foster care is 17 months and very few children return home within six months.

of discrimination that may be harder to quantify or address through reforms by CPS.

IV.B. Screener Analysis Sample and Balance Checks

The screener analysis sample is constructed from 558,434 unique hotline calls of white and Black children in Michigan between January 2017 and June 2019. We first drop observations with a missing screener identifier ($N = 13,885$). To minimize noise in our measures of screener and investigator tendencies, we drop calls assigned to screeners with fewer than 100 calls in the sample ($N = 413$) and screened-in investigations assigned to investigators with fewer than 200 cases ($N = 161,494$). To ensure that all calls in the sample are quasi-randomly assigned both to screeners and investigators, we drop calls involving sexual abuse as these are handled by a non-random subset of investigators ($N = 15,909$). We also drop screened-in cases with missing child ZIP code information since quasi-random assignment of investigators is conditional on geography ($N = 30,462$). Finally, we drop repeat reports within one year since repeat screened-in investigations tend to be assigned to the initial investigator ($N = 129,701$).

The resulting analysis sample, summarized in columns 1–3 of Table I, consists of 206,570 hotline calls involving 190,776 children and 162 screeners. 65% of children are white and around half are female; the average child is 8.5 years old. 24% of all calls included at least one physical abuse allegation, while 88% included at least one neglect allegation and 1% included an unspecified maltreatment allegation. 57% of calls come from a mandated reporter (e.g., a teacher, police officer, or doctor), while nearly 20% come from a family member.

The foster care placement rate in the full population of calls is 2.0% (1.7% for white children and 2.6% for Black children). This incorporates a screening-in rate of 63% among Black cases and 58% among white cases. White children are more likely to experience subsequent maltreatment when left at home (either due to being screened-out, or screened-in but not placed in foster care) in the full population of calls: 13.7% of white children experience subsequent maltreatment within six months, compared to 12.6% of Black children.

Table A.I checks an implication of the as-good-as-random assignment of screeners: that observable child and case characteristics are uncorrelated with the screening tendencies of the assigned screener. Separately by race, each column reports point estimates from an OLS regression of an indicator equal to one if the call was screened-in (Columns 1 and 3) and the screener's screening-in tendency (Columns 2 and 4) on child and call characteristics and day-by-shift fixed effects. As expected due to the rotational assignment of screeners, a rich set of characteristics are not jointly predictive of the instrument ($p = 0.122$ for Black children and $p = 0.135$ for white children) despite being very predictive of the decision to screen-in (with F-statistics of 244 and 632, respectively).

IV.C.   Investigator Analysis Sample and Balance Checks

The investigator analysis sample is constructed similarly, making only the necessary sample restrictions for the investigator-level analysis. We start with the dataset of screened-in calls from Gross and Baron (2022), extended to June 2019. We focus on the 374,776 unique investigations in this dataset that did not involve repeat reports within one year and that were assigned to a primary CPS investigator who handled at least 200 investigations of white and Black children. As in the screener analysis sample, we drop cases that involved sexual abuse ($N = 25,907$), cases with missing child ZIP code information ($N = 46,965$), and cases involving children not classified as white or Black ($N = 48,523$). Finally, we drop cases in rotations with only one investigation ($N = 1,830$) and cases assigned to investigators who handled either investigations of only white or only Black children ($N = 6,975$).

The resulting analysis sample, summarized in columns 4–6 of Table I, consists of 244,576 investigations involving 203,438 children and 814 investigators. 69% of children are white and 48% are female; the average child is around seven years old. Nearly 30% of investigations include a physical abuse allegation with roughly 85% including a neglect allegation. The rate of cases including a physical abuse allegation is higher for Black children, while the rate of cases including a neglect allegation is higher for white children. Around 92% of alleged perpetrators include a parent or stepparent.

Overall, 3.4% of investigated (screened-in) children in this sample are placed in foster care (3% of white children and 4.2% of Black children). Around 16% of investigated but non-placed children are re-investigated for child maltreatment within six months. As in the population of all hotline calls, white children are more likely to be re-investigated within 6 months in the set of screened-in calls. The gap in subsequent maltreatment rates is larger in the investigator sample, however: 16.9% of white children compared to 14.5% of Black children.

Table A.II checks an implication of the as-good-as-random assignment of investigators: that observable child and case characteristics are uncorrelated with the foster care placement tendencies of the assigned investigator. As is the case for screeners, a rich set of characteristics are not jointly predictive of the placement instrument ($p = 0.175$ for Black children and $p = 0.166$ for white children) despite being very predictive of the decision to place in foster care (with F-statistics of 169 and 315, respectively). Moreover, in Table A.III, we use the more limited sample period where we observe investigators' demographic information to show that investigators of a given race/gender are not differentially likely to be assigned to same-race/same-gender cases.

IV.D. Descriptive Disparity Analysis

As an initial analysis of screening and placement racial disparities, we estimate descriptive regressions of screening and placement decisions (conditional on the call being screened-in) on an indicator for a child being Black controlling for a variety of child and call characteristics. Table A.IV presents these estimates. Columns 1 and 3 present estimates from simple bivariate regressions without any controls. The 5.3 percentage point screening disparity and the 1.2 percentage point placement disparity correspond to the gaps previously discussed in Table I.

In columns 2 and 4 of Table A.IV, we additionally control for rotation fixed effects—day-by-shift in the screening decision and ZIP code-by-year in the placement decision—and the observable characteristics listed in Table I. The screening disparity shrinks to a 3.7 percentage point gap (6% of the mean screening rate), while the placement disparity shrinks to a 1 percentage point gap (29% of the mean placement rate).

A significant disparity in screening and placement rates thus remains among observably similar children and cases. But the implications of these controlled disparities for UDs are at this point unclear: we cannot adjust these descriptive regressions for subsequent maltreatment potential. Consequently, the overall disparities in Columns 1 and 3 of Table A.IV may suffer from OVB and either over- or under-state the true level of UD across child welfare screeners and investigators in Michigan. The controlled disparities in Columns 2 and 4 may furthermore suffer from included variables bias (IVB) if the controls include mediators of UD. We next present UD estimates that overcome both challenges, following the Section III. approach.

V. Main Results

We present our primary findings in four stages. First, we present UD estimates for screeners (Section V.A.) and investigators (Section V.B.). Second, we present and decompose estimates of placement UD (Section V.C.). Third, we show robustness to a series of relaxations of the exclusion restriction (Section V.D.). Fourth, we document heterogeneity by subsequent maltreatment potential and consider its welfare implications (Section V.E.).

V.A. Unwarranted Disparity in Screener Decisions

As shown in Section III., the identification challenge in estimating UDs in screeners' decisions reduces to the challenge of identifying two key parameters: white and Black mean risk in the full population of calls. We begin by presenting estimates of these two key parameters. We then use these parameters to estimate average screener UD.

Using the screener sample, we extrapolate variation in the race-specific placement and

subsequent maltreatment rates across the quasi-randomly assigned screeners. Panel A of Figure II plots this variation, with a binned scatterplot of strata-adjusted estimates of screeners' placement and subsequent maltreatment rates among calls that did not result in foster care placement. The figure reveals that subsequent maltreatment risk is higher for white cases and that a large number of screeners have a placement rate close to zero, suggesting plausible grounds for extrapolation. The intercepts of each pair of specifications yield estimates of the key mean risk parameters, μ_b and μ_w.[21] Given the low treatment rates in this setting, a linear, quadratic, and a local linear extrapolation all yield very similar estimates of the intercept. We generally use linear extrapolation throughout for simplicity.

Column 1, Panel A of Table II reports estimates of race-specific subsequent maltreatment risk, μ_b and μ_w, from our linear specification. The average maltreatment rate in the six months following an investigation is 0.129 (SE=0.001) in the population of Black children and 0.141 (SE=0.001) in the population of white children. That is, in the full population of hotline calls, we estimate that white children are 1.2 percentage points more likely to experience subsequent maltreatment in the home relative to Black children.

We combine these race-specific mean risk estimates with estimates of screeners' screen-in and conditional subsequent maltreatment rates to estimate UDs in each screener's decisions. We then average across screeners, weighting by their individual caseloads to obtain an estimate of the overall average screener UD (Δ^S). We find a weighted average UD of 0.050 (SE=0.001), suggesting that, on average, calls involving Black children are screened-in at a 5 percentage points higher rate than calls involving white children with identical potential for future maltreatment (Column 1, Panel B). This represents a disparity of roughly 8% of the overall screen-in rate of 60%. The estimated UD is about 35% larger than the controlled observational disparity of 3.7 percentage points (Table A.IV), which underscores the importance of accounting for the selectively observed maltreatment potential.

To examine the sensitivity of our estimates of screener UD to different values of race-specific maltreatment risk, we also construct the non-parametric bounds described in Section III.. We first compute a range of possible mean maltreatment risk parameters, given by the overall average rates of placement and future maltreatment in the full population of calls. The lower bound in this range is obtained by assuming that no children placed in foster care would have experienced subsequent maltreatment, while the upper bound is obtained by assuming that all children placed in foster care would have experienced maltreatment in their homes. Applying this logic to the values in Table I, we estimate Black mean risk bounds of $\mu_b \in [0.125, 0.151]$

---

[21]Throughout this section, we weight each specification inversely by the variance of estimation error in each screener's and investigator's subsequent maltreatment rates. However, as we show below, our estimates are nearly identical if we do not weight the regressions.

and white mean risk bounds of $\mu_w \in [0.137, 0.154]$. We then estimate the range of overall average screener UDs given all combinations of (μ_b, μ_w) in these bounds. Panel A of Figure III shows robust evidence of average screener UD. The range of possible UDs, as measured in the case-weighted average Δ^S, is tightly estimated to be from around 4.95 percentage points to around 5.1 percentage points. The black lines in the figure show that our estimates of average screener UD from linear, quadratic, and local linear extrapolations fall around the middle of the non-parametric bounds.[22]

V.B. Unwarranted Disparity in Investigator Decisions

We follow a similar approach to estimating unwarranted disparity in investigators' decisions: we first estimate white and Black mean risk in the population of screened-in calls using the investigator sample. We then use these parameters to estimate average investigator UD.

In Panel B of Figure II, we extrapolate variation in the race-specific placement and subsequent maltreatment rates across the quasi-randomly assigned investigators. Column 2, Panel A of Table II reports estimates of $\mu_w^{S=1}$ and $\mu_b^{S=1}$. The average future maltreatment rate is 0.155 (SE=0.003) in the population of screened-in Black children and 0.175 (SE=0.003) in the population of screened-in white children. As in the population of all hotline calls, screened-in white children are more likely to experience subsequent maltreatment relative to screened-in Black children. However, the difference in maltreatment risk is larger in the screened-in sample (2 percentage points versus 1.2 percentage points among all calls).

Column 2 of Panel B reports the corresponding estimates of average overall investigator UD (Δ^D). On average, investigators place Black children at a 1.7 percentage point higher rate than white children with identical potential for future maltreatment (SE=0.002). This represents a disparity of roughly 50% relative to the placement rate among screened-in calls of 3.4%, and is nearly 70% larger than the controlled observational disparity in Table A.IV.

We also construct non-parametric bounds of average investigator UD, following the same logic we used to bound average screener UD but now focusing only on screened-in cases. We estimate Black mean risk bounds among screened-in calls of $\mu_b^{S=1} \in [0.139, 0.181]$ and white mean risk bounds of $\mu_w^{S=1} \in [0.164, 0.194]$. We then estimate the range of overall average investigator UDs in Panel B of Figure III, given all combinations of $(\mu_b^{S=1}, \mu_w^{S=1})$ in these bounds. The range of possible UDs, as measured in the case-weighted average Δ^D, is

---

[22]Tighter bounds on average screener UD are obtained by using the placement and future maltreatment rates of screeners with low placement rates. Table A.V shows that restricting to screeners with an estimated (strata-adjusted) placement rate of 0.01 or lower yields mean risk bounds of $\mu_b \in [0.127, 0.137]$ and $\mu_w \in [0.138, 0.148]$. These allow us to construct a tighter bound on average screener UD of $[0.050, 0.051]$.

estimated to be from around 1.2 percentage points to around 2.1 percentage points.[23]

Foster care placement ultimately arises via a combination of screeners' decisions to screen-in the call and investigators' decisions to place children into foster care. We next link the UDs in screener and investigator decisions by a decomposition of overall placement UD (Δ^P). Panel C of Table II shows estimates of overall placement UD from Equations (8) and (9).

Calls involving Black children are 1.1 percentage points more likely to end up in foster care than calls involving white children with identical maltreatment potential (SE=0.001). This represents a disparity of roughly 55% relative to the placement rate among all calls of 2%.

Our two decompositions in Panel C show that screener decisions account for between 13% and 19% of overall placement UD, with investigators driving the remainder. The fact that call screeners drive a significant share of eventual UD in foster care placement is somewhat surprising, since only a small share of screened-in investigations result in placement. This finding highlights the importance of a systems-based analysis of discrimination in high-stakes settings like CPS: our estimates show that eliminating unwarranted disparity in foster care placement rates may require intervention at both phases of CPS involvement. Policies that are able to reduce unwarranted disparity in only the investigation or screening phase leave behind significant discrimination. At the same time, the fact that investigators are the main contributors to overall placement UD underscores the importance of further unpacking UD in investigators' decisions, which we do in Section VI..

V.D.   Relaxing the Exclusion Restriction

The primary threat to the validity of our UD estimates, given quasi-random-assignment, is the possibility of direct effects of screeners and investigators on subsequent at-home maltreatment potential. Here we discuss several tests and extensions of our main estimation procedure to check sensitivity to such exclusion restriction violations. Taken together, these exercises suggest our baseline UD estimates are robust to a wide range of possible violations.

We first examine the exclusion restriction for the screener analysis. Given their narrow mandate and responsibilities, screeners have little scope for affecting future maltreatment potential except through the screening-in decision. However, in principle, this decision

---

[23]Like Table A.V, Table A.VI shows that tighter bounds on average investigator UD can also be obtained by using the placement and future maltreatment rates of investigators with low placement rates among screened-in children. Restricting our focus to investigators with an estimated (strata-adjusted) placement rate of 0.01 or lower yields a bound on average investigator UD of $[0.016, 0.017]$.

could directly affect maltreatment potential in a way that introduces bias in our analysis. Specifically, screening-in a call could see the underlying Y_i^* changed through the effect of investigation on home conditions or behavior.[24] Such screening-in effects could bias our estimates of μ_r, which come from pooling together screened-out calls with calls that are screened-in but do not result in foster care placement.

We develop a procedure in Online Appendix B to test for such direct screening-in effects. This procedure uses the fact that we can estimate the marginal effect of being screened-in on measured at-home maltreatment, $(1 - P_i)Y_i^*$, via a standard IV approach leveraging quasi-random screener assignment. We can then bound the marginal effect of being screened-in on Y_i^* using the fact that it is binary and the fact that we can measure the analogous marginal effect on P_i. Applying this procedure, we estimate tight bounds on the direct screening-in effect of $[-0.006, 0.014]$ (albeit with a standard error of 0.030 on each bound). Hence, we do not find strong evidence to suggest the act of screening-in meaningfully affects at-home maltreatment potential—supporting the exclusion restriction for our screener analysis.

We next turn to a set of specification checks that examine the plausibility of the exclusion restriction for the investigation phase. Here, the scope for unmeasured direct effects is conceivably larger since investigators can interact with the child's parents over an extended period and may refer the family to targeted services as an alternative to foster care. If such services have meaningful effects on at-home maltreatment potential, or if the extended exposure to investigators causes parents to change their behavior in a manner relevant to at-home maltreatment potential, our estimates of the key screened-in mean risk parameters $\mu_r^{S=1}$, which only consider the placement decision, could be biased.[25]

We develop a series of extensions in Online Appendix B that relax the investigator exclusion restriction by allowing for direct effects of targeted services or investigator contact more broadly. We first show that without any exclusion restriction we can still bound a particular measure of unwarranted disparity: one that conditions on a measure of subsequent maltreatment potential that may be impacted by investigators' differential provision of services or unobserved contact effects. These "exclusion-free" bounds come from a version of the Figure III bounds that only uses aggregate (i.e., state-level) rates of placement and subsequent at-home maltreatment, and thus make no assumptions on investigator assignment

---

[24]The impact of CPS investigation without foster care placement on subsequent maltreatment is theoretically ambiguous. On the one hand, a CPS investigation may deter parents from maltreating their children in the future. On the other hand, CPS investigations could be stressful and destabilizing for families.

