Racial Discrimination in Child Protection*

E. Jason Baron† Joseph J. Doyle, Jr.‡ Natalia Emanuel§

Peter Hull¶ Joseph Ryan‖

July 2023

Abstract

Ten percent of Black children in the U.S. spend time in foster care—twice the rate of white children. We estimate unwarranted disparities in foster care placement decisions, adjusting for differences in the potential for future maltreatment leveraging the quasi-random assignment of cases to investigators. Using a sample of nearly 220,000 maltreatment investigations, we find that Black children are 1.7 percentage points (50%) more likely to be placed into foster care following an investigation than white children conditional on subsequent maltreatment potential. This disparity is entirely driven by white investigators and by cases where maltreatment potential is present, in which Black children are twice as likely to be placed as white children (12% vs. 6%). These results suggest white children may be harmed by "under-placement" in high-risk situations via the leniency that white investigators afford to white parents. Leveraging the additional quasi-random assignment of hotline call screeners, we find that both screeners and investigators are responsible for unwarranted disparities in placement, with investigators amplifying the disparity for cases with subsequent maltreatment potential and mitigating it for lower-risk cases. This finding highlights the importance of "systems-based" analyses of inequity in high-stakes decisions, where discrimination can compound across multiple decision-makers.

*We thank Desmond Ang, Peter Arcidiacono, David Arnold, Bocar Ba, Anthony Bald, Pat Bayer, Peter Blair, Aislinn Bohren, Chris Campos, Will Dobbie, Matt Gentzkow, Ezra Goldstein, Felipe Goncalves, Marie-Pascale Grimon, Max Gross, Alex Imas, Larry Katz, Chris Mills, Aurelie Ouss, Ashesh Rambachan, Katherine Rittenhouse, Evan Rose, Soum Shukla, Seth Zimmerman, and seminar participants at the Becker Friedman Institute; the Cowles Labor and Public Economics Conference; Duke University; the Federal Reserve Bank of New York; the Paris School of Economics; the Race, Racism, and Structural Inequality Conference; the São Paulo School of Economics; the Society of Labor Economists; and the Western Economic Association for helpful comments. The findings do not necessarily represent the opinion of the Federal Reserve Bank of New York or the Federal Reserve System. All errors are our own.

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

‡MIT and NBER. Email: jjdoyle@mit.edu.

§Federal Reserve Bank of New York. Email: natalia@nataliaemanuel.com.

¶Brown University and NBER. Email: peter_hull@brown.edu.

‖University of Michigan. Email: joryan@umich.edu.

I Introduction

Child protective services (CPS) aim to prevent child maltreatment by investigating reported cases of abuse or neglect and placing children in foster care when deemed necessary to ensure their safety. In the U.S., CPS involvement is remarkably common: 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 (Kim et al., 2017), and Black children are twice as likely to spend time in foster care (10%, compared to 5% of white children; Wildeman and Emanuel, 2014).

There is enormous interest in these racial disparities and the extent to which they might reflect discrimination in the decisions of CPS investigators. This question has been studied for at least fifty years, garnering significant popular and media attention.[1] 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.[2]

Interest in these disparities reflects the fact that CPS actions can have a tremendous impact on the lives of children and parents, and involve a difficult trade-off. Leaving children in high-risk situations may lead to subsequent maltreatment, which is associated with impaired physical and mental health (Lansford et al., 2006), decreased educational attainment and future earnings (Currie and Spatz Widom, 2010), and increased criminal activity (Currie and Tekin, 2012; Doyle and Aizer, 2018).[3] At the same time, 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 (Doyle, 2007, 2008; Bald et al., 2022a,b; Baron and Gross, 2022; Grimon, 2023; Gross and Baron, 2022; Helénsdotter, 2022). Discrimination in placement decisions thus stands to exacerbate inequities in many long-term outcomes.

Attributing disparities to racial discrimination is fundamentally challenging, however. The central mandate of CPS investigators is to place a child in foster care when the potential for future maltreatment in the home is high. Unconditional disparities in foster care placement

---

[1]Academic work includes: Billingsley and Giovannoni (1972); Chibnall et al. (2003); Roberts (2009); Drake et al. (2011); Font et al. (2012); Pryce et al. (2019); Dettlaff and Boyd (2020); Reddy et al. (2022). For a popular account, see Newman, A (2022), *Is N.Y.'s Child Welfare System Racist? Some of Its Own Workers Say Yes*, The New York Times. Accessed at: https://www.nytimes.com/ (11/23/2022).

[2]See Kelly, J (2022), *UN Committee Suggests the US Change or Repeal Major Child Welfare Policies*, The Imprint. Accessed at: https://imprintnews.org/ (5/6/2023); White, S and Persson, S (2022), *Racial Discrimination in Child Welfare Is a Human Rights Violation–Let's Talk About It That Way*, The American Bar Association. Accessed at: https://www.americanbar.org/ (5/6/2023).

[3]Peterson et al. (2018) estimate that the lifetime cost of new child abuse and neglect cases in the U.S. each year is as high as 10.9% of GDP.

rates may therefore suffer from omitted variables bias (OVB) by not adjusting for differences in future maltreatment potential. This bias is difficult to address since maltreatment potential in the home is only selectively observed among children who are not placed in foster care. Studies that condition disparities on other traits, such as poverty, may further suffer from included variables bias (IVB) if discrimination operates indirectly through such traits.[4]

A further empirical challenge arises from the multi-phase nature of CPS systems, where foster care outcomes are technically determined in two phases. Hotline screeners first choose to "screen-in" calls and launch investigations before investigators decide whether to place screened-in children in foster care. A large theoretical literature notes the importance of accounting for the ways in which discrimination can be sustained, amplified, or mitigated across decision-makers in such multi-phase systems (Pincus, 1996; Powell, 2008). But bringing to data such a "systems-based" analysis of discrimination in high-stakes decisions is usually difficult, in part because of the potential for OVB and IVB at each phase of the system.

This paper—the first quasi-experimental analysis of racial discrimination in foster care decisions—develops and applies new empirical tools for overcoming these persistent challenges. We measure discrimination as unwarranted disparities (UDs): racial disparities in foster care placement rates conditional on a child's potential for future maltreatment in the home. This measure captures 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 (Bohren et al., 2022).[5]

In the first part of the study, we estimate UDs in the placement decisions of Michigan CPS investigators using nearly 220,000 screened-in calls from 2008 to 2016. Our initial focus on the post-screening phase follows much of the academic literature and recent policy debates, and is natural since investigators make the ultimate placement recommendations.

An important question for the analysis is how to measure subsequent child maltreatment. Our primary UD measure compares the foster care placement rates of white and Black children with the same potential for a subsequent child maltreatment investigation in the home within six months—a common proxy for subsequent maltreatment in the child welfare literature (Antle et al., 2009; Putnam-Hornstein et al., 2021). This proxy is imperfect insofar as it may miss unreported maltreatment, it may be influenced by discriminatory practices among reporters, and it is a binary measure that does not distinguish between different levels of maltreatment

[4]Examples of observational disparity analyses in this setting include Paxson and Waldfogel (1999, 2002); Putnam-Hornstein et al. (2013); Shaw et al. (2008); Wulczyn et al. (2013); Billingsley and Giovannoni (1972); Chibnall et al. (2003); Font et al. (2012); Courtney et al. (1996); Drake et al. (2011).

[5]As we discuss below, our UD measure aligns with the legal theory of disparate impact as well as notions of algorithmic fairness in the computer science literature (Arnold et al., 2021, 2022).

severity. We take these issues seriously in the analysis, showing that our results are similar when using other proxies for subsequent maltreatment that reflect varying time frames and levels of severity (including subsequent foster care placement) and across reporter types.[6] For ease of exposition, we refer to the baseline six-month re-investigation proxy as "subsequent maltreatment" throughout the paper.

The key identification challenge in measuring UDs is the selective observability of a child's potential for subsequent maltreatment in the home. We address this challenge by leveraging variation across quasi-randomly assigned investigators who differ in their tendency to place children into foster care. To build intuition for this approach, consider a randomly assigned investigator whose placement rate is near zero. By virtue of random assignment, the subsequent maltreatment rates observed among Black and white children left at home by this investigator are close to the average rates among all Black and white children. These race-specific maltreatment rates capture the correlation between future maltreatment potential and race, and they can be used to correct for OVB in unconditional placement disparities.

Absent such an investigator, we estimate the key race-specific maltreatment rates by extrapolating from the observed maltreatment rates among quasi-randomly assigned investigators with low placement tendencies. This identification strategy builds on Arnold et al. (2022), who use quasi-random bail judge assignment to measure discrimination in pretrial release decisions, as well as a broader literature on "identification at infinity" in sample selection models. The CPS setting is particularly well-suited to this strategy: placement rates are low, making extrapolation more credible while also allowing us to construct highly informative bounds on the average maltreatment rates and UDs without any extrapolation.

We find significant evidence of unwarranted disparity in investigators' placement decisions: Black children are 1.7 percentage points (50%) more likely to be placed in foster care than white children with identical potential for subsequent maltreatment in the home. Correcting for the selective observability of maltreatment potential is substantively important: UD estimates are nearly 90% *larger* than placement disparities from an observational analysis that conditions on child and investigation traits.[7] Estimates from a simple model of investigator decision-making suggest that UDs arise from racial bias and not accurate

[6] We prefer subsequent investigation as our baseline proxy because it is not directly influenced by the initial investigator. Since re-investigations within short time frames tend to be re-assigned to the initial investigator, subsequent placement recommendation is potentially endogenous. Importantly, our analysis is not premised on the view that differences in re-investigation rates are unaffected by discrimination in other phases of the CPS system or society as a whole. We return to this point extensively below.