[25]These bias concerns may be heightened for two reasons. First, placement is mechanically related to the decision to assign preventative services (as this is the alternate path). Second, preventative services may impact the reporting of future maltreatment, affecting our Y_i^* measure. Our specification checks below account for these possible sources of bias.

or excludability. As shown in Table A.VII, we obtain from this exercise bounds on overall average investigator UD of $[0.009, 0.015]$ with standard errors of around 0.001 on each bound. These are similar to (if somewhat smaller than) our baseline estimates of average investigator UD.[26]

The exclusion-free bounds can be seen to capture a policy-relevant measure of unwarranted disparity: the differential rate that investigators place Black children into foster care vs. white children of equal maltreatment risk, conditional on any other interventions that might affect this risk. To gauge the importance of such conditioning, we extend this approach in two ways. First, we adjust the exclusion-free bounds to remove any effects of targeted services on maltreatment potential using the observed service assignment rate (of 6.8%) and *a priori* bounds on the service takeup rate. Intuitively, when assignment and takeup rates are low, such adjustment would have minimal effects on the exclusion-free bounds even for arbitrarily large effects of targeted services on maltreatment potential. Panel A of Table A.VII shows we obtain average investigator UD bounds of $[0.008, 0.017]$ with standard errors of around 0.001 on each bound when assuming a takeup rate of 10%, which is on the high end of our estimates of targeted service takeup in this setting (as discussed in Section II.). For these bounds to include zero, takeup would have to be much higher than this conservative estimate—around 60%, again allowing for arbitrarily large service effects.[27]

Second, we adjust the exclusion-free bounds to remove any unobserved direct effects of investigator contact on subsequent maltreatment potential. Intuitively, such effects would bias our estimates of $\mu_r^{S=1}$ up or down relative to the true race-specific maltreatment risk one would observe in the absence of investigator contact. We thus test how the exclusion-free bounds change as we allow for different ranges of bias in our estimates. Panel B of Table A.VII shows we obtain average investigator UD bounds of $[0.007, 0.018]$ with standard errors of around 0.001 when we allow bias in the range of ± 1 percentage point, which is approximately the estimated upper bound of screening-in effects discussed above. For these bounds to include zero, the bias would have to be much higher—in the range of ± 5 percentage points. Notably, the endpoints of these bounds arise in these simulations when Black and white children experience opposite-signed effects of investigator contact: the lower bound crosses zero when investigator contact decreases Black maltreatment potential by around 5 percentage points

---

[26]As we discuss in Online Appendix B, these bounds do not take into account potentially non-random sorting of investigators across offices and may thus differ from our baseline bounds, which adjust for rotation strata. In practice, as discussed in footnote 13, strata adjustment plays little role in our baseline estimates.

[27]In practice, we expect such effects to be small given previous IV estimates in Gross and Baron (2022). Replicating their analysis in our sample and using the current definition of subsequent maltreatment, we estimate a marginal effect of targeted service assignment on subsequent maltreatment potential of -0.006, with a standard error of 0.027.

while simultaneously *increasing* white maltreatment potential by around 5 percentage points.

V.E. *Heterogeneity by Maltreatment Potential*

Recall that our baseline placement UD measure (Δ^P) is a weighted average of two conditional disparities: the disparity among children with potential for subsequent maltreatment (Δ_1^P), and the disparity among children without potential for subsequent maltreatment (Δ_0^P). The same is true for our measures of UDs in screener (Δ^S) and investigator decisions (Δ^D). While the Δ averages are of significant policy interest, capturing the expected level of UD when the decision-maker encounters a representative pool of children with unknown future maltreatment potential, here we document striking heterogeneity in some of their subcomponents.

To set the stage for these results, consider the interpretation of the two subcomponents of Δ^P. A finding of $\Delta_0^P > 0$ would suggest that Black children without future maltreatment potential are "over-placed" in foster care relative to white children without future maltreatment potential. This disparity would be unambiguously harmful to Black children, since family separation is costly and they would have been safe remaining in their homes. In contrast, a finding of $\Delta_1^P > 0$ may suggest that the relatively higher placement rate of Black children is actually protective, since they experience subsequent maltreatment in their homes and previous research in our context shows that the causal effects of foster care are beneficial for both Black and white children—especially in cases with maltreatment potential in the home (Baron and Gross, 2022; Gross and Baron, 2022). Thus, a finding of $\Delta_1^P > 0$ could be interpreted as "under-placement" of white children relative to equally-risky Black children.

Table III shows significant heterogeneity in the UDs by the unobserved level of maltreatment potential. The average estimate of overall placement UD (1.1 percentage points) is concentrated in the population of children with maltreatment potential, with an estimate of Δ_1^P of 4 percentage points. This means 7.9% of Black children with maltreatment potential are placed in foster care—twice the placement rate experienced by white children at 3.9% (Table A.VIII). In contrast, there is a much smaller disparity among children without maltreatment potential, with an estimate of Δ_0^P of around 0.5 percentage points.

The systems-based analysis reveals a further interesting pattern: ultimate placement UDs are largely driven by disparities in the population of calls with subsequent maltreatment potential, despite significant UDs in screener decisions in both subpopulations (of $Y_i^* = 1$ and $Y_i^* = 0$) of 3.8 and 5.2 percentage points, respectively. We thus find that the heterogeneity in placement UDs is due to the actions of investigators, who make the ultimate placement recommendation. Specifically, we find that investigators amplify initial screener UDs in the subpopulation with $Y_i^* = 1$, despite observing the same or more information and deliberating

over a much longer time frame. We estimate an average investigator UD of nearly 6 percentage points in this subpopulation. At the same time, investigators mitigate initial screener UDs in the subpopulation with $Y_i^* = 0$. We estimate a statistically insignificant investigator UD of 0.8 percentage points in this subpopulation. Table A.VIII shows these results follow from very different investigator placement rates across these subpopulations. Investigators place 13.4% and 7.6% of Black and white screened-in children with maltreatment potential, respectively, but only 2.9% vs 2.1% of Black and white screened-in-children without maltreatment potential, respectively. This is likely because of investigators' skill at inferring underlying risk.

Welfare Implications

These results provide a better understanding of racial discrimination in child protection, which may help guide the appropriate policy response. As mentioned above, a question that our findings may speak to is whether Black children are "over-placed" relative to white children or whether white children are "under-placed." Properly evaluating this question requires an understanding of the benefits and costs of placement for marginal cases.

Two pieces of evidence suggest that in our context foster care placement improves the well-being of screened-in children, reducing subsequent maltreatment and improving longer-term outcomes. First, the quasi-experimental variation in Panel B of Figure II shows that the likelihood of subsequent maltreatment is decreasing in investigator placement rates. Maltreatment is known to carry substantial costs for children's short- and long-term outcomes (Currie and Spatz Widom, 2010; Currie and Tekin, 2012; Soares et al., 2021; Doyle and Aizer, 2018), and failing to place high-risk cases means these costs are more likely to be realized.

Second, among marginal cases where investigator assignment matters for foster care placement decisions, prior research in our setting suggests foster care improves longer-term, welfare-relevant outcomes for screened-in children. Gross and Baron (2022) and Baron and Gross (2022) show that foster care placement causes significant declines in later-in-life criminal justice contact as well as improvements in schooling outcomes such as attendance, test scores, high school graduation, and postsecondary enrollment. These benefits are similar for white and Black children; if anything, the estimates are larger for white children in percent terms. Moreover, comparisons across investigator types in Baron and Gross (2022) suggest that the effects are most positive for children assigned to low-placement-rate investigators. This demonstrates the intuitive notion that marginal cases at particularly high-risk are more likely to benefit from placement. Our current results show that racial disparities in placement are found precisely in these high-risk cases ($Y^* = 1$).[28]

---

[28]While the causal effects of foster care for children on the margin may vary by context (e.g., see Doyle (2007,

Thus, our findings suggest that high-risk cases may benefit from placement both in terms of short-run maltreatment risk and longer-run outcomes, such as improvements in educational outcomes and reductions in criminal justice involvement. A full welfare consideration would include a wider set of benefits and costs, as well as the unobserved willingness to pay by parents to avoid child placement in foster care. Nevertheless, in contrast to the motivation for numerous reforms recently proposed in child welfare policy—which are based on the premise that Black children tend to be over-placed relative to white children—our findings seem consistent with white children at high risk of continued maltreatment being relatively under-placed compared to Black children. Consider, for example, a policy that lowers the placement rate of Black children to equalize white and Black placement rates among screened-in children with identical maltreatment potential.[29] A back-of-the-envelope calculation based on our estimates, combined with observed racial disparities in adult criminal justice contact in Michigan and estimates of the causal effects of foster care on adult crime from Baron and Gross (2022), suggest such a policy would increase the Black-white adult conviction gap by 10% (see Online Appendix C for details).

VI.   Unpacking Investigator UDs

The rest of this paper further unpacks investigator UDs, given their primary role in shaping overall foster care placement UD. We first study heterogeneity in investigator UDs by investigator characteristics (Section VI.A.) and examine possible drivers (Section VI.B.). We then summarize a variety of additional robustness checks (Section VI.C.) and an extension which probes external validity with national data (Section VI.D.).

VI.A.   Heterogeneity by Investigator Characteristics

Table IV reports estimates from OLS regressions of the estimated Δ_{j1}^D on several investigator characteristics. We focus on investigator UDs among children with subsequent maltreatment potential, where investigator UD is concentrated. Results are similar for the overall Δ_j^D's.[30]

Column 1 of Table IV documents significant racial concordance effects in investigator decisions:

2008)), the effects are likely to be most positive precisely in cases with subsequent maltreatment potential.

[29]Such a policy has spurred various recently proposed reforms, exemplified by initiatives such as the Minnesota African American Family Preservation Act, Nassau and Kent County's Blind Removals programs, and policy recommendations outlined in the New York State Bar Association's "Resolution addressing systemic racism in the child welfare system of the State of New York."

[30]As discussed in Section IV., a subset of MDHHS data (from 2008 to 2017) contains investigators' names, which we use to predict demographic characteristics. We focus on the 699 investigators for whom we have such information in this section. Table A.IX shows that our estimates of race-specific maltreatment risk and investigator UD are nearly identical in this sample.

white investigators tend to place white children in high-risk situations at lower rates than Black children in high-risk situations, and vice-versa for Black investigators. Specifically, white investigators are estimated to have an average Δ_{j1}^{D} of 8.4 percentage points (the control mean in this specification) while Black investigators are estimated to have an average Δ_{j1}^{D} of -5.7 percentage points (the sum of the control mean and the Black investigator coefficient estimate). These estimates suggest that investigators of a given race may give the "benefit of the doubt" to families of their same race in high-risk situations. Because the vast majority of investigators in Michigan are white, however, on average Black children have higher placement rates in cases with maltreatment potential.

The table also shows that other investigator traits are associated with smaller disparities. Female investigators, investigators working in urban counties, and investigators with a caseload featuring an above-median share of Black children have lower levels of UD in this subpopulation.

VI.B. Potential Drivers

As noted in Section III., UDs can arise from direct discrimination on the basis of race (either via taste-based discrimination or statistical discrimination), as well as indirect discrimination through non-race characteristics. We first examine the scope for indirect discrimination by testing whether adjusting for child-specific traits (such as age, gender, and family size) as well as investigation-specific traits (such as the nature of the allegations and relationship to the alleged perpetrator) changes the UD estimates. To do this, we residualize the estimated investigator- and race-specific placement and subsequent maltreatment rates by the child and investigation characteristics in Column 4 of Table I. We then recompute mean maltreatment risk and investigator UDs with these adjusted rates.

Table A.X suggests limited scope for indirect discrimination on non-race characteristics. Adjusting for child- and investigation-specific traits leads to very similar estimates of average investigator UD relative to our baseline estimates in Table II (1.6 and 1.7 percentage points, respectively). This finding suggests UD is similar across these non-race characteristics; indeed, we find UD present among every subgroup of children. Table A.XI shows analyses conducted separately for specific subgroups of children. While estimates of UD are larger for female children relative to male children, younger children relative to older children, and investigations that do not involve physical abuse relative to those that do, all estimates are at or above 1 percentage point and statistically significant. In other words, we find meaningful UD across all observed subgroups.

We next study the potential drivers of direct discrimination. Without imposing additional

structure on the quasi-experimental variation, it is difficult to disentangle racial bias and statistical discrimination (Hull, 2021). In Online Appendix D, we estimate a structural model of investigator decision-making akin to that in Arnold, Dobbie, and Hull (2022), which parameterizes the quasi-experimental variation via a series of marginal treatment effect frontiers. Estimates from this model suggest that racial bias, either from racial preferences or inaccurate beliefs, is the primary driver of UD in foster care placement decisions. We find little evidence for statistical discrimination: estimates suggest that investigators act on similarly-precise signals of maltreatment potential by race. Furthermore, estimates of mean maltreatment risk by race suggest investigators should, if accurately statistically discriminating, place white children at higher rates than Black children with identical maltreatment potential, as opposed to the lower placement rates that we observe.

VI.C.   Additional Robustness Checks

Strata-adjustment and weighting in the extrapolations. Throughout the paper, we account for the fact that investigators are only quasi-randomly assigned conditional on geography and time using a linear adjustment that accounts for strata fixed effects. Column 2 of Table A.XIII shows that our results are nearly identical if we do not adjust for strata fixed effects. Moreover, in our baseline extrapolations, we inversely weight by the variance of estimation error in each investigator's re-investigation rate. Column 3 shows that our results are robust to not weighting the specification.

Empirical Bayes shrinkage. In principle, estimation error may attenuate the estimated relationship between placement rates and rates of subsequent at-home maltreatment across investigators. In practice, such attenuation is likely to be minimal since we restrict attention to investigators who see a large number of cases. To verify this, we conduct a standard empirical Bayes "shrinkage" correction to the placement rate estimates (separately by race). Column 4 shows that this exercise yields very similar estimates to our main results, consistent with there being minimal bias from the first-step estimation error.

Omitted payoff bias. Our main outcome throughout the study is whether the child was re-investigated for alleged child maltreatment in the home within six months of the focal investigation.[31] Table A.XIV considers robustness to alternative time frames for re-investigation, finding that the specific period considered has little impact. In Columns 2 through 5 of the table we instead use re-investigation within two, three, four or five months.[32]

---

[31]As noted in Section II., investigator manuals often instruct workers to place a child in foster care if they believe the child is in imminent risk, which we interpret as focusing investigators on the likelihood of maltreatment in the short run rather than the long run.

[32]We omit the first month after the start of an investigation since investigators have 30 days to complete

Reassuringly, we find very similar estimates of investigator UD across all horizons.

We next examine robustness of our proxy for subsequent maltreatment potential. Table A.XV shows that we obtain very similar results when we instead consider a range of other outcomes, including the potential for a subsequent *substantiated* investigation and the potential for subsequent foster care placement. While we prefer subsequent re-investigation as Y_i^* because other measures may be endogenously determined through re-assignment to the initial investigator, it is reassuring that we find similar levels of UD by conditioning on more severe (though potentially endogenous) maltreatment proxies.

A related concern is the potential disconnect between actual maltreatment versus reported maltreatment. For example, subsequent investigations could be partially driven by racial biases in the reporting of child maltreatment. Prior research suggests that Black children may be disproportionately likely to be reported conditional on case severity (Lane et al., 2002). Such bias would tend to understate investigator UD in our context by inflating the risk of a subsequent investigation for screened-in Black children left at home.

Nevertheless, we explore this concern using the screener sample, which contains the category of the initial maltreatment reporter. When we estimate UDs separately for children referred by mandated reporters (social workers, educational, medical, and law enforcement personnel) and non-mandated reporters (e.g., neighbors or other family members), we find similar estimates of UD across these two reporter types (Table A.XVI). These results suggest that our main estimates are not driven by biases from a particular reporter type.

Maltreatment in foster care. Another possible concern with the interpretation of our baseline UD measures, which condition only on a child's potential for subsequent maltreatment in the home, is that children may also face maltreatment in foster care. If maltreatment in foster care is common, and if investigators select children for foster care placement "on gains" of maltreatment potential inside vs. outside of foster care, then a more appropriate UD measure might condition on both potential outcomes. Specifically, questions of possible "over-" and "under-placement" of Black and white children might be better answered by estimating disparities in foster care placement among Black and white children with the same potential reduction in subsequent maltreatment when placed.

Two key empirical facts suggest that such an analysis is unlikely to affect our main conclusions. First, subsequent maltreatment is rare among children placed into foster care: in our data, only around 3% of children in foster care see a subsequent investigation while in foster care, with fewer than 0.8% seeing a substantiated investigation.[33] Second, investigators' removal

---

the focal investigation, and we do not observe a disposition date in the data.

[33]These numbers line up with federal data on substantiated maltreatment in foster care (USDHHS, 2021).

decisions are uncorrelated with the potential for maltreatment in foster care: Figure A.I shows that the slope of the regression relating maltreatment rates in foster care to investigator placement rates is insignificant for both white and Black children. This finding is unsurprising, given the CPS mandate for investigators to focus on maltreatment potential in the home and the fact that investigators have minimal involvement in the decision over where to place a child when removed from home.[34] Taken together, these facts suggest investigators select "on levels" of maltreatment potential in the home (their stated mandate) and not "on gains" from reduced maltreatment potential in care (which they have limited information and control over). In this case, our baseline estimates also capture disparities among Black and white children with the same potential reduction in subsequent maltreatment when placed into foster care.

VI.D.   External Validity

Because CPS is administered at the state or local level and the decision-making process of CPS investigators may vary across states, a natural question is whether our key findings are generalizable outside of Michigan. To examine this question, we use a nationwide dataset from the National Child Abuse and Neglect Data System (NCANDS, 2023) that has fewer variables, but covers most of the US from 2008 to 2019 (matching our main sample period). This dataset contains information on child maltreatment investigations in most states in the U.S. The data include key variables such as the child's race, subsequent CPS investigations, and whether the investigation resulted in foster care placement. While the data do not contain unique investigator identifiers (which prevents us from using extrapolation methods), we apply the aggregate non-parametric bounds discussed in Section V.D. to estimate state-specific investigator UDs. Online Appendix E details the data and approach, and shows that we can replicate our main findings for Michigan using this more limited dataset.

Non-parametric bounds show that our primary findings are generalizable to most U.S. states. Overall, we estimate nationwide average investigator UD bounds of $[0.003, 0.017]$. That is, the average state in the NCANDS data places screened-in Black children in foster care at higher rates than screened-in white children facing the same future maltreatment potential. Figure IV further plots the average estimate in state-specific UD bounds, separately for cases with and without maltreatment potential. As in Michigan, investigator UD in cases without maltreatment potential is near zero in most states while UD in cases with maltreatment potential tends to be as large or larger. We estimate a nationwide average investigator UD

---

[34]Foster children are supervised by a separate set of case managers, not the original investigators. This fact also aligns with earlier findings by Baron and Gross (2022) showing that investigators' placement tendencies are uncorrelated with children's experiences in foster care (such as length of stay, number of different placements, and placement type).

among cases without maltreatment potential of 0.6 percentage points, while the estimated average investigator UD among cases with maltreatment potential is 3.7 percentage points.

VII. Conclusion

This paper develops an empirical framework for studying how discrimination arises and evolves across multi-phase systems. We apply this framework in the context of child protection, a high-stakes and often-debated system, leveraging the quasi-random assignment of both CPS screeners and investigators. We find substantial unwarranted disparity in both phases of the CPS system we study, with investigators inheriting and amplifying initial disparities by screeners. Overall, we find that screeners account for up to 19% of total unwarranted disparity in foster care placement, with investigators accounting for the remainder. Strikingly, in our context, this UD is concentrated in the population of children with subsequent maltreatment potential—a finding that generalizes to most other U.S. states.

Our results have important policy implications for tackling discrimination in foster care placement. The fact that UD arises from both phases of the CPS system suggests that addressing racial disparities may require systemic intervention: policies that target discrimination by focusing only on one phase of CPS may leave untouched significant unwarranted disparity. Furthermore, the finding that UD is concentrated in high-risk cases is relevant to often-discussed policies focused on raising the threshold to place Black children in foster care. Since foster care appears protective to marginally placed children in this setting, such interventions may disproportionately harm Black children by keeping them in risky home environments. Our empirical framework yields a tractable foundation for the important next step of developing appropriate policy responses.

The methods in this paper may also prove useful in other settings where large and pervasive racial disparities have been documented, such as employment, healthcare, housing, and criminal justice. Like child protection, decisions in these settings usually arise across multiple phases, and it is often challenging to disentangle discrimination from selection. Recent work in such settings has developed a variety of quasi-experimental tools to estimate causal parameters. Bringing our framework to these settings, in combination with these tools, may yield new insights on persistent inequities and help design effective policy interventions.

References

Aigner, Dennis J, and Glen G Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *Ilr Review* 30 (2): 175–187.

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

Andrews, Isaiah, James H Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11 727–753.

Antle, Becky F, Anita P Barbee, Dana N Christensen, and Dana J Sullivan. 2009. "The prevention of child maltreatment recidivism through the solution-based casework model of child welfare practice." *Children and Youth Services Review* 31 (12): 1346–1351.

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

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

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

Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, 3–33, Princeton University Press.

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

Baron, E Jason, 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.

Becker, Gary S. 1957. *The Economics of Discrimination*. University of Chicago Press.

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

Bohren, J Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope. 2020. "Inaccurate Statistical Discrimination: An Identification Problem." *Review of Economics and Statistics* Forthcoming.

Bohren, J Aislinn, Peter Hull, and Alex Imas. 2022. "Systemic Discrimination: Theory and Measurement." National Bureau of Economic Research Working Paper 29820.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *Quarterly Journal of Economics* 131 (4): 1753–1794.

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

Brown, Anna, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. "Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Canay, Ivan A, Magne Mogstad, and Jack Mountjoy. 2022. "On the Use of Outcome Tests for Detecting Bias in Decision Making." National Bureau of Economic Research Working Paper 27802.

Casanueva, Cecilia, Stephen Tueller, Melissa Dolan, Mark Testa, Keith Smith, and Orin Day. 2015. "Examining Predictors of Re-reports and Recurrence of Child Maltreatment Using Two National Data Sources." *Children and Youth Services Review* 48 1–13.

Chamberlain, Gary. 1986. "Asymptotic Efficiency in Semi-Parametric Models with Censoring." *Journal of Econometrics* 32 (2): 189–218.

Chouldechova, Alexandra, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. "A Case Study of Algorithm-Assisted Decision Making in Child Maltreatment Hotline Screening Decisions." In *Conference on Fairness, Accountability and Transparency*, 134–148, PMLR.

Currie, Janet, and Cathy Spatz Widom. 2010. "Long-Term Consequences of Child Abuse and Neglect on Adult Economic Well-Being." *Child Maltreatment* 15 (2): 111–120.

Currie, Janet, and Erdal Tekin. 2012. "Understanding the Cycle: Childhood Maltreatment and Future Crime." *Journal of Human Resources* 47 (2): 509–549.

DHHS. 2016. "Protection from Race, Color or National Origin Discrimination in the Child Welfare System."Technical report, Office for Civil Rights, Department of Health and Human Services, https://www.hhs.gov/sites/default/files/child-welfare-title-vi-mepa-factsheet.pdf.

Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram Pathania. 2021. "Measuring Bias in Consumer Lending." *Review of Economic Studies* 88 (6): 2799–2832.

DOJ. 2016. "Title VI Legal Manual: Section VII."Technical report, The United States Department of Justice, https://www.justice.gov/crt/fcs/T6Manual7.

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.

Doyle, Joseph J, and Anna Aizer. 2018. "Economics of Child Protection: Maltreatment, Foster Care, and Intimate Partner Violence." *Annual Review of Economics* 10 87–108.

Drake, Brett, Jennifer M Jolley, Paul Lanier, John Fluke, Richard P Barth, and Melissa Jonson-Reid. 2011. "Racial bias in Child Protection? A Comparison of Competing Explanations Using National Data." *Pediatrics* 127 (3): 471–478.

Feigenberg, Benjamin, and Conrad Miller. 2022. "Would Eliminating Racial Disparities in Motor Vehicle Searches have Efficiency Costs?" *Quarterly Journal of Economics* 137 (1): 49–113.

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

Font, Sarah A, Lawrence M Berger, and Kristen S Slack. 2012. "Examining Racial Disproportionality in Child Protective Services Case Decisions." *Children and Youth Services Review* 34 (11): 2188–2200.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2023. "Judging Judge Fixed Effects." *American Economic Review* 113 (1): 253–277.

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.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111 (5): 1406–1441.

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

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): 170–199.

Harrington, Emma, and Hannah Shaffer. 2022. "Brokers of Bias in the Criminal Justice System: Do Prosecutors Compound or Attenuate Racial Disparities in Policing?", Working paper.

Heckman, James. 1990. "Varieties of Selection Bias." *American Economic Review* 80 (2): 313–318.

Hull, Peter. 2021. "What Marginal Outcome Tests Can Tell Us about Racially Biased Decision-Making." National Bureau of Economic Research Working Paper 28503.

Kelly, John. 2022. "UN Committee Suggests the US Change or Repeal Major Child Welfare Policies." *The Imprint.*

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.

Lane, Wendy G, David M Rubin, Ragin Monteith, and Cindy W Christian. 2002. "Racial Differences in the Evaluation of Pediatric Fractures for Physical Abuse." *JAMA* 288 (13): 1603–1609.

Loury, Glenn C. 1976. "A dynamic theory of racial income differences." Working paper.

Luh, Elizabeth. 2022. "Not so Black and White: Uncovering Racial Bias from Systematically Misreported Trooper Reports." *Working Paper.*

MDHHS. 2020. *Children's Protective Services Policy Manuals.* https://mplp.org/sites/default/files/2020-08/child-protective-services-manual.pdf.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R Walters. 2021. "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables." *American Economic Review* 111 (11): 3663–3698.

Mueller-Smith, Michael. 2015. "The Criminal and Labor Market Impacts of Incarceration." Working paper.

NCANDS. 2023. "Dataset 150, 156, 165, 169, 178, 188, 195, 204, 210, 220, 233, 237, 253, 263."Technical report, National Data Archive on Child Abuse and Neglect (NDACAN), https://www.ndacan.acf.hhs.gov/datasets/datasets-list-ncands-child-file.cfm.

Norris, Samuel. 2019. "Examiner Inconsistency: Evidence from Refugee Appeals." Working paper.

Olea, José Luis Montiel, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3): 358–369.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4): 659–661.

Pincus, Fred L. 1996. "Discrimination Comes in Many Forms: Individual, Institutional, and Structural." *American Behavioral Scientist* 40 (2): 186–194.

Powell, John A. 2008. "Structural Racism: Building upon the Insights of John Calmore." *North Carolina Law Review* 86 791–816.

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.

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, Barbara Needell, Bryn King, and Michelle Johnson-Motoyama. 2013. "Racial and Ethnic Disparities: A Population-Based Examination of Risk Factors for Involvement with Child Protective Services." *Child Abuse & Neglect* 37 (1): 33–46.

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.

Rambachan, Ashesh. 2022. "Identifying Prediction Mistakes in Observational Data." 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, Kelsey V. 2019. "Foster Care and Child Welfare." Working paper.

Small, Mario L., and Devah Pager. 2020. "Sociological Perspectives on Racial Discrimination." *Journal of Economic Perspectives* 34 (2): 49–67.

Soares, Sara, Vania Rocha, Michelle Kelly-Irving, Silvia Stringhini, and Silvia Fraga. 2021. "Adverse Childhood Events and Health Biomarkers: A Systematic Review." *Frontiers in Public Health* 9 649825.

USDHHS. 2021. "Child Welfare Outcomes Report Data." Technical report, Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services, https://cwoutcomes.acf.hhs.gov/cwodatasite/recurrence/index.

White, Shereen, and Stephanie Persson. 2022. "Racial Discrimination in Child Welfare Is a Human Rights Violation–Let's Talk About It That Way." *American Bar Association.*

Wildeman, Christopher, and Natalia Emanuel. 2014. "Cumulative Risks of Foster Care Placement by Age 18 for US Children, 2000–2011." *PloS one* 9 (3): e92785.

Figure I: Child Protection in Michigan
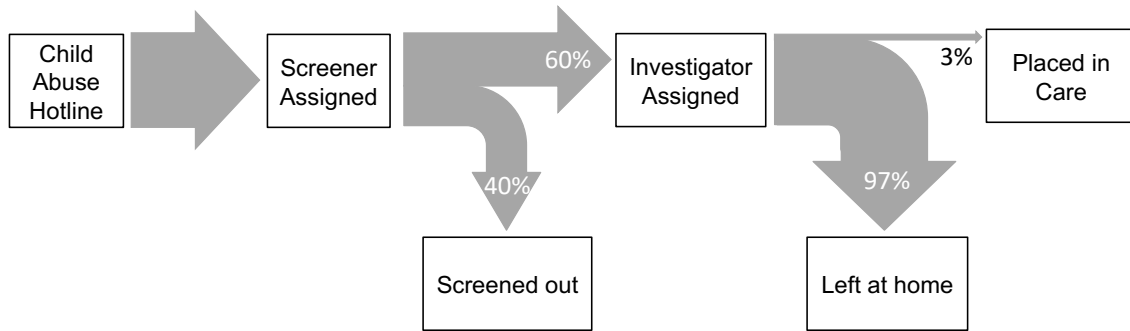
Notes: The figure describes the child protection process in Michigan. Both screeners and investigators are quasi-randomly assigned, as described in the text. The percentages on screening-in and out refer to all calls received; percentages thereafter refer to investigated cases.

Figure II: Screener and Investigator Placement and Conditional Maltreatment Rates
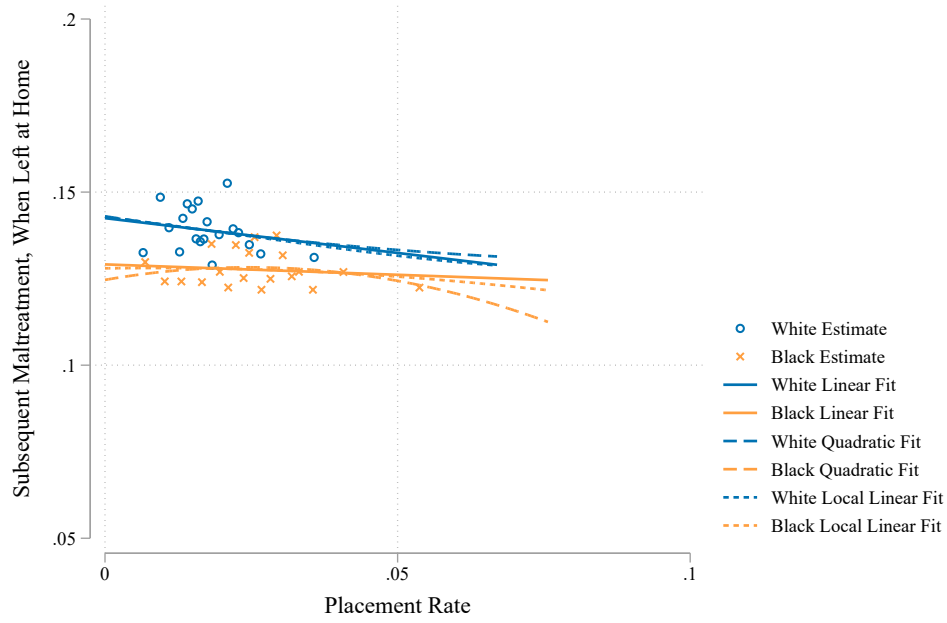
Panel A: All Hotline Calls
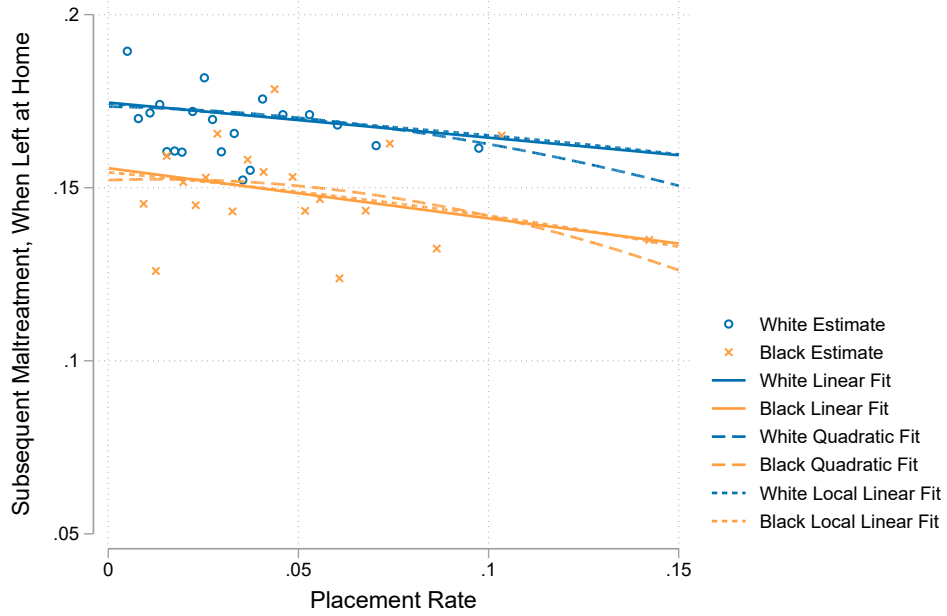
Panel B: Screened-in Calls

Notes. Panel A shows a binscatter (with 20 equal-sized bins) of placement rates, by race, for the 162 screeners in the screener analysis sample against rates of a subsequent maltreatment investigation within six months for children left at home. All estimates adjust for day by shift fixed effects. The figure also plots race-specific linear, quadratic, and local linear curves of best fit, obtained from screener-level regressions that inverse-weight by the variance of the estimated re-investigation rate among children left at home. Panel B repeats this exercise, but for the 814 investigators in the investigator analysis sample.