[7] The fact that observational disparities understate the level of UD reflects that OVB attenuates disparities in our context: Black children in our sample are around 2 percentage points (13%) less likely to see future maltreatment when left at home than white children.

statistical discrimination, differences in investigator skill, or non-race factors.

We then document two striking forms of heterogeneity in the UDs. First, we find that the placement disparity is concentrated among children with subsequent maltreatment potential in the home, with Black children placed in foster care at twice the rate of white children in this subpopulation (12% versus 6%). In contrast, the placement disparity is small and statistically insignificant in the subpopulation of children without maltreatment potential. Second, investigators display a racial concordance effect in cases with maltreatment potential, being significantly less likely to place children of their own race than children of another race. Since the vast majority of investigators in Michigan are white (86%), this concordance effect yields higher conditional placement rates for Black children.

The finding that unwarranted disparity is concentrated among high-risk cases implies that a higher placement rate may actually offer protection to Black children relative to white children. Indeed, prior research 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 systems, which often focus on the possibility that Black children are "over-placed" in foster care.[8] While it is true that Black children in Michigan are disproportionately placed in foster care relative to white children in homes with future maltreatment potential, white children may be harmed by "under-placement" in these high-risk situations. The finding of a racial concordance effect 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. A simple back-of-the-envelope calculation shows that lowering the placement rate of Black children to equalize placement rates across race would lead to a 7% increase in the number of Black children who are subsequently maltreated in the home.

These findings are not unique to Michigan CPS. Nationwide (though more limited) data from the National Child Abuse and Neglect Data System (NCANDS, 2023) allow us to construct non-parametric bounds on foster care placement UD for almost all states in the U.S. This supplementary analysis replicates our main findings for Michigan while showing qualitatively similar patterns in most other states. Nationwide, UD in low-risk cases tends to be small while UD in high-risk cases is often as large or larger than in Michigan.

[8] See, for example, the Minnesota African American Family Preservation Act, as well as policy recommendations in the New York State Bar Association's "Resolution addressing systemic racism in the child welfare system of the State of New York."

In the final part of the study, we extend our quasi-experimental approach to conduct a systems-based analysis of unwarranted disparity in foster care placement. Our primary analysis, by design, isolates UDs in investigator decisions holding fixed initial screening decisions that determine whether or not to launch an investigation. Exploiting a novel source of variation—the quasi-random assignment of CPS hotline screeners—and imposing some additional structure allows us to additionally trace UDs across both phases and study how screeners and investigators jointly contribute to discrimination in foster care placement.

The systems-based analysis shows that our primary findings hold in the population of all calls, including those that were screened out of investigation. In total, hotline calls involving Black children are more likely to result in foster care placement than calls involving white children with identical maltreatment potential, with the UD entirely driven by calls with future maltreatment potential. Notably, this pattern holds despite UDs in the initial screening phase for cases *both* with and without maltreatment potential. We reconcile these results by decomposing the total UD into components due to screeners and investigators, building on the framework of Bohren et al. (2022). Because investigators are skilled at inferring risk, they place a small share of children in cases without maltreatment potential—thereby mitigating initial screener UDs in low-risk cases. In contrast, investigators amplify initial UDs in high-risk cases, such that both screeners and investigators contribute to the potential "under-placement" of white children. Our decomposition shows that screeners account for up to 14% of this effect, showing that eliminating UD may require intervention at both phases of the system.

This study contributes to several related literatures. First, we add to the literature examining CPS systems by conducting the first quasi-experimental study of racial discrimination in foster care decisions. We also leverage 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. (2022b) for a review), much less is known about the broader effects of CPS.[9]

Second, we add to a recent methodological literature that explores how the quasi-random assignment of decision-makers can be used to estimate different forms of bias and discrimination in high-stakes decisions, such as pretrial release (e.g., Arnold et al. (2018), Hull (2021), Arnold et al. (2022), Rambachan (2022), and Canay et al. (2022)), traffic stops (e.g., Goncalves and Mello (2021) and Feigenberg and Miller (2022)) and lending (e.g., Dobbie et al. (2021)).[10] Our analysis benefits from the foster care setting featuring many decision-makers

[9]A related literature explores the impact of algorithmic decision tools within CPS: see, e.g., Chouldechova et al. (2018), Brown et al. (2019), Grimon and Mills (2022) and Rittenhouse et al. (2022).