Figure III: Bounds for Average Screener and Investigator UD
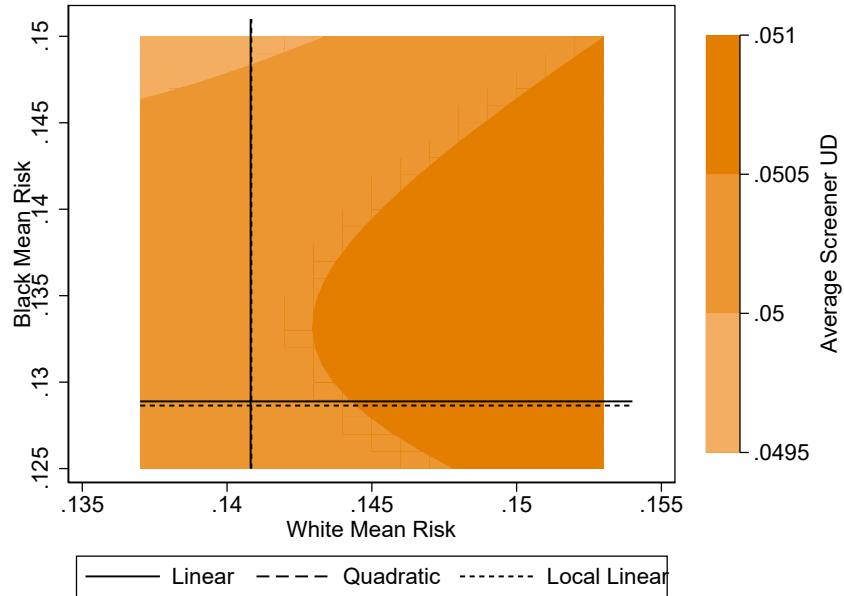
Panel A: Screeners
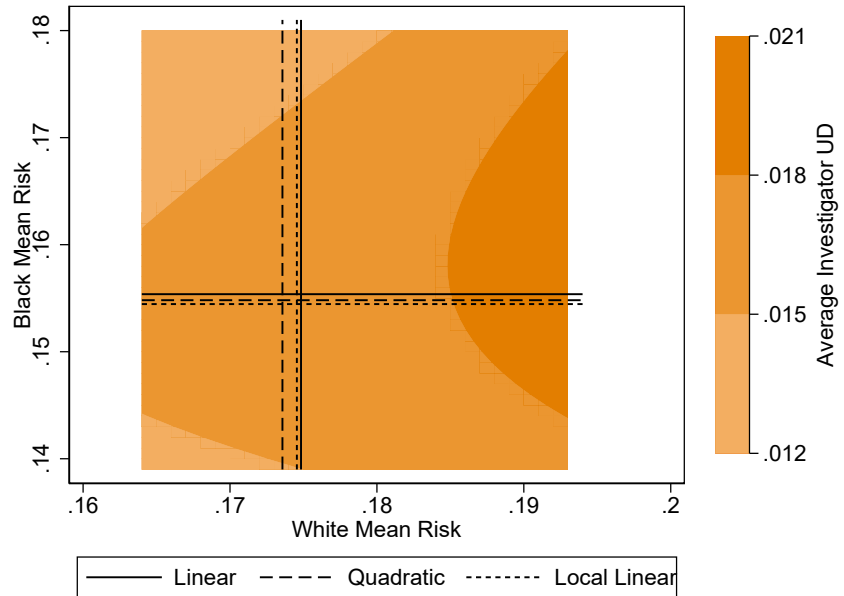
Panel B: Investigators

Notes. Panel A shows how our estimates of average screener UD change under different estimates of Black and white mean risk in the full population of calls. The mean risk estimates obtained from the linear, quadratic, and local linear extrapolations in Panel A of Figure II are indicated by solid, dashed, and dotted lines, respectively. The ranges of Black and white mean risk reflect the bounds implied by the average placement and subsequent maltreatment rates in the screener sample: $\mu_b \in [0.125, 0.151]$ and $\mu_w \in [0.137, 0.154]$. Panel B repeats this exercise, but for average investigator UD. The ranges of Black and white mean risk in the population of screened-in calls implied by the average placement and subsequent maltreatment rates in the investigator sample are: $\mu_b^{S=1} \in [0.139, 0.181]$ and $\mu_w^{S=1} \in [0.164, 0.194]$.

Figure IV: Nationwide Estimates of Average Investigator UD, by Maltreatment Potential

> .05
.01 − .05
0 − .01
−.01 − 0
< −.01
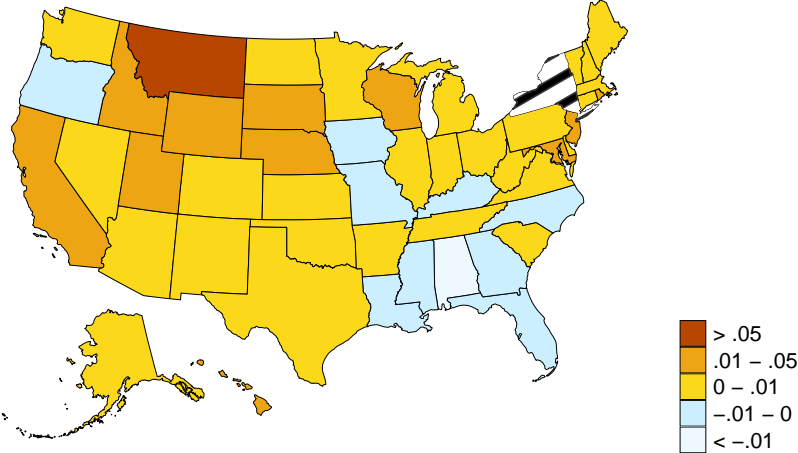
Panel A: No maltreatment potential ($Y^* = 0$)

> .05
.01 − .05
0 − .01
−.01 − 0
< −.01
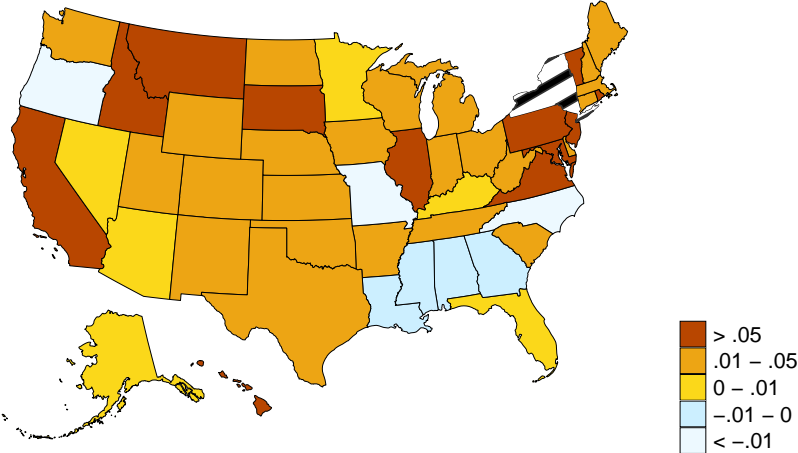
Panel B: Maltreatment potential ($Y^* = 1$)

Notes. This figure uses national data from NCANDS to show the mean of the state-specific investigator UD bounds between 2008 and 2019. Panel A presents estimates for cases without maltreatment potential while Panel B presents estimates for cases with maltreatment potential. See Online Appendix E for additional details regarding the dataset and methodology.

Table I: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	Screener Sample			**Investigator Sample**		
	All Children	**Black Children**	**White Children**	**All Children**	**Black Children**	**White Children**
Panel A: Child characteristics						
White	0.647	0.000	1.000	0.692	0.000	1.000
Female	0.491	0.488	0.492	0.482	0.483	0.482
Age during investigation	8.456	7.941	8.737	6.918	6.595	7.062
Panel B: Call characteristics						
Call included:						
Physical abuse allegation	0.241	0.250	0.236	0.285	0.301	0.278
Neglect allegation	0.878	0.871	0.882	0.851	0.842	0.855
Maltreatment allegation	0.010	0.008	0.011	0.177	0.176	0.178
Reporter category:						
Mandated reporter	0.574	0.561	0.581			
Family member	0.199	0.176	0.211			
Other	0.228	0.264	0.208			
Alleged perpetrator includes:						
Parent/step-parent				0.920	0.917	0.921
Non-parent relative				0.051	0.059	0.047
Someone unrelated				0.109	0.103	0.112
Number of children in the home				3.071	3.386	2.931
Panel C: Treatment						
Screen-in rate	0.599	0.634	0.581	1.000	1.000	1.000
Foster care placement rate	0.020	0.026	0.017	0.034	0.042	0.030
Panel D: Outcomes, if not placed						
Re-investigated within 6 months	0.135	0.128	0.139	0.162	0.145	0.169
Number of calls	206,570	72,877	133,693	244,576	75,346	169,230
Number of children	190,776	67,348	123,428	203,438	63,990	139,448

Notes. This table summarizes the screener and investigator analysis samples. The screener sample includes 206,570 unique hotline calls in Michigan from 2017 to 2019, assigned to 162 distinct hotline screeners. The investigator sample includes 244,576 unique investigations (screened-in calls) from 2008 to 2019, assigned to 814 unique investigators. Reporter information is available only from 2017 onward, which is why it is only presented for the screener sample. Similarly, the alleged perpetrator is coded by the investigator, making it available only for screened-in calls. A mandated reporter is defined as someone who is either educational personnel, medical personnel, a social worker, or a member of law enforcement. Neglect allegations include physical neglect, medical neglect, failure to protect, and improper supervision. Rows describing the nature of the allegation and the alleged perpetrator can sum to more than one because the categories are not mutually exclusive.

Table II: Estimates of Mean Maltreatment Risk and Unwarranted Disparity

	(1)	(2)
	Panel A: Subsequent maltreatment risk	
	All calls	**Screened-in calls**
Black children	0.129	0.155
	(0.001)	(0.003)
White children	0.141	0.175
	(0.001)	(0.003)
	Panel B: Unwarranted disparity (UD)	
	Screeners	**Investigators**
Average across decision-makers	0.050	0.017
	(0.001)	(0.002)
	Panel C: Placement UD and decompositions	
	Equation (8)	**Equation (9)**
Placement UD	0.011	0.011
	(0.001)	(0.001)
Screener share (%)	12.5	18.6
	(2.5)	(2.0)
Investigator share (%)	87.5	81.4
	(2.5)	(2.0)
Number of screeners	162	162
Number of investigators	814	814

Notes. The table reports estimates of mean maltreatment risk and unwarranted disparity in both phases of the CPS system. Panel A reports estimates of race-specific average subsequent maltreatment risk, both in the full population of calls and in the population of screened-in calls. These estimates come from the linear extrapolations in Figure II. Panel B reports estimates of average overall screener and investigator UD (weighted by the number of cases assigned to each screener and investigator, respectively). Panel C reports overall placement UD in the full population of cases, as well as the share of placement UD that is due to screeners' decisions according to Equations (8) and (9). Robust standard errors for estimates of maltreatment risk in the population of all calls, as well as average screener UD, are two-way clustered at the child and screener level. Robust standard errors for estimates of maltreatment risk in the population of screened-in calls, as well as average investigator UD, are two-way clustered at the child and investigator level. Robust standard errors for estimates of overall placement UD, as well as the percent of placement UD due to screeners and investigators, reflect uncertainty in the underlying estimates of both average screener and investigator UDs. All standard errors are obtained via a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table III: Estimates of UD by Subsequent Maltreatment Potential

	(1) Weighted average	(2) With maltreatment potential	(3) Without maltreatment potential
Screener UD	0.050	0.038	0.052
	(0.001)	(0.004)	(0.001)
Investigator UD	0.017	0.058	0.008
	(0.002)	(0.023)	(0.005)
Placement UD	0.011	0.040	0.005
	(0.001)	(0.013)	(0.003)
Number of screeners	162	162	162
Number of investigators	814	814	814

Notes. This table presents estimates of average screener, investigator, and placement UD separately by maltreatment potential. Specifically, the first row presents estimates of Δ^S, Δ_1^S, and Δ_0^S, respectively. We then present estimates of Δ^D, Δ_1^D, and Δ_0^D. Finally, we combine estimates of screener and investigator UDs to estimate overall placement UD (Δ^P, Δ_1^P, and Δ_0^P) according to Equations (8) and (9). Robust standard errors for estimates of screener UDs are two-way clustered at the child and screener level. Robust standard errors for estimates of investigator UDs are two-way clustered at the child and investigator level. Robust standard errors for estimates of placement UD reflect uncertainty in the underlying estimates of both screener and investigator UDs. All standard errors are obtained via a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table IV: Regressions of Investigator UDs On Investigator Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Black investigator	-0.141						-0.084
	(0.063)						(0.064)
Other race	-0.224						-0.204
	(0.135)						(0.138)
Female investigator		-0.118					-0.106
		(0.047)					(0.049)
New investigator			-0.037				-0.031
			(0.046)				(0.051)
Urban county				-0.152			-0.064
				(0.043)			(0.043)
High share of Black children					-0.158		-0.103
					(0.043)		(0.042)
Low placement rate						-0.009	-0.019
						(0.046)	(0.052)
Control mean	0.084	0.145	0.085	0.135	0.139	0.067	0.063
Number of investigators	699	699	699	699	699	699	699

Notes. This table reports OLS estimates of regressions of estimates of Δ_{j1}^D (based on a linear extrapolation) on investigator characteristics. As discussed in Section IV., a subset of MDHHS data (from 2008 to 2017) contains the names of investigators, which we use to predict demographic characteristics. This table focuses on the 699 investigators for whom we have such information. Except for Column 7, the "control mean" refers to the average value of dependent variable, weighted by the number of investigations assigned to each investigator, for the omitted category. The omitted category in Column 1 is an indicator variable for whether the investigator is white. 86% of investigators in our sample are white, 12% are Black, and 2% are another race, such as Hispanic or Asian. Column 7 reports the (weighted) average value of the dependent variable across all investigators in this sample. New investigators are defined as those hired during our estimation period. There are 457 new investigators in our sample. Investigators with a high share of Black children are those whose caseloads have an above-median share of Black children. Low placement rate investigators are those with a placement rate below the sample median, controlling for rotation fixed effects. All specifications are weighted by the number of investigations assigned to the specific investigator. Robust standard errors, twoway clustered by child and investigator are reported in parentheses, and are obtained via a bootstrapping procedure (with 500 replications).