[10]See also Chan et al. (2022) and Angelova et al. (2023) for related quasi-experimental approaches to evaluating decision-maker skill.

with very low treatment rates, which allows for both precise non-parametric inferences on overall UDs and the statistical power to distinguish between disparities among high- or 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.

Third, this is the first quasi-experimental analysis of how discrimination perpetuates and compounds across multiple decision-makers in a system. A large theoretical literature emphasizes this possibility and its implications for policy (see Bohren et al. (2022) for a review). These insights, while potentially valuable in many areas within economics (e.g., studies of discrimination in the criminal justice system or the labor market), are often hard to bring to data because of the non-random decision-making at either or both phases of a system (e.g., police officer and prosecutor decisions). We provide a practical framework for how to conduct such an analysis when the multiple decision-makers are quasi-randomly assigned.[11]

The remainder of the paper is organized as follows. Section II describes the CPS setting. Section III describes our analysis sample and presents motivating results. Section IV develops our UD measures and empirical approach. Section V presents our main findings on investigators' decisions. Section VI introduces a novel screening instrument to study how UDs propagate through CPS. Section VII concludes.

II Setting

The CPS system aims to protect children from maltreatment in their home environment; Figure 1 summarizes the process in Michigan and most states. The process 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).[12] Anyone can make a report to the hotline, though the most common reporters are educators and law enforcement personnel (Benson et al., 2022).

Calls to the CPS hotline are answered by screeners who assess whether the allegation of maltreatment conforms with state law and guidance from Michigan's Department of Health

[11] In an innovative paper, 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.

[12] Most CPS investigations involve at least one allegation of neglect. Definitions of neglect tend to be broad and vary across states, leading to concerns that investigations of neglect may be driven by poverty-based hardship alone (e.g., Raz and Sankaran (2019)). However, there is mounting evidence that this is not the case. The vast majority of states, including Michigan, do not consider *involuntary neglect* (i.e. deprivation due to financial inability alone) as neglect for the purposes of a CPS investigation (Rebbe, 2018). Examples of neglect instead include exposure to family violence and unsafe supervision due to parental substance abuse. Neglect allegations furthermore tend to be co-reported along physical and emotional abuse (Palmer et al., 2022). Indeed, the association between neglect investigations and later-in-life outcomes tends to be at least as negative as that of physical abuse investigations (Font and Maguire-Jack, 2020).

and Human Services (MDHHS). Screeners sit centrally in two offices (one in Grand Rapids, MI and one in Detroit, MI), both with the same CPS hotline number. Conditional on exact day by shift, incoming calls are randomly assigned to screeners with no exceptions. Incoming calls enter a queue, and the hotline system routes the call to the available screener who has been waiting the longest since her last call. Calls typically last about 15 minutes. Screeners have substantial discretion in whether to screen-in a call, though they follow general guidelines. Screeners are instructed to screen-in calls to minimize the likelihood of subsequent maltreatment if the call is screened-out. 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 and screeners do not systematically learn the eventual outcome of a given investigation.

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 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 notable exceptions to 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 isolate the quasi-random variation from rotations with the child's ZIP code by investigation year ("rotation") fixed effects, excluding cases of sexual abuse and those involving children who had been the subject of an investigation in the year before the report. Once a case is assigned, the investigator has 24 hours to begin an investigation, 72 hours to establish face-to-face contact with the alleged child victim, and 30 days to complete the investigation.

Investigators make two primary decisions. First, an investigator must decide whether there is enough evidence to substantiate the allegation. This determination is based on interviews with the child maltreatment reporter and family members, as well as police and medical reports. In Michigan, around three quarters of investigations are unsubstantiated over our study period. 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. Under CPS guidelines, the only justification for placement is a potential for subsequent maltreatment in the home: investigators are instructed to place the child in foster care if the child is in imminent danger of maltreatment in the home, but to otherwise keep the child with their family.[13] Investigators have immense discretion over foster care placement. While there

[13]For example, Michigan's Department of Health and Human Services' *Children's Protective Services Policy*

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 responses to match their priors (Gillingham and Humphreys, 2010; Bosk, 2015).

If the investigator determines that the potential for maltreatment is high, she requests 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. Regardless of the placement decision, investigators can recommend preventative referrals to support the family. These referrals range from food banks to substance abuse or parenting classes, though parents are not typically compelled to use them. Past work has found minimal impact of these prevention-referral decisions (Baron and Gross, 2022; Gross and Baron, 2022).

Around 3% of screened-in cases 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. Children spend around 17 months in foster care on average, after which most are reunified with their parents.