Discrimination in Multi-Phase Systems:
Evidence from Child Protection

E. Jason Baron, Joseph J. Doyle, Jr., Natalia Emanuel,
Peter Hull, and Joseph Ryan

Online Appendix

Table of Contents

A Supplemental Figures and Tables

Figure A.I: Investigator Placement and Subsequent Maltreatment Rates in Foster Care
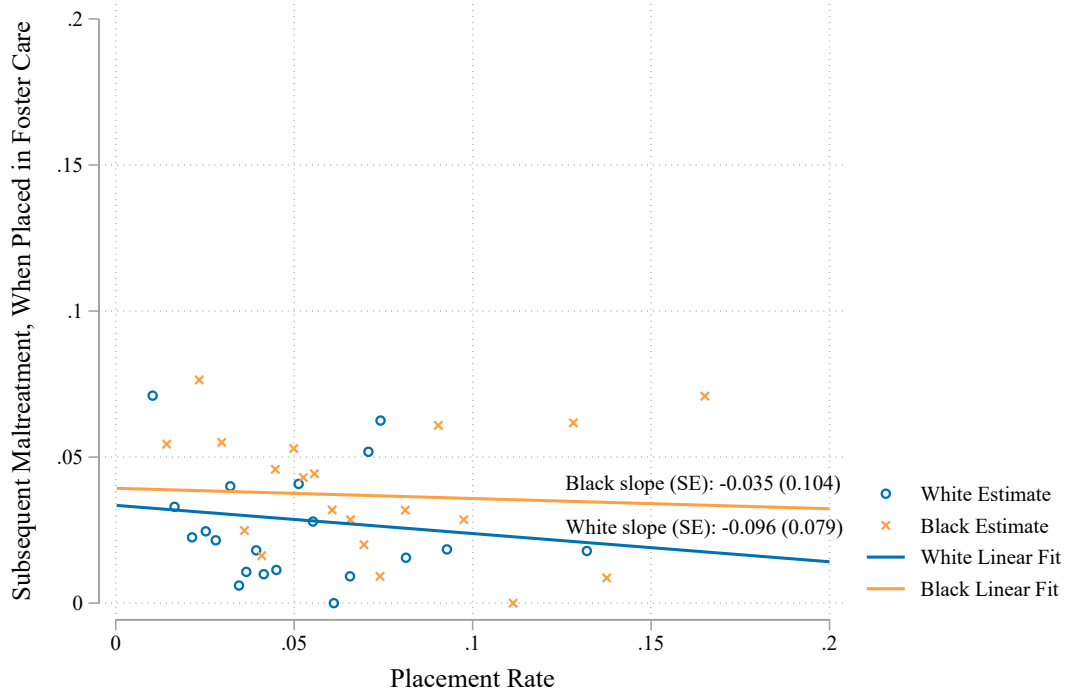
Notes. The figure shows a binscatter (with 20 equal-sized bins) of placement rates, by race, for the 814 investigators in the investigator analysis sample against rates of a subsequent maltreatment investigation within six months for children *placed in foster care*. All estimates adjust for ZIP code by year fixed effects. The figure also plots the race-specific linear curves of best fit, obtained from investigator-level regressions that inverse-weights by the variance of the estimated re-investigation rate among children placed in foster care. Robust standard errors for the slope of this line are reported in parentheses.

Figure A.II: Non-Parametric Bounds of Average Investigator UD (MDHHS vs. NCANDS)
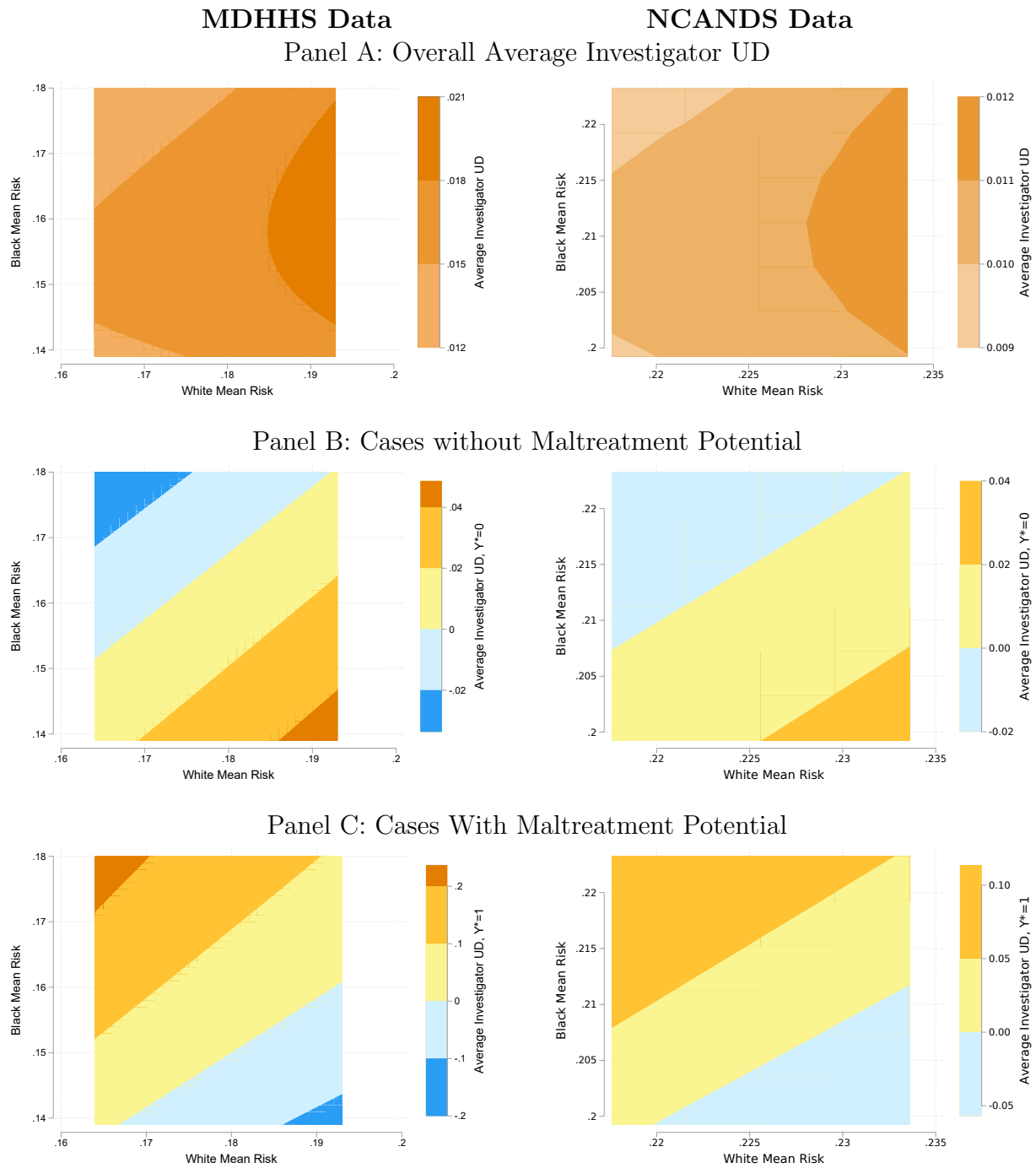
Notes. This figure shows non-parametric bounds of average investigator UD using our primary administrative dataset from MDHHS (left-hand side) and more limited NCANDS data (right-hand side). The figures show how estimates of average investigator UD change under different estimates of Black ($\mu_b^{S=1}$) and white ($\mu_w^{S=1}$) mean risk. The ranges of Black and white mean risk reflect the bounds implied by the average placement and subsequent maltreatment rates among screened-in cases in the respective sample.

Table A.I: Screener Balance Tests

Dependent variable:	(1) Black children Screened-in	(2) Screening Tendency	(3) White children Screened-in	(4) Screening Tendency
Female	-0.005	0.000	-0.011	-0.000
	(0.003)	(0.000)	(0.003)	(0.000)
Age during investigation = 1	-0.039	0.000	-0.049	-0.000
	(0.010)	(0.001)	(0.009)	(0.001)
Age during investigation = 2	-0.063	0.000	-0.050	0.000
	(0.010)	(0.001)	(0.009)	(0.001)
Age during investigation = 3	-0.060	-0.000	-0.077	-0.000
	(0.011)	(0.001)	(0.009)	(0.001)
Age during investigation = 4	-0.079	-0.002	-0.084	0.000
	(0.011)	(0.001)	(0.009)	(0.001)
Age during investigation = 5	-0.103	0.000	-0.112	-0.001
	(0.010)	(0.001)	(0.008)	(0.001)
Age during investigation = 6	-0.124	0.001	-0.129	0.001
	(0.011)	(0.001)	(0.009)	(0.001)
Age during investigation = 7	-0.130	-0.002	-0.143	0.000
	(0.012)	(0.001)	(0.009)	(0.001)
Age during investigation = 8	-0.130	-0.001	-0.138	-0.001
	(0.011)	(0.001)	(0.009)	(0.001)
Age during investigation = 9	-0.130	-0.001	-0.141	-0.001
	(0.011)	(0.001)	(0.009)	(0.001)
Age during investigation = 10	-0.145	-0.001	-0.155	-0.000
	(0.010)	(0.001)	(0.009)	(0.001)
Age during investigation = 11	-0.165	-0.000	-0.158	0.001
	(0.011)	(0.001)	(0.008)	(0.001)
Age during investigation = 12	-0.157	-0.000	-0.176	0.000
	(0.012)	(0.001)	(0.009)	(0.001)
Age during investigation = 13	-0.164	-0.001	-0.175	-0.000
	(0.011)	(0.001)	(0.009)	(0.001)
Age during investigation = 14	-0.183	-0.000	-0.193	-0.000
	(0.013)	(0.001)	(0.009)	(0.001)
Age during investigation = 15	-0.177	0.001	-0.211	-0.000
	(0.012)	(0.001)	(0.009)	(0.001)
Age during investigation = 16	-0.189	-0.002	-0.190	0.000
	(0.011)	(0.001)	(0.009)	(0.001)
Age during investigation = 17	-0.180	-0.001	-0.178	-0.000
	(0.013)	(0.001)	(0.010)	(0.001)
Age during investigation = 18	-0.154	0.003	-0.188	-0.001
	(0.017)	(0.001)	(0.012)	(0.001)
Physical abuse allegation	0.101	0.002	0.094	0.001
	(0.008)	(0.001)	(0.005)	(0.001)
Neglect allegation	0.006	0.001	0.006	0.001
	(0.009)	(0.001)	(0.007)	(0.001)
Maltreatment allegation	0.123	0.003	0.151	0.003
	(0.027)	(0.003)	(0.016)	(0.001)
Education personnel	0.214	0.002	0.209	0.004
	(0.061)	(0.007)	(0.047)	(0.004)
Family member	0.192	0.002	0.187	0.005
	(0.061)	(0.007)	(0.047)	(0.004)
Medical personnel	0.172	0.002	0.208	0.005
	(0.061)	(0.007)	(0.047)	(0.004)
Counselor	0.112	0.005	0.127	0.004
	(0.063)	(0.008)	(0.046)	(0.004)
Law enforcement	0.317	0.002	0.321	0.007
	(0.060)	(0.007)	(0.047)	(0.004)
Court	0.184	0.002	0.204	0.003
	(0.066)	(0.007)	(0.052)	(0.005)
MDHHS	0.211	0.000	0.223	0.004
	(0.061)	(0.007)	(0.048)	(0.005)
Provider	0.066	0.003	0.118	0.006
	(0.063)	(0.007)	(0.051)	(0.005)
Other	0.200	0.001	0.213	0.005
	(0.060)	(0.007)	(0.048)	(0.004)
Exact time of call	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
First-stage F-statistic	244		632	
F-statistic from joint test	128.44	1.33	185.61	1.32
P-value from joint test	0.000	0.122	0.000	0.135
Number of calls	72,877	72,877	133,693	133,693

Notes. This table reports OLS estimates of regressions of the dependent variable on child/call characteristics. The regressions are estimated on the screener analysis sample. Each screener's screening-in tendency is a leave-one-out rate among all calls in the sample that were quasi-randomly assigned to the screener. Specifically, we first regress the screened-in indicator on day by shift fixed effects. Separately by race, we then calculate the leave-one-out average residual from this regression for each screener. The p-values reported at the bottom of each column are from F tests of the joint significance of the variables listed in the rows. We also report robust Kleibergen-Paap F-statistics, which in the just-identified case are equivalent to the effective F-statistics of Olea and Pflueger (2013) (Andrews, Stock, and Sun, 2019). Robust standard errors, two-way clustered at the child and screener levels, are reported in parentheses.

Dependent variable:	(1) Black children Placed	(2) Placement Tendency	(3) White children Placed	(4) Placement Tendency
Female	-0.000	0.000	-0.001	0.000
	(0.001)	(0.000)	(0.001)	(0.000)
Age during investigation = 1	-0.067	0.000	-0.056	-0.000
	(0.005)	(0.000)	(0.003)	(0.000)
Age during investigation = 2	-0.067	-0.000	-0.064	-0.000
	(0.005)	(0.000)	(0.003)	(0.000)
Age during investigation = 3	-0.071	-0.000	-0.066	-0.001
	(0.004)	(0.001)	(0.003)	(0.000)
Age during investigation = 4	-0.071	-0.000	-0.065	-0.001
	(0.005)	(0.000)	(0.003)	(0.000)
Age during investigation = 5	-0.069	0.000	-0.067	-0.000
	(0.004)	(0.000)	(0.003)	(0.000)
Age during investigation = 6	-0.068	-0.000	-0.069	-0.001
	(0.005)	(0.001)	(0.003)	(0.000)
Age during investigation = 7	-0.073	0.000	-0.070	-0.001
	(0.005)	(0.000)	(0.003)	(0.000)
Age during investigation = 8	-0.075	-0.000	-0.073	-0.001
	(0.005)	(0.001)	(0.003)	(0.000)
Age during investigation = 9	-0.076	-0.000	-0.073	-0.001
	(0.004)	(0.000)	(0.003)	(0.000)
Age during investigation = 10	-0.074	-0.000	-0.070	-0.001
	(0.005)	(0.001)	(0.003)	(0.000)
Age during investigation = 11	-0.073	-0.000	-0.072	-0.001
	(0.005)	(0.001)	(0.003)	(0.000)
Age during investigation = 12	-0.064	-0.000	-0.070	-0.001
	(0.005)	(0.001)	(0.003)	(0.000)
Age during investigation = 13	-0.061	0.000	-0.068	-0.001
	(0.005)	(0.001)	(0.003)	(0.000)
Age during investigation = 14	-0.056	-0.000	-0.066	-0.001
	(0.005)	(0.001)	(0.003)	(0.000)
Age during investigation = 15	-0.061	0.001	-0.066	-0.000
	(0.006)	(0.001)	(0.004)	(0.000)
Age during investigation = 16	-0.054	-0.001	-0.068	-0.001
	(0.006)	(0.001)	(0.004)	(0.000)
Age during investigation = 17	-0.056	-0.001	-0.069	-0.001
	(0.011)	(0.001)	(0.007)	(0.001)
Age during investigation = 18	-0.077	-0.003	-0.083	0.002
	(0.013)	(0.002)	(0.009)	(0.001)
Physical abuse allegation	0.015	0.001	0.007	0.000
	(0.003)	(0.000)	(0.002)	(0.000)
Neglect allegation	0.029	0.001	0.022	0.000
	(0.004)	(0.000)	(0.002)	(0.000)
Maltreatment allegation	0.010	-0.000	0.012	0.000
	(0.003)	(0.001)	(0.002)	(0.000)
Alleged perpetrator includes the parent/step-parent	0.048	0.002	0.033	0.001
	(0.007)	(0.001)	(0.003)	(0.000)
Alleged perpetrator includes a non-parent relative	0.015	0.001	0.011	0.000
	(0.007)	(0.001)	(0.003)	(0.000)
Alleged perpetrator includes someone unrelated	0.038	0.001	0.025	0.000
	(0.006)	(0.001)	(0.003)	(0.000)
Number of children in allegation	-0.004	0.000	-0.002	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)
First-stage F-statistic	169		315	
F-statistic from joint test	27.53	1.25	42.60	1.26
P-value from joint test	0.000	0.175	0.000	0.166
Number of investigations	75,346	75,346	169,230	169,230

Notes. This table reports OLS estimates of regressions of the dependent variable on child/investigator characteristics. The regressions are estimated on the screener analysis sample. Each investigator's placement tendency is a leave-one-out rate among all investigations in the sample that were quasi-randomly assigned to the investigator. Specifically, we first regress the placement indicator on ZIP code by year fixed effects. Separately by race, we then calculate the leave-one-out average residual from this regression for each investigator. The p-values reported at the bottom of each column are from F tests of the joint significance of the variables listed in the rows. We also report robust Kleibergen-Paap F-statistics, which in the just-identified case are equivalent to the effective F-statistics of Olea and Pflueger (2013) (Andrews, Stock, and Sun, 2019). Robust standard errors, two-way clustered at the child and investigator levels, are reported in parentheses.