III Data

III.A Data Sources and Analysis Sample

Our primary analysis uses data from the Michigan Department of Health and Human Services, consisting of the universe of maltreatment investigations (screened-in calls) in Michigan between January 2008 and June 2017. The data include details of each investigation, such as the allegation report date, allegation types as coded by the screener, the child's ZIP code, and indicators for substantiation and foster care placement. The files also include information on the child's age, sex, race, and ethnicity, as well as the name of the investigator.[14] We use the investigators' names to predict their race, ethnicity and gender.[15]

Manuals reads: "placement of children out of their homes should occur only if their well-being cannot be safeguarded with their families" (p.3). It 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).

[14] A more recent MDHHS dataset (from January 2017—December 2019) contains information on both the hotline screeners assigned to each call and the outcomes of screened-out calls. In this subset we can thus follow an initial hotline call through both the screening and foster care placement decisions, though we do not observe the characteristics of CPS investigators. As discussed in Section VI, we use this dataset to study how UDs persist and compound across the the two phases of the CPS system.

[15] Specifically, we impute investigator race/ethnicity via the *ethnicolr* package in Python, which uses U.S. census data, the Florida voting registration data, and Wikipedia data collected by Skiena and Ward (2014) to predict race/ethnicity based on first and last name. We predict investigator gender using the R package *predictrace*, which uses data from the U.S. Social Security Administration (SSA) to determine the probability

We construct our analysis sample from these data as follows. We begin with the 573,349 unique investigations of children in Michigan between January 2008 and June 2017 that did not involve either sexual abuse or repeat reports within one year since these cases are not quasi-randomly assigned. We then drop a relatively small number of cases with missing child ZIP code information ($N = 9,970$), since quasi-random assignment of investigators is conditional on geography. We further drop cases involving children not classified as white or Black ($N = 91,085$), cases assigned to investigators who handled fewer than 200 investigations of white and Black children ($N = 234,430$), and cases assigned to investigators who handled either investigations of only white or only Black children ($N = 7,237$). The latter two restrictions ensure a relatively large sample for studying investigator heterogeneity. We then drop children for whom we cannot observe child welfare outcomes for at least six months after the focal investigation ($N = 11,623$), as this will be our primary outcome of interest. Finally, we drop observations in rotations with only one investigation ($N = 1,300$).

The resulting analysis sample consists of 217,704 investigations of 181,928 unique children. There are 699 unique investigators, each assigned to 312 cases and 6,644 rotations on average. Panel A of Table 1 summarizes this sample. Overall, 70% of children in our sample are white and 48% are female. The average child is nearly seven years old. A large share of children (46%) have a previous investigation, with around one previous investigation on average.

Panel B of Table 1 shows that 43% of investigations include a physical or substance abuse allegation with nearly 53% including an improper supervision allegation. These rates are higher for white children, while the rate of investigations including a physical neglect allegation (around 45%, on average) is higher for Black children. Around 91% of alleged perpetrators include a parent or stepparent.[16]

Panel C of Table 1 shows that 3.4% of investigated children in our sample are placed in foster care. There are clear racial disparities in this rate: 4.3% of Black children are placed in foster care, relative to 3.1% of white children. That is, on average, investigated Black children are 1.2 percentage points (35%) more likely to be placed in foster care.

Finally, Panel D of Table 1 summarizes future maltreatment outcomes among children not placed in foster care. Two months after the start of an investigation, around 7% of non-placed children are re-investigated for child maltreatment. This rate rises steadily with time, with around 17% of non-placed children re-investigated for child maltreatment after six months. We take this six-month outcome as our primary measure of subsequent maltreatment, though we explore robustness to the other shorter-horizon outcomes. Our focus on these short-run

that a given first name is of a particular gender.

[16] These categories are not mutually exclusive since there can be multiple allegations and perpetrators.

outcomes is consistent with investigator manuals instructing workers to place a child in foster care if they believe the child is in "imminent risk."[17]

Our primary maltreatment measure considers whether the child was re-investigated, which entails both a report to CPS and a decision by a hotline screener to begin an investigation. This is a common measure of subsequent maltreatment in the child welfare literature (Antle et al., 2009; Putnam-Hornstein and Needell, 2011; Casanueva et al., 2015; Putnam-Hornstein et al., 2015, 2021). As mentioned above, only 60% of all hotline calls are assigned for investigation. While a subsequent investigation is not the only potential proxy for subsequent maltreatment, it is our preferred measure because it is not impacted by decisions of the initial investigator. Recall that subsequent investigations within a few months are often re-assigned to the initial investigator, while neither the decision to report nor to screen-in a case—both of which are necessary for a re-investigation—involve the investigator. Nevertheless, we show below that our results are robust to considering alternative maltreatment outcomes, such as whether the child was a substantiated victim of child maltreatment within 6 months, whether the child was placed in foster care within 6 months, or whether the child had an investigation for physical abuse (as opposed to neglect). Our estimates can be seen to isolate policy-relevant measures of unwarranted disparity, with the outcome representing a substantial intervention (including a 30-day investigation of the family). For ease of exposition, we refer to our primary outcome of re-investigation within six months as "subsequent maltreatment."

Importantly, our analysis is not premised on the view that differences in re-investigation rates are unaffected by discrimination at other points of the system or in society more broadly. Indeed, we find evidence of discrimination in hotline screening decisions below (Section VI). Differences in re-investigation risk could moreover be driven by the over-reporting of Black children to CPS. For example, prior research suggests that Black children may be disproportionately likely to be reported by medical personnel conditional on case severity (Lane et al., 2002).[18] Both of these scenarios 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 UDs that may be reliably targeted by policy, holding fixed other forms of discrimination that may be harder to quantify or address through reforms to CPS.

Panel D of Table 1 shows that, when left at home, white children are more likely than Black

[17]We omit the first month after the start of an investigation since we do not observe a disposition date in the data and, as mentioned above, investigators have 30 days to complete the focal investigation.

[18]As we discuss below, however, our UD estimates are similar across reporter types (e.g., mandated reporters such as educational, medical, and law enforcement personnel versus non-mandated reporters such as neighbors or other family members), suggesting that our results are not driven by the biases of particular reporter types.

children to experience subsequent maltreatment. Within six months of the focal investigation, 17% of white children see subsequent maltreatment, compared to 15% of Black children. Taken at face value, this outcome disparity may suggest white children have a higher level of "need" for foster care placement—which in turn suggests the placement rate disparity in Panel C *understates* the conditional disparity among children with the same level of need. Importantly, however, the outcomes in Panel D are only measured among children endogenously left at home. Since the average stay in foster care in our setting is 17 months long, and very few children return home within six months, we do not observe subsequent maltreatment in the home within six months if a child is temporarily placed in foster care. We are thus unable to directly adjust for potential maltreatment outcomes when assessing placement disparities.

III.B Descriptive Disparity Analysis

As an initial analysis of the overall foster care placement disparity, we estimate descriptive regressions of placement decisions on an indicator for the child's race controlling for a variety of child and investigation characteristics. Specifically, we estimate ordinary least squares (OLS) regressions of the form:

$$D_i = \alpha + \beta B_i + X_i'\gamma + \epsilon_i \tag{1}$$

where D_i is an indicator equal to one if child i is placed in foster care following a maltreatment investigation, B_i is an indicator equal to one if the child is Black, and X_i is a vector of controls. Estimates of β thus capture placement disparities among observably similar cases.

Column 1 of Table 2 presents estimates from a simple bivariate regression of D_i on B_i without any controls. The 1.2 percentage point disparity in this specification corresponds to the gap previously discussed in Panel C of Table 1. Column 2 of Table 2 shows this disparity is unchanged with rotation (ZIP code by year) fixed effects. Column 3 further adds controls for the child and focal investigation characteristics in Panels A and B of Table 1. The disparity in this specification shrinks slightly to 0.9 percentage points (27% of the mean placement rate).

A significant disparity in placement rates thus remains among observably similar children and cases. At 27% of the overall placement rate, this 0.9 percentage point disparity is meaningful. But the implications of this controlled disparity for UDs are at this point unclear: we cannot adjust these descriptive regressions for subsequent maltreatment potential, the evaluation of which is the sole objective of child welfare investigators. Consequently, the overall disparity in Column 1 of Table 2 may suffer from OVB and either over- or under-state the true level of UD across child welfare investigators in Michigan. The controlled disparities in Columns 2 and 3 may furthermore suffer from included variables bias (IVB) if the controls include mediators of UD. We next discuss the empirical strategy we will use to overcome both challenges.

IV Methods

IV.A Unwarranted Disparity Measure

We formalize our discrimination measure and estimation approach by considering a population of screened-in cases (indexed by i), each assigned to one of J investigators (indexed by j). Each case involves either a Black or white child, indicated by $R_i \in \{b, w\}$. Each child has a potential for future maltreatment $Y_i^* \in \{0, 1\}$, with $Y_i^* = 1$ indicating future maltreatment when the child is not placed into foster care. Without directly observing Y_i^*, investigators make a decision $D_{ij} \in \{0, 1\}$ to place child i in foster care, with $D_{ij} = 1$ indicating placement when child i is assigned to investigator j. We study discrimination in these potential placement decisions, leaving unspecified for now the assignment process of investigators.

We measure discrimination as UDs: racial disparities in foster care placement rates conditional on a child's true potential for subsequent maltreatment in the home Y_i^*. This measure builds on Arnold et al. (2021, 2022), who study discrimination in the bail context by racial disparities in pretrial release rates conditional on a defendant's true potential for pretrial misconduct. As in that context, our UD discrimination measure is natural given clear decision-maker objectives: under CPS investigator guidelines, the only justification for foster care placement is a potential for subsequent maltreatment in the home. Arnold et al. (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.[19]

Importantly, unwarranted disparity can arise from both "direct" discrimination on the basis of race itself and from "indirect" discrimination through non-race characteristics (such as poverty levels). The former source, which has historically been the focus in economics, includes discrimination from biased preferences and beliefs (e.g. Becker (1957); Bordalo et al. (2016); Bohren et al. (2020)) and statistical discrimination (e.g. Phelps (1972); Arrow (1973); Aigner and Cain (1977)). The latter source, which historically has been more often studied outside of economics, can arise when non-race characteristics embed discrimination from the past or other domains (Bohren et al., 2022). We explore the role of each of these potential drivers using a simple model of investigator decision-making below.

[19]Disparate impact is one of two main legal doctrines of discrimination in U.S. case law, which 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 et al. (2022) for more background and discussion of relevant case law.

To build up to our UD measure, we first define the placement rate disparity for each investigator j among Black and white children without future maltreatment potential:

$$\Delta_{j0} = E[D_{ij} \mid R_i = b, Y_i^* = 0] - E[D_{ij} \mid R_i = w, Y_i^* = 0], \tag{2}$$

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

$$\Delta_{j1} = E[D_{ij} \mid R_i = b, Y_i^* = 1] - E[D_{ij} \mid R_i = w, Y_i^* = 1]. \tag{3}$$

We measure the overall UD by averaging these two conditional disparities:

$$\Delta_j = \Delta_{j0}(1 - \bar{\mu}) + \Delta_{j1}\bar{\mu}, \tag{4}$$

with weights given by the average future maltreatment risk in the population, $\bar{\mu} = E[Y_i^*]$. The Δ_j average thus captures the expected level of UD when investigator j encounters a representative pool of children, with unknown future maltreatment potential. We measure the system-wide level of UD by the case-weighted average of Δ_j across investigators. By holding the potential population of assigned children fixed, variation in Δ_j around this average is meaningfully attributed to investigator heterogeneity.[20]

While our initial focus is on Δ_j, in additional analyses we separately study the two individual components, Δ_{j0} and Δ_{j1}, as this can lead to a more nuanced analysis of unwarranted disparity in foster care placement. A finding of $\Delta_{j0} > 0$, for example, 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 in their homes. However, a finding of $\Delta_{j1} > 0$ would suggest that, while Black children are placed at higher rates than white children, the higher placement rate may actually be protective to Black children, since they experience subsequent maltreatment potential in their homes and previous research in our context shows that the causal effects of foster care are positive 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 $\Delta_{j1} > 0$

[20]Our measures of discrimination, which condition on the potential for child maltreatment in the home, follow naturally from investigators' mandate to focus on these outcomes. Another possible measure would condition on a child's potential reduction in maltreatment if placed into foster care. These measures are likely to be very similar, since maltreatment is very rare while in foster care: 0.88% of children entering foster care in Michigan in 2017 were maltreated during their stay (USDHHS, 2017; Biehal, 2014). Investigators' placement tendencies are also uncorrelated with children's experiences in foster care, such as length of stay, number of different placements, and placement type (Baron and Gross, 2022), suggesting that investigators' decisions are not influenced by their expectations of children's foster care experiences. Moreover, Baron and Gross (2022) show that the causal effects of foster care are similar for Black and white children at the margin of placement.

could be interpreted as "under-placement" of white children relative to Black children.

The fundamental challenge in estimating UDs is the selective observability of maltreatment potential. Among children who are not placed in foster care, Y_i^* is directly observed by 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 (2)-(4).[21]

We address this selection challenge, following Arnold et al. (2022), by first re-writing the components of equations (2)-(4) in terms of a set of directly estimable moments and two unknown parameters: the average future maltreatment risk in the population of Black and white children, $\mu_b = Pr(Y_i^* = 1 \mid R_i = b)$ and $\mu_w = Pr(Y_i^* = 1 \mid R_i = w)$. Specifically:

$$E[D_{ij} \mid R_i = r, Y_i^* = 1] = 1 - \frac{E[(1 - D_{ij})Y_i^* \mid R_i = r]}{E[Y_i^* \mid R_i = r]} = 1 - \frac{(1 - \phi_{jr})\psi_{jr}}{\mu_r} \tag{5}$$

and

$$E[D_{ij} \mid R_i = r, Y_i^* = 0] = 1 - \frac{E[(1 - D_{ij})(1 - Y_i^*) \mid R_i = r]}{E[1 - Y_i^* \mid R_i = r]} = 1 - \frac{(1 - \phi_{jr})(1 - \psi_{jr})}{1 - \mu_r} \tag{6}$$

for each j and r, where $\phi_{jr} = Pr(D_{ij} = 1 \mid R_i = r)$ is investigator j's potential foster care placement rate for children of race r and $\psi_{jr} = Pr(Y_i^* = 1 \mid D_{ij} = 0, R_i = r)$ is the subsequent maltreatment rate of children of race r who would be left at home by investigator j.