Table A.III: Additional Balance Tests for Investigator Assignment

Dependent variable:	(1) White Investigator	(2) Black Investigator	(3) Female Investigator
White child	0.0016		
	(0.0035)		
Black child		0.0001	
		(0.0032)	
Female child			0.0012
			(0.0017)
Constant	0.8640	0.1097	0.6925
	(0.0125)	(0.0109)	(0.0157)
Number of investigations	217,704	217,704	217,704

Notes. This table reports OLS estimates of regressions of the dependent variable on child characteristics. The regressions are estimated on the sample of investigations from 2008–2017, for whom we have information on their demographics. All regressions control for ZIP code by investigation year fixed effects. Robust standard errors, two-way clustered at the child and investigator levels, are reported in parentheses.

Table A.IV: Descriptive Regressions of Screening and Foster Care Placement

Dependent variable:	(1) Screened-in	(2) Screened-in	(3) Placed	(4) Placed
Black child	0.053	0.037	0.012	0.010
	(0.003)	(0.003)	(0.001)	(0.001)
Baseline controls		✓		✓
Rotation FE		✓		✓
Dependent variable mean	0.599	0.599	0.034	0.034
Number of observations	206,570	206,570	244,576	244,576

Notes. Columns 1 and 2 report OLS estimates of regressions of an indicator equal to one if the call was screened-in on an indicator equal to one if the call involved a Black child. These regressions are estimated on the screener analysis sample. Column 1 reports the results of a bivariate regression, while Column 2 additionally controls for the child and call characteristics in Column 1 of Table I, as well as day by shift fixed effects. Columns 3 and 4 report OLS estimates of regressions of an indicator equal to one if the child was placed in foster care on an indicator equal to one if the investigation involved a Black child. These regressions are estimated on the investigator analysis sample. Column 3 reports the results of a bivariate regression, while Column 4 additionally controls for the child and investigation characteristics in Column 4 of Table I, as well as zip code by year fixed effects. Robust standard errors are two-way clustered at the child and screener levels in the first two columns, and two-way clustered at the child and investigator levels in the last two columns. All standard errors are presented in parentheses.

Table A.V: Bounds on Maltreatment Risk and Screener UDs

	(1) From 0.03 Placement Rate	(2) From 0.02 Placement Rate	(3) From 0.01 Placement Rate
	Panel A: Subsequent maltreatment risk		
Black children	[0.124,0.154]	[0.125,0.145]	[0.127,0.137]
	(0.001,0.001)	(0.001,0.001)	(0.001,0.001)
White children	[0.133,0.163]	[0.137,0.157]	[0.138,0.148]
	(0.001,0.001)	(0.001,0.001)	(0.001,0.001)
	Panel B: Screener UD		
Mean across screeners	[0.050,0.052]	[0.050,0.051]	[0.050,0.051]
	(0.001,0.001)	(0.001,0.001)	(0.001,0.001)
Number of screeners	162	162	162

Notes. Panel A reports bounds on race-specific average re-investigation risk in the full population of calls, while Panel B reports bounds on average (case-weighted) screener UD. To estimate bounds on mean risk, Column 1 uses a local linear fit of re-investigation rates among screeners with a 3% placement rate of Black and white children. Columns 2 and 3 form bounds from screeners with placement rates of 2% and 1%, respectively. Bounds are formed under the assumption that either none or all of the children placed in foster care in each column have potential for re-investigation. Panel B searches within these bounds to find the combination of Black and white mean risk that minimize or maximize each screener UD statistic. Robust standard errors on the endpoints of each set of bounds, two-way clustered at the child and screener levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.VI: Bounds on Screened-in Maltreatment Risk and Investigator UDs

	(1) From 0.03 Placement Rate	(2) From 0.02 Placement Rate	(3) From 0.01 Placement Rate
Panel A: Subsequent maltreatment risk			
Black children	[0.147,0.177]	[0.149,0.169]	[0.151,0.161]
	(0.002,0.002)	(0.002,0.002)	(0.003,0.003)
White children	[0.166,0.196]	[0.169,0.189]	[0.171,0.181]
	(0.002,0.002)	(0.002,0.002)	(0.002,0.002)
Panel B: Investigator UD			
Mean across investigators	[0.013,0.020]	[0.015,0.019]	[0.016,0.017]
	(0.002,0.002)	(0.002,0.002)	(0.002,0.002)
Number of investigators	814	814	814

Notes. Panel A reports bounds on race-specific average re-investigation risk in the population of screened-in calls, while Panel B reports bounds on average (case-weighted) investigator UD. To estimate bounds on mean risk, Column 1 uses a local linear fit of re-investigation rates among investigators who remove 3% of screened-in Black and white children. Columns 2 and 3 form bounds from investigators who remove 2% and 1% of screened-in Black and white children, respectively. Bounds are formed under the assumption that either none or all of the children placed in foster care in each column have potential for re-investigation. Panel B searches within these bounds to find the combination of Black and white mean risk that minimize or maximize each investigator UD statistic. Robust standard errors on the endpoints of each set of bounds, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.VII: Exclusion-Free Bounds of Investigator UD, With and Without Direct Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Without effects from targeted services					
	Services take-up rate					
Baseline	10%	20%	30%	40%	50%	60%
$[0.009, 0.015]$	$[0.008, 0.017]$	$[0.006, 0.018]$	$[0.005, 0.022]$	$[0.003, 0.024]$	$[0.001, 0.029]$	$[-0.001, 0.032]$
$(0.001, 0.001)$	$(0.001, 0.001)$	$(0.001, 0.001)$	$(0.001, 0.001)$	$(0.001, 0.001)$	$(0.001, 0.002)$	$(0.001, 0.001)$
	Panel B: Without effects from investigator contact					
	Range of contact effects					
Baseline	$\pm 1pp$	$\pm 2pp$	$\pm 3pp$	$\pm 4pp$	$\pm 5pp$	$\pm 6pp$
$[0.009, 0.015]$	$[0.007, 0.018]$	$[0.005, 0.021]$	$[0.003, 0.025]$	$[0.001, 0.031]$	$[-0.003, 0.036]$	$[-0.006, 0.043]$
$(0.001, 0.001)$	$(0.001, 0.001)$	$(0.001, 0.001)$	$(0.001, 0.001)$	$(0.001, 0.002)$	$(0.001, 0.002)$	$(0.001, 0.002)$

Notes. Panel A reports the exclusion-free bounds described in Section V.D. and formalized in Online Appendix B, as well as the extension which removes any possible direct effects of targeted services. Panel B similarly reports the exclusion-free bounds and the extension which removes any possible unobserved contact effects. Robust standard errors on the endpoints of each set of bounds, two-way-clustered at the child and investigator level, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.VIII: Estimates of Screen-in and Placement Rates by Subsequent Maltreatment Potential

	(1) Weighted average	(2) With maltreatment potential	(3) Without maltreatment potential
Panel A: Conditional screen-in rates			
Black children	0.632	0.634	0.632
	(0.001)	(0.004)	(0.001)
White children	0.582	0.595	0.579
	(0.001)	(0.003)	(0.001)
Panel B: Placement rates among screened-in calls			
Black children	0.048	0.134	0.029
	(0.002)	(0.018)	(0.004)
White children	0.031	0.076	0.021
	(0.001)	(0.012)	(0.003)
Panel C: Placement rates among all calls			
Black children	0.030	0.079	0.019
	(0.001)	(0.012)	(0.003)
White children	0.018	0.039	0.014
	(0.001)	(0.007)	(0.002)
Number of screeners	162	162	162
Number of investigators	814	814	814

Notes. This table presents estimates of conditional treatment rates in each phase of the CPS system. Panel A presents estimates of $E[S_i|R_i = r, Y_i^* = y]$. Column 1 presents these estimates averaged across cases with and without maltreatment potential (weighted by the relevant μ). Column 2 focuses on cases with maltreatment potential ($E[S_i|R_i = r, Y_i^* = 1]$), while Column 3 focuses on cases without maltreatment potential ($E[S_i|R_i = r, Y_i^* = 0]$). Panels B and C repeat this exercise for $E[D_i|R_i = r, Y_i^* = y, S_i = 1]$ and $E[P_i|R_i = r, Y_i^* = y]$, respectively. Panel A presents robust standard errors two-way clustered at the child and screener level. Robust standard errors in Panel B are two-way clustered at the child and investigator level. Robust standard errors for estimates in Panel C reflect uncertainty in the underlying estimates in panels A and B. All standard errors are obtained via a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.IX: Estimates of Maltreatment Risk and Investigator UD (2008-2017)

	(1) **Baseline**	(2) **2008-2017**
	Panel A: Subsequent maltreatment risk	
Black children	0.155	0.155
	(0.003)	(0.003)
White children	0.175	0.176
	(0.003)	(0.002)
	Panel B: Investigator UD	
Mean across investigators	0.017	0.017
	(0.002)	(0.002)
Number of investigators	814	699

Notes. This table summarizes estimates of mean maltreatment risk among screened-in calls and average (case-weighted) investigator UD. Column 2 presents estimates from the sample of screened-in investigations from 2008–2017, for which we have information on investigators' demographic information. Panel A reports estimates of race-specific average re-investigation risk, while Panel B reports estimates of average investigator UD. Estimates in Panel A come from a linear extrapolation of the variation in Panel B of Figure II. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.X: Covariate-Adjusted Estimates of Maltreatment Risk and Investigator UD

	(1) Baseline	(2) Covariate-Adjusted
	Panel A: Subsequent maltreatment risk	
Black children	0.155	0.138
	(0.003)	(0.002)
White children	0.175	0.168
	(0.003)	(0.002)
	Panel B: Investigator UD	
Mean across investigators	0.017	0.016
	(0.002)	(0.002)
Number of investigators	814	814

Notes. This table summarizes estimates of mean maltreatment risk among screened-in calls and average (case-weighted) investigator UD for different re-investigation time horizons. Panel A reports estimates of race-specific average re-investigation risk, while Panel B reports estimates of average investigator UD. Estimates in Panel A come from a linear extrapolation of the variation in Panel B of Figure II, but where placement and re-investigation rates adjust for the covariates in Column 4 of Table I in addition to ZIP code by year fixed effects. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.XI: Heterogeneity in Investigator UD by Child and Investigation Characteristics

	(1) All Children	(2) Male Children	(3) Female Children	(4) Young Children	(5) Old Children	(6) Physical Abuse	(7) No Physical Abuse
	Panel A: Subsequent maltreatment risk						
Black children	0.155	0.137	0.170	0.150	0.123	0.138	0.146
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.008)	(0.002)
White children	0.175	0.160	0.198	0.187	0.138	0.166	0.165
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.006)	(0.002)
	Panel B: Investigator UD						
Mean across investigators	0.017	0.014	0.017	0.018	0.011	0.010	0.013
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)
Number of investigators	814	838	757	657	901	223	1277

Notes. This table summarizes estimates of race-specific mean risk among screened-in calls and average investigator UD, separately by child and investigation characteristics. For each subgroup, we require that an investigator handled at least 100 cases in order to be included in the sample. Therefore, the number of investigators varies across the columns depending on how many investigators in the sample met the requirement. We define young as a child who is 7 years old or younger (the median age in our sample). Estimates come from a linear extrapolation of the variation in Panel B of Figure II. However, the extrapolations are done within the given characteristic. Panel A reports estimates of race-specific average maltreatment risk, while Panel B reports estimates of average (case-weighted) investigator UD. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.XII: Hierarchical MTE Model Estimates

	(1) White Children	(2) Black Children	(3) Diff.
Mean Risk	0.216	0.192	0.024
	(0.001)	(0.002)	(0.002)
Mean Marginal Outcome	0.929	0.875	0.053
	(0.018)	(0.019)	(0.026)
Mean Signal Quality	4.579	3.049	1.530
	(2.992)	(1.126)	(3.219)
Marginal Outcome Std. Dev.	0.178	0.229	-0.051
	(0.019)	(0.014)	(0.022)
Number of investigators	699	699	–

Notes. This table reports simulated minimum distance estimates of moments of the MTE model. The model is estimated on the sample of investigations from 2008 to 2017, for which we have information on investigators' demographic information. Robust standard errors, two-way clustered at the child and the investigator level, are obtained by a bootstrapping procedure (with 500 replications times) and appear in parentheses.

Table A.XIII: Robustness of Estimates of Maltreatment Risk and Investigator UD

	(1) Baseline	(2) No Strata Adjustment	(3) No Weighting	(4) Empirical Bayes Shrinkage
	Panel A: Subsequent maltreatment risk			
Black children	0.155	0.145	0.147	0.158
	(0.003)	(0.003)	(0.004)	(0.004)
White children	0.175	0.170	0.166	0.183
	(0.003)	(0.002)	(0.003)	(0.003)
	Panel B: Investigator UD			
Mean across investigators	0.017	0.016	0.015	0.016
	(0.002)	(0.002)	(0.002)	(0.002)
Number of investigators	814	814	814	814

Notes. This table summarizes estimates of mean maltreatment risk among screened-in calls and average (case-weighted) investigator UD across different specifications. Estimates in Column 2 come from a linear extrapolation of the variation in Panel B of Figure II, but where the placement and subsequent maltreatment rates are estimated without adjusting for ZIP code by year fixed effects. Estimates in Column 3 come from a linear extrapolation of the variation in Panel B of Figure II, but, unlike our baseline extrapolation, the specification is not inversely weighted by the variance of each investigator's subsequent maltreatment rate. Finally, estimates in Column 4 come from a linear extrapolation of the variation in Panel B of Figure II, but where we shrink estimates of the placement rate using conventional Empirical Bayes shrinkage. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.XIV: Estimates of Maltreatment Risk and Investigator UD (Shorter Time Horizons)

	(1) Baseline	(2) Inv. within 5 months	(3) Inv. within 4 months	(4) Inv. within 3 months	(5) Inv. within 2 months
	Panel A: Subsequent maltreatment risk				
Black children	0.155	0.133	0.110	0.085	0.059
	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
White children	0.175	0.153	0.128	0.098	0.065
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
	Panel B: Investigator UD				
Mean across investigators	0.017	0.016	0.016	0.015	0.015
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of investigators	814	814	814	814	814

Notes. This table summarizes estimates of mean maltreatment risk among screened-in calls and average (case-weighted) investigator UD for different re-investigation time horizons. Panel A reports estimates of race-specific average re-investigation risk, while Panel B reports estimates of average investigator UD. Estimates in Panel A come from a linear extrapolation of the variation in Panel B of Figure II. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (500 times) and appear in parentheses.

Table A.XV: Estimates of Maltreatment Risk and Investigator UD (Alternative Outcomes)

	(1) Baseline	(2) Confirmed victim within 6 months	(3) Placed within 6 months	(4) Abuse inv. within 6 months	(5) Neglect Inv. within 6 months	(6) Confirmed abuse within 6 months	(7) Confirmed neglect within 6 months
	Panel A: Subsequent maltreatment risk						
Black children	0.155	0.032	0.008	0.042	0.133	0.007	0.030
	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)
White children	0.175	0.040	0.009	0.045	0.150	0.008	0.036
	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
	Panel B: Investigator UD						
Mean across investigators	0.017	0.014	0.014	0.015	0.016	0.015	0.014
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of investigators	814	814	814	814	814	814	814

Notes. This table summarizes estimates of mean maltreatment risk among screened-in calls and average (case-weighted) investigator UD for different outcome variables. Panel A reports estimates of race-specific average re-investigation risk, while Panel B reports estimates of average investigator UD. Estimates in Panel A come from a linear extrapolation of the variation in Panel B of Figure II. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

Table A.XVI: Heterogeneity in Investigator UD by Reporting Source

	(1) Baseline	(2) Call from Mandated Reporter	(3) Call from Non-mandated Reporter
	Panel A: Subsequent maltreatment risk		
Black children	0.155	0.176	0.149
	(0.003)	(0.009)	(0.009)
White children	0.175	0.194	0.179
	(0.003)	(0.007)	(0.010)
	Panel B: Investigator UD		
Mean across investigators	0.017	0.018	0.018
	(0.002)	(0.004)	(0.006)
Number of investigators	814	121	75

Notes. This table summarizes estimates of mean maltreatment risk among screened-in calls and average (case-weighted) investigator UD. Columns 2 and 3 come from screened-in calls in the screener analysis sample, for which we have information on the reporting source of the maltreatment allegation. A mandated reporter is defined as someone who is either educational personnel, medical personnel, a social worker, or a member of law enforcement. Panel A reports estimates of race-specific average re-investigation risk, while Panel B reports estimates of average investigator UD. Estimates in Panel A come from a linear extrapolation of the variation in Panel B of Figure II, but estimated on the subset of screened-in calls in the screener sample. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (with 500 replications) and appear in parentheses.

B   Exclusion Restriction Tests and Extensions

Bounding Screening-In Effects

As described in Section V.D., the primary concern for the exclusion restriction in the screener analysis is the possibility of direct effects of screening-in decisions on future maltreatment potential. To formalize this concern, let $Y_i^{*,S=1}$ denote child i's potential for future at-home maltreatment when the case is screened-in and let $Y_i^{*,S=0}$ correspondingly denote the potential at-home maltreatment outcome when the case is screened-out. Among the full set of cases that do not result in foster care placement (with $P_i = 0$), we observe:

$$Y_i^* = (1 - S_i)Y_i^{*,S=0} + S_iY_i^{*,S=1} \tag{16}$$

where $S_i \in \{0,1\}$ again indicates the screening-in of case i. This expression shows that extrapolations of rates of at-home maltreatment Y_i^* across screeners with different rates of screening S_i could potentially be biased when screening-in effects $Y_i^{*,S=1} - Y_i^{*,S=0}$ are non-zero.

To develop our strategy for bounding such effects, note that we can use (16) to write:

$$
\begin{aligned}
(1 - P_i)Y_i^* &= (1 - P_i)(1 - S_i)Y_i^{*,S=0} + (1 - P_i)S_iY^{*,S=1} \\
&= (1 - S_i)Y_i^{*,S=0} + S_i\left[(1 - D_i)Y^{*,S=1}\right],
\end{aligned}
\tag{17}
$$

where $D_i \in \{0,1\}$ again indicates the placement decision among screened-in cases. The second line follows from the fact that $P_i = S_iD_i$, so $(1 - P_i)(1 - S_i) = (1 - S_iD_i)(1 - S_i) = 1 - S_i$ and $(1 - P_i)S_i = (1 - S_iD_i)S_i = (1 - D_i)S_i$. Equation (17) gives a potential outcome model for observed outcome $(1 - P_i)Y_i^*$ in terms of the treatment S_i, with an untreated potential outcome of $Y_i^{*,S=0}$ and a treated potential outcome of $(1 - D_i)Y^{*,S=1}$.

Consider a binary instrument Z_i which is as-good-as-randomly assigned, only affects $(1 - P_i)Y_i^*$ through the screening-in treatment S_i, and satisfies the usual first-stage monotonicity condition. By Equation (17), an IV regression of $(1 - P_i)Y_i^*$ on S_i that instruments with this Z_i identifies a local average treatment effect:

$$\frac{E[(1 - P_i)Y_i^*|Z_i = 1] - E[(1 - P_i)Y_i^*|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]} = E[(1 - D_i)Y_i^{*,S=1} - Y_i^{*,S=0}|S_{i1} > S_{i0}], \tag{18}$$

where S_{iz} indicates the potential screening status when $Z_i = z$. When nobody is placed into foster care ($D_i = 0$), this expression captures the average effect of a screened-in investigation on subsequent maltreatment among instrument compliers: i.e., $E[Y^{*,S=1} - Y_i^{*,S=0}|S_{i1} > S_{i0}]$. Otherwise, we can bound this effect using additional information on the average foster care

placement rate of compliers, $E[D_i \mid S_{i1} > S_{i0}]$.

To derive these bounds, first note that the average $Y_i^{*,S=0}$ for compliers is identified by an IV regression of $(1 - S_i)(1 - P_i)Y_i^*$ on $1 - S_i$ which instruments with Z_i:

$$\frac{E[(1 - S_i)(1 - P_i)Y_i^*|Z_i = 1] - E[(1 - S_i)(1 - P_i)Y_i^*|Z_i = 0]}{E[1 - S_i|Z_i = 1] - E[1 - S_i|Z_i = 0]} = E[Y_i^{*,S=0}|S_{i1} > S_{i0}]. \quad (19)$$

Further observe, since $Y_i^{*,S=1}$ is binary:

$$E[Y_i^{*,S=1}|S_{i1} > S_{i0}]$$
$$\in \left[E[(1 - D_i)Y_i^{*,S=1}|S_{i1} > S_{i0}], 1 - E[(1 - D_i)(1 - Y_i^{*,S=1})|S_{i1} > S_{i0}] \right]. \quad (20)$$

The lower bound in Equation (20) is identified by an IV regression of $S_i(1 - P_i)Y_i^*$ on S_i which instruments with Z_i:

$$\frac{E[S_i(1 - P_i)Y_i^*|Z_i = 1] - E[S_i(1 - P_i)Y_i^*|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]} = E[(1 - D_i)Y_i^{*,S=1}|S_{i1} > S_{i0}]. \quad (21)$$

The upper bound can moreover be written:

$$E[(1 - D_i)(1 - Y_i^{*,S=1})|S_{i1} > S_{i0}] = 1 - E[D_i|S_{i1} > S_{i0}] - E[(1 - D_i)Y_i^{*,S=1}|S_{i1} > S_{i0}]. \quad (22)$$

This is a function of the identified $E[(1 - D_i)Y_i^{*,S=1}|S_{i1} > S_{i0}]$ and $E[D_i|S_{i1} > S_{i0}]$, which is identified by an IV regression of $S_i D_i$ on S_i which instruments with Z_i:

$$\frac{E[S_i D_i|Z_i = 1] - E[S_i D_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]} = E[D_i|S_{i1} > S_{i0}]. \quad (23)$$

We apply these bounds with Z_i indicating the assignment of case i to a screener with an above-median tendency to screen-in cases, computed as in Table A.I. Specifically, we use this instrument to estimate complier-average D_i and $(1 - D_i)Y_i^{*,S=1}$ by IV regressions of $S_i D_i$ and $S_i(1 - P_i)Y_i^*$, respectively, on S_i instrumenting by Z_i and adjusting for screener randomization strata. We use these to form bounds on complier-average $Y_i^{*,S=1}$, following Equations (20) and (22). Finally, we combine these bounds with an estimate of complier-average $Y_i^{*,S=0}$—obtained from an IV regression of $(1 - S_i)(1 - P_i)Y_i^*$ on $1 - S_i$, with the same instrument and controls—to bound the complier-average screening-in effect $Y_i^{*,S=1} - Y_i^{*,S=0}$.

The estimated lower and upper bounds from this exercise are, respectively, -0.006 and 0.014 with a standard error (clustered by child and screener) of 0.030 on both.

Exclusion-Free Investigator UD Bounds

To formalize the exclusion-free UD bounds described in Section V.D., consider a population of screened-in cases i of race R_i with placement status $D_i \in \{0,1\}$ and subsequent maltreatment potential $Y_i^* \in \{0,1\}$. As with Equations (10)-(11), we have:

$$E[D_i \mid R_i = r, Y_i^* = 1] = 1 - \frac{E[(1 - D_i)Y_i^* \mid R_i = r]}{E[Y_i^* \mid R_i = r]} \tag{24}$$

and

$$E[D_i \mid R_i = r, Y_i^* = 0] = 1 - \frac{E[(1 - D_i)(1 - Y_i^*) \mid R_i = r]}{1 - E[Y_i^* \mid R_i = r]}, \tag{25}$$

The numerators in these expressions are directly estimable, since Y_i^* is observed when $D_i = 0$. The mean risk parameters in the denominators are furthermore bounded:

$$E[Y_i^* \mid R_i = r] \in \left[E[(1 - D_i)Y_i^* \mid R_i = r], 1 - E[(1 - D_i)(1 - Y_i^*) \mid R_i = r] \right], \tag{26}$$

with the upper and lower bounds again identified. We combine Equations (24)-(26) to bound:

$$\overline{\Delta}^D = \left(E[D_i \mid R_i = b, Y_i^* = 0] - E[D_i \mid R_i = w, Y_i^* = 0] \right) \left(1 - E[Y_i^*] \right) \tag{27}$$
$$+ \left(E[D_i \mid R_i = b, Y_i^* = 1] - E[D_i \mid R_i = w, Y_i^* = 1] \right) E[Y_i^*].$$

Three points are worth noting about these exclusion-free bounds on $\overline{\Delta}^D$. First, they only require estimates of aggregate (state-level) placement rates and at-home maltreatment rates by race. They thus relax all assumptions on investigator assignment along with the exclusion restriction. Second, and relatedly, the $\overline{\Delta}^D$ estimand differs from our baseline investigator UD measure Δ^D by not accounting for the possibly non-random sorting of investigators across offices. If, for example, investigators with higher placement rates tend to work in offices that see a higher share of white cases then $\overline{\Delta}^D$ will tend to understate Δ^D. Third, the $\overline{\Delta}^D$ measure of investigator UD conditions on a Y_i^* which may be potentially affected by investigators through targeted services or unobserved contact effects.

Estimated bounds on $\overline{\Delta}^D$ are reported in Appendix Table A.VII along with two extensions that adjust the exclusion-free bounds to remove a possible influence of direct investigator effects. To formalize the first extension, which removes effects of targeted services, let $T_i \in \{0, 1\}$ indicate the observed assignment of such services in case i. Write $Y_i^* = Y_i^*(T_i)$ where $Y_i^*(t)$

22

denotes potential subsequent maltreatment in case i when $T_i = t$. Consider the estimand:

$$\overline{\Delta}^{D,T=0} = (E[D_i \mid R_i = b, Y_i^*(0) = 0] - E[D_i \mid R_i = w, Y_i^*(0) = 0]) \, (1 - E[Y_i^*(0)]) \qquad (28)$$
$$+ (E[D_i \mid R_i = b, Y_i^*(0) = 1] - E[D_i \mid R_i = w, Y_i^*(0) = 1]) \, E[Y_i^*(0)].$$

This UD measure is identical to $\overline{\Delta}^D$ except that it conditions on potential subsequent maltreatment in the absence of targeted services, $Y_i^*(0)$, rather than the service-inclusive maltreatment potential Y_i^*. As before, we bound $\overline{\Delta}^{D,T=0}$ by first rewriting its components:

$$E[D_i \mid R_i = r, Y_i^*(0) = 1] = 1 - \frac{E[(1 - D_i)Y_i^*(0) \mid R_i = r]}{E[Y_i^*(0) \mid R_i = r]} \qquad (29)$$

and

$$E[D_i \mid R_i = r, Y_i^*(0) = 0] = 1 - \frac{E[(1 - D_i)(1 - Y_i^*(0)) \mid R_i = r]}{1 - E[Y_i^*(0) \mid R_i = r]}. \qquad (30)$$

Unlike before, the numerators in these expressions are only partially identified. Write:

$$E[(1 - D_i)Y_i^*(0) \mid R_i = r] = E[Y_i^*(0) \mid D_i = 0, R_i = r]Pr(D_i = 0 \mid R_i = r) \qquad (31)$$

and

$$E[(1 - D_i)(1 - Y_i^*(0)) \mid R_i = r] = (1 - E[Y_i^*(0) \mid D_i = 0, R_i = r])Pr(D_i = 0 \mid R_i = r). \qquad (32)$$

Here $Pr(D_i = 0 \mid R_i = r)$ is directly estimable, and with $Y_i^* = Y_i^*(0) + T_i(Y_i^*(1) - Y_i^*(0))$:

$$E[Y_i^*(0) \mid D_i = 0, R_i = r] = E[Y_i^* \mid D_i = 0, R_i = r] - E[T_i \mid D_i = 0, R_i = r] \times \theta \qquad (33)$$

where $\theta = E[Y_i^*(1) - Y_i^*(0) \mid T_i = 1, D_i = 0, R_i = r]$ denotes the average effect of targeted service assignment on those assigned and not removed from home. Given bounds on this parameter, $\theta \in [\theta^L, \theta^U]$, and direct estimates of both $E[Y_i^* \mid D_i = 0, R_i = r]$ and $E[T_i \mid D_i = 0, R_i = r]$, we can thus bound:

$$E[Y_i^*(0) \mid D_i = 0, R_i = r] \in E[Y_i^* \mid D_i = 0, R_i = r] - E[T_i \mid D_i = 0, R_i = r] \times [\theta^U, \theta^L]. \qquad (34)$$

Finally, as before, the denominators in Equations (29) and (30) are bounded by functions of the numerators:

$$E[Y_i^*(0) \mid R_i = r] \in [E[(1 - D_i)Y_i^*(0) \mid R_i = r], 1 - E[(1 - D_i)(1 - Y_i^*(0)) \mid R_i = r]]. \qquad (35)$$

Thus, to bound $\overline{\Delta}^{D,T=0}$, we first use *a priori* bounds on θ to bound the numerators in Equations (29) and (30) following Equations (31), (32), and (34). We then use these to bound the denominators in Equations (29) and (30) following Equation (35). Finally, we search within both sets of bounds to find the combination that minimizes and maximizes Equation (28).

In Panel A of Appendix Table A.VII we construct these bounds by putting *a priori* bounds on service takeup rates, allowing for arbitrarily large effects of services among those that receive them. Specifically, we set $[\theta^L, \theta^U] = [-\tau, \tau]$ where τ is a range of possible takeup rates from 0.1 to 0.6. The lower bound of $-\tau$ corresponds to a scenario where services reduce maltreatment potential (from $Y_i^* = 1$ to $Y_i^* = 0$) in all cases that receive them, while the upper bound corresponds to a scenario where services increase maltreatment potential in all such cases.

The second extension in Panel B of Appendix Table A.VII follows similarly, except that we bound a version of $\overline{\Delta}^D$ that conditions on a measure of potential subsequent maltreatment in the absence of unobserved contact effects (again written $Y_i^*(0)$, in a slight abuse of notation). Here we use bounds of:

$$E[Y_i^*(0) \mid D_i = 0, R_i = r] \in E[Y_i^* \mid D_i = 0, R_i = r] - [\pi^U, \pi^L]. \tag{36}$$

where $[\pi^L, \pi^U]$ are *a priori* bounds on the contact effects. We then bound $E[Y_i^*(0) \mid R_i = r]$ and the corresponding unwarranted disparity measure following identical steps. In practice, we set $[\pi^L, \pi^U] = [-\pi, \pi]$ where π ranges from 1 to 6 percentage points.

C   IMPACTS ON WELFARE-RELEVANT OUTCOMES

This section assesses how unwarranted racial disparities in the decisions of CPS investigators impact racial gaps in contact with the criminal justice system by age 19, the main outcome in Baron and Gross (2022). It also examines the effects of proposed policies aimed at equalizing the placement rates of screened-in children by race, a recent consideration in the policy sphere.

To do so, we combine *(i)* observed racial gaps in adult convictions in the Michigan data, *(ii)* estimates of race-specific effects of foster care on later-in-life convictions, and *(iii)* estimates of investigator UDs from the current study. While we focus on adult convictions, being a particularly important and well-measured outcome, this analysis shows how similar calculations can be applied with other outside estimates of *(i)* and *(ii)*.

The inputs to this analysis are as follows. In the Michigan data used in Baron and Gross (2022), 9% of Black children and 7% of white children are convicted by age 19. Using IV, Baron and Gross (2022) estimate that foster care placement decreases adult conviction by

25 percentage points (60%) for Black children and by 24 percentage points (86%) for white children. In the current study, unwarranted racial disparities among screened-in children are primarily concentrated among children with future maltreatment potential ($Y_i^* = 1$), who make up 17% of white children and 15% of Black children (see Table II). In these subpopulations, 13% of Black children and 7% of white children are placed in foster care (see Table A.VIII).

To assess how much these disparities in placement rates impact gaps in conviction by age 19, we will assume that the race-specific LATEs in Baron and Gross (2022)—which capture conviction effects among children at the margin of placement—are a reasonable approximation of the average race-specific causal effects of foster care among screened-in children with $Y_i^* = 1$. This assumption seems reasonable, since our mean marginal outcome estimates in Table A.XII suggest 88-92% of children at the margin of placement have $Y_i^* = 1$.