Since placement is directly observed, and since future maltreatment outcomes directly reveal future maltreatment potential Y_i^* among children not placed in foster care, the ϕ_{jr} and ψ_{jr} moments in equations (5) and (6) are not affected by the selective observability of Y_i^*. In fact, when investigators are randomly assigned, these moments are identified simply by the race-specific placement rates and at-home maltreatment rates among cases assigned to each investigator.[22] Thus, in quasi-experimental data the challenge of identifying Δ_j reduces to the challenge of identifying the two remaining parameters in equations (5) and (6): μ_b and μ_w.

IV.B Identification Strategies

We consider two strategies, building on Arnold et al. (2022), for estimating the two key mean risk parameters, μ_b and μ_w. Each strategy yields corresponding estimates of the UDs.

[21] Since the average stay in foster care is 17 months and very few children return home within six months, it is usually impossible to see if a child placed in foster care is maltreated at home within six months.

[22] Formally, $\phi_{jr} = E[D_{ij} \mid R_i = r] = E[D_i \mid Z_{ij} = 1, R_i = r]$ and $\psi_{jr} = E[Y_i^* \mid D_{ij} = 0, R_i = r] = E[Y_i \mid D_i = 0, Z_{ij} = 1, R_i = r]$ where Z_{ij} indicates the assignment of case i to investigator j, D_i indicates realized foster care placement for case i, and Y_i indicates realized future maltreatment for case i. Here we assume simple random assignment, such that Z_{ij} is independent of (D_{ij}, Y_i^*, R_i). We discuss how we handle conditional random assignment below.

Bounding with Observed Subsequent Maltreatment. Our first strategy forms non-parametric bounds on each μ_r using the directly estimable moments ϕ_{jr} and ψ_{jr}. For each investigator j, a lower bound is given by:

$$\mu_{jr}^L \equiv (1-\phi_{jr})\psi_{jr} = E[(1-D_{ij})Y_i^* \mid R_i = r] \leq E[Y_i^* \mid R_i = r] = \mu_r. \tag{7}$$

Intuitively, this lower bound is derived by assuming none of the children who would be placed in foster care by investigator j (and thus for whom we cannot observe future maltreatment potential Y_i^*) would have had future maltreatment. Similarly, an upper bound is given by:

$$\begin{aligned}\mu_{jr}^U \equiv 1-(1-\phi_{jr})(1-\psi_{jr}) &= 1-E[(1-D_{ij})(1-Y_i^*) \mid R_i = r] \\ &\geq 1-E[1-Y_i^* \mid R_i = r] = \mu_r,\end{aligned} \tag{8}$$

Intuitively, this upper bound is derived by assuming all of the children who would be placed in foster care by investigator j would have had future maltreatment.

The $[\mu_{jr}^L, \mu_{jr}^U]$ bounds can be estimated from the placement rates and future maltreatment rates of each quasi-randomly assigned investigator j, as detailed below, or from an average of these rates across investigators. Tighter bounds are obtained by focusing on investigators with lower placement rates, since:

$$\mu_{jr}^U - \mu_{jr}^L = 1-(1-\phi_{jr})(1-\psi_{jr}) - (1-\phi_{jr})\psi_{jr} = \phi_{jr}. \tag{9}$$

Because placement in foster care is relatively rare, $\phi_{jr} = Pr(D_{ij} = 1 \mid R_i = r)$ is relatively small for many investigators such that our bounds are likely to be informative. Intuitively, the mean future maltreatment rate in the full population cannot be too far away from the observed maltreatment rates among children left at home when placement is relatively rare.

Extrapolating Maltreatment Variation Across Investigators. Our second strategy for estimating the two mean risk parameters extrapolates observed variation in subsequent maltreatment rates across as-good-as-randomly assigned investigators. To build intuition for this approach, suppose investigators are completely randomly assigned and there is an investigator j^* who places virtually no children of either race in foster care such that $\phi_{j^*r} = Pr(D_{ij^*} = 1 \mid R_i = r) \approx 0$. For each race, this investigator's future maltreatment rate among children left at home would be close to the mean maltreatment risk of that race:

$$\psi_{j^*r} = E[Y_i^* \mid D_{ij^*} = 0, R_i = r] \approx E[Y^* \mid R_i = r] = \mu_r \tag{10}$$

Moreover, by random assignment, ϕ_{j^*r} and ψ_{j^*r} are identified by the realized race-specific placement rates and future maltreatment