Because the effects of foster care on conviction are beneficial for $Y_i^* = 1$ children of either race, eliminating investigator UDs would only increase the racial gap in adult convictions. Suppose, for example, we reduced the placement rate of screened-in Black children with $Y_i^* = 1$ to equal that of screened-in white children. Currently, about 1.2% of screened-in white children are placed in foster care (7% of $Y_i^* = 1$ children, who make up around 17% of the white child sample), compared to 2% of screened-in Black children (13% of $Y_i^* = 1$ children, who make up around 15% of the white child sample). Therefore, in this counterfactual, we would reduce the share of Black children placed into foster care by roughly 0.8 percentage points. Multiplying this decline by the LATE of foster care on later-in-life convictions for Black children predicts that the racial gap in adult convictions would increase from a 2 percentage point gap to a 2.2 percentage point gap. That is, based on these estimates, 9.2% of Black children would be convicted by age 19 compared to 7% of white children.

Similarly, consider a counterfactual where we eliminate unwarranted racial disparity by increasing the share of white children placed in foster care from 1.2% to 2%: a 0.8 percentage point increase. Multiplying this increase by the LATE of foster care on later-in-life convictions for white children predicts that the gap in adult convictions would similarly increase from a 2 percentage point gap to a 2.2 percentage point gap. That is, based on these estimates, 9% of Black children would be convicted by age 19 vs. 6.8% of white children.

Similar counterfactuals can be conducted with other estimates of foster care impacts. We focus on the estimates in Baron and Gross (2022), due to the similarity of settings and sample periods between their study and ours. However, this exercise shows how other estimates might result in different counterfactuals. Doyle (2007, 2008), for example, finds more adverse effects of foster care among marginal placements, though it is less clear how these map to

the long-run effects of children with $Y_i^* = 1$ in the current study given potentially different margins and different foster care systems. Nevertheless, this exercise emphasizes that the welfare implications of foster care placement disparities depend crucially on the particular subpopulation in which disparities exist (e.g., $Y_i^* = 1$ versus $Y_i^* = 0$) and the causal effects of foster care in the relevant subpopulation.

D Potential Drivers of Investigator UD

The main analysis shows unwarranted disparity in investigators' foster care placement decisions in Michigan, both on average and for various subgroups. Such disparities could arise from racially biased preferences and beliefs (e.g., Becker (1957); Bordalo et al. (2016); Bohren et al. (2020)) or accurate statistical discrimination (e.g., Phelps (1972); Arrow (1973); Aigner and Cain (1977)), as well as indirect discrimination through non-race characteristics (e.g. Bohren, Hull, and Imas, 2022). Our primary analysis suggests that non-race characteristics play a limited role, leaving on the table classic models of race-based bias and statistical discrimination.

This section imposes additional structure on the quasi-experimental variation in order to understand the role that racial bias and statistical discrimination play in shaping investigator UD. We follow the model and estimation approach in Arnold, Dobbie, and Hull (2022). Specifically, we fit a hierarchical marginal treatment effect (MTE) model to the quasi-experimental variation in investigator placement and subsequent maltreatment rates. We first present a behavioral model of individual investigator placement decisions and then show that this model parameterizes a set of investigator- and race-specific MTE frontiers that capture racial bias and statistical discrimination. We then estimate the model via simulated minimum distance (SMD), matching moments of the quasi-experimental variation to those implied by the model.

Model Setup

Assume each investigator j observes a noisy signal of subsequent maltreatment potential for case i, $\nu_{ij} = Y_i^* + \eta_{ij}$, with conditionally normally distributed noise: $\eta_{ij} \mid Y_i^*, (R_i = r) \sim N(0, \sigma_{jr}^2)$. The "quality" (i.e. precision) of risk signals $\tau_{ij} = 1/\sigma_{jr}$ is allowed to vary both across investigators and race. Investigators with a higher τ_{jr} can be seen as being more skilled at inferring potential for subsequent maltreatment, either by having a richer information set or by a higher ability to infer true potential conditional on an information set. We assume that investigators form accurate posterior risk predictions from the noisy signal and the child's race: $p_j(\nu_{ij}, R_i) = Pr(Y_i^* = 1 | \nu_{ij}, R_i)$. Each investigator further has a subjective benefit of

leaving at home children of race r, $\pi_{jr} \in (0,1)$. Investigators leave at home all children for whom this benefit exceeds the posterior risk cost, yielding the following decision rule:

$$D_{ij} = \mathbb{1}[\pi_{jR_i} \geq p_j(\nu_{ij}, R_i)] \tag{37}$$

Racial bias, as in Becker (1957), arises when an investigator has a different subjective benefit from leaving at home white and Black children with the same posterior risk (e.g., $\pi_{jb} < \pi_{jw}$). Racial bias leads to UDs against the group with the lower benefit, since the investigator generally makes different decisions for children with the same maltreatment potential.[35]

Statistical discrimination, as in Aigner and Cain (1977), arises when investigators set the same threshold by race but discriminate because risk predictions are impacted by differences across race in either μ_r or τ_{jr}. Differences in μ_r will tend to lead to higher placement rates for children in the group with higher risk, resulting in UDs against that group. However, statistical discrimination due to differences in τ_{jr} has an ambiguous effect on UDs. For instance, if $\pi_{jr} > \mu_r$ for each r, then noisier signals will lead fewer children of that race being placed in foster care, since investigators will put more weight on μ_r.

Note that the model allows both racial bias and statistical discrimination to arise indirectly from non-race characteristics, such as income or maltreatment type. For instance, an investigator may inadvertently set race-specific thresholds by penalizing certain types of neglect (such as improper supervision) that may be correlated with race—though as mentioned above we find little empirical support for non-race characteristics driving UD in our context.

To estimate the model, we first re-write Equation (37) as $D_{ij} = \mathbb{1}[\Pi_{jR_i} \geq U_{ij}]$ where $U_{ij} \mid R_i$ is conditionally uniformly distributed by applying a probability transformation to $p_j(\nu_{ij}, R_i)$. This defines a conditional MTE frontier:

$$\mu_{jr}(t) = E[Y_i^* \mid U_{ij} = t, R_i = r] \tag{38}$$

where $\mu_{jr}(t)$ represents the effect of being left at home on subsequent maltreatment (Y_i^*) for children of race r that investigator j perceives to be at the $(t \times 100\text{th})$ percentile of risk. $\Pi_{jr} = E[D_{ij} \mid R_i = r]$ parameterizes the leave-at-home rate of investigator j, and $\int_0^{\Pi_{jr}} \mu_{jr}(t)dt = E[Y_i^* \mid D_{ij} = 1, R_i = r]$.

---

[35]Inaccurate racial stereotyping tends to be observationally equivalent to racial animus, and can therefore similarly result in UDs (Arnold, Dobbie, and Yang, 2018; Arnold, Dobbie, and Hull, 2022; Hull, 2021; Bohren et al., 2020). In particular, Arnold, Dobbie, and Hull (2022) show that this model with accurate beliefs and biased risk thresholds is observationally equivalent to a model with biased priors on Y_i^* and equal risk thresholds by race.

Differences in an investigator's MTE curves by race, evaluated at her leave-at-home threshold, Π_{jr}, yields a marginal outcome test for racial bias in her leave-at-home decisions (Arnold, Dobbie, and Hull, 2022; Arnold, Dobbie, and Yang, 2018; Hull, 2021). This is because leave-at-home impacts on subsequent maltreatment, at the margin of leaving at home, capture an investigator's specific leave-at-home benefits: $\mu_{jr}(\Pi_{jr}) = \pi_{jr}$. Therefore, the race-specific MTEs will be equal when the investigator is racially unbiased. Alternatively, marginal white children will have higher rates of subsequent maltreatment if the investigator is racially biased against Black children.

Investigator- and race-specific MTE frontiers can also be used to quantify statistical discrimination. The mean risk of race r is given by integrating the MTE frontier of any investigator: $\mu_r = \int_0^1 \mu_{jr}(t)dt$. As Arnold, Dobbie, and Hull (2022) show, the slopes of these curves furthermore capture the quality of an investigator's risk signals: an investigator with $\tau_{jw} > \tau_{jb}$, for instance, will have a steeper-sloping $\mu_{jw}(\cdot)$ than $\mu_{jb}(\cdot)$.

Because the parameterization of investigator skill and preferences in this model is very flexible, we face an underidentification challenge. We follow Arnold, Dobbie, and Hull (2022) in overcoming this challenge by parameterizing the distribution of investigator signal quality. This parameterization allows for heterogeneous MTE curves across investigators (which amounts to a first-stage monotonicity assumption and uniform investigator skills), and leads to a hierarchical MTE model.

SMD Estimator

The model parameterization uses the fact that $p_j(\nu, r)$ is strictly increasing in ν, and is therefore invertible by race:

$$D_{ij} = \mathbb{1}[\pi_{jR_i} \geq p_j(\nu_{ij}, R_i)] = \mathbb{1}[\kappa_{jR_i} \geq Y_i^* + \eta_{ij}] \tag{39}$$

where $\kappa_{jr} = p_j^{-1}(\pi_{ij}, r)$ is a normalized signal threshold. We model κ_{jr} and $\ln\tau_{jr}$ as being joint-normally distributed (independently across investigators) with a separate mean and variance by race. The log normality of τ_{jr} imposes the constraint of positive signal precision. This yields a higher-level parameter vector Θ containing μ_r and the means, variances, and covariances or κ_{jr} and $\ln\tau_{jr}$.

We estimate the model by a minimum distance procedure based on the intuition in Section V.B. in Arnold, Dobbie, and Hull (2022): we find the values of Θ that can best match key features of the distribution of model-implied leave-at-home and subsequent maltreatment rates to the corresponding features of estimated leave-at-home and subsequent maltreatment

rates in Panel B of Figure II. The features we match are the race-specific mean and variance of investigator leave-at-home rates, and the race-specific intercept and slope from quadratic regressions of investigator subsequent maltreatment rates on investigator leave-at-home rates. As in our primary analysis, we adjust for ZIP code by year fixed effects to ensure investigator assignment is as good as random.

Results

Table A.XII reports SMD estimates of the race-specific moments: the mean maltreatment risk and the first and second moments of both marginal leave-at-home outcomes and signal quality across investigators.

We find that white children have a 2.4 percentage points (SE=0.002) higher mean subsequent maltreatment risk relative to Black children. This estimate is extremely similar to the estimates in Table II (2 percentage points).

We also find that white children have higher mean marginal leave-at-home outcomes relative to Black children, implying racial bias per the discussion above. Mean subsequent maltreatment risk is 0.929 (SE=0.018) for marginal white children, compared to 0.875 (SE=0.019) for marginal Black children. The difference in marginal outcomes is a statistically significant 5.3 percentage points (SE=0.026).

We find an average signal quality of 4.579 (SE=2.992) for white children, and 3.049 (SE=1.126) for Black children. The difference between the two (1.530) is not statistically significant, though it is imprecisely estimated. Given this imprecision, we further probe whether signal quality varies by race by testing that the slopes in Figure II are equal by race.[36] Because all three extrapolations are virtually identical, we focus on the linear extrapolation for this exercise. We fail to reject the null hypothesis of equal slopes ($p = 0.70$).

These results suggest that statistical discrimination is not a primary driver of UDs in our setting: average signal quality is similar by race, and average risk is higher for white children, which rules out first-order statistical discrimination as the reason behind higher placement rates for Black children. Rather, the results in this section suggest that either racial bias or inaccurate racial stereotyping may be the primary drivers of UDs in our context.

---

[36] As discussed in Arnold, Dobbie, and Hull (2022), differences in signal quality by race in this model would manifest as different slopes by race in Figure II.

E Nationwide Estimates From NCANDS Data

This section explores the generalizability of our key findings on investigator UD to other states. Section V. showed that screened-in Black children in Michigan are placed at higher rates than screened-in white children, conditional on maltreatment potential, and that this UD is primarily driven by cases with maltreatment potential. A natural question is whether these findings are unique to Michigan.

To explore generalizability, we use more limited data from the National Child Abuse and Neglect Data System Child files (NCANDS, 2023), sourced from the National Data Archive on Child Abuse and Neglect at Cornell University. This dataset can be accessed by researchers free of charge via application. NCANDS is a voluntary data collection system that gathers information regarding reports of child abuse and neglect that were investigated by CPS (i.e., it contains only "screened-in" calls). We use data from 2008 to 2019, consistent with the sample period of our main analyses.

The NCANDS data include information about whether a child was placed in foster care and the child's race. Importantly, however, the data do not include investigator identifiers. This prevents us from using quasi-random investigator assignment or the extrapolation methods to point-identify investigator UDs in each state. We instead leverage the fact that the non-parametric bounds in Panel B of Figure III can be applied to state-level aggregate statistics in order to bound race-specific maltreatment potential.[37] Recall that a lower-bound on $\mu_r^{S=1}$ can be derived by assuming that all of the children of race r who were placed in foster care (and for whom we cannot observe future maltreatment potential) would have been safe in their homes. An upper bound can be derived by assuming that all such children would have experienced subsequent maltreatment in their homes.

Because the NCANDS data do not contain investigator identifiers, we apply the non-parametric bounds to state-level aggregate statistics: the overall race-specific average placement rate and the overall average subsequent maltreatment rate (measured as a subsequent investigation within six months) among those not placed in foster care. As discussed in Online Appendix B, a subtlety to the interpretation of these aggregate-data investigator UDs is that they do not adjust for investigator rotation (ZIP code by year) fixed effects, as in our main analysis. In principle, the aggregate-data UDs might therefore be driven

---

[37]We make the same general sample restrictions in the NCANDS data as in our investigator analysis sample, dropping cases of sexual abuse and repeat investigations since these tend to not be quasi-randomly assigned in Michigan and many other states. We keep only white and Black children in the data and drop a small number of observations with invalid child numeric identifiers (which affects roughy 0.35% of cases in Michigan over this time period). Because we do not observe investigator identifiers, however, we are unable to restrict the sample to cases assigned to investigators who were assigned at least 200 cases over the sample period.

by the differential sorting of cases to investigators over time or across regions within a state. As we show in the main analysis, adjusting for the rotation fixed effects has little effect on results in Michigan (Table A.XIII). But we are unable to directly test this in other states.[38]

Recall that the width of the bounds of race-specific risk among screened-in children in a given state will be equal to the race-specific placement rate among screened-in children in that state. The bounds are therefore likely to be informative, since treatment rates tend to be low in CPS. For example, only seven states in the NCANDS dataset have overall placement rates that are above 10% during this time period.

Our main findings using administrative data from Michigan (presented in Section V.) are evident in NCANDS data for Michigan as well, despite differences in the analysis sample. Figure A.II summarizes this comparison. The left-hand side of the figure shows non-parametric bounds using administrative data from Michigan, both for overall investigator UD (as previously reported in Panel B of Figure III) and separately for cases with and without maltreatment potential. The right-hand side replicates these results using NCANDS data.

The UD range in the MDHHS data is [0.012, 0.021] while the range in the NCANDS data is [0.009, 0.012]. Thus, the two datasets show that, in Michigan, overall UD in investigator's placement decisions is positive: Screened-in Black children are up to 1.2 percentage points more likely to be placed than screened-in white children with the same maltreatment potential in the NCANDS data. Moreover, the overall investigator UD tends to be primarily driven by disparities in cases with maltreatment potential. The average estimate within the bounds in the MDHHS data is 0.8 percentage points in cases without maltreatment potential and 6 percentage points in cases with maltreatment potential. The average estimates within the bounds in the NCANDS data are 0.5 percentage points and 3 percentage points, respectively.

Having shown that we can replicate our main findings in the more limited NCANDS data, we extend the analysis to other states to examine whether our findings are generalizable. Using the same approach as above, we estimate non-parametric bounds for each state in the NCANDS data for 2008 to 2019, separately for cases with and without maltreatment potential. We then take the average estimate in each state-specific bounds and plot this estimate in Figure IV. Section VI.C. discusses these results.

---

[38]As mentioned in Online Appendix B, a related virtue of these aggregate-data UDs is that they do not require quasi-random investigator assignment. Investigators have been shown to be quasi-randomly assigned to cases in many states besides Michigan, such as Illinois (Doyle, 2007, 2008), Rhode Island (Bald et al., 2022), and South Carolina (Roberts, 2019). But this is likely not the case in every CPS system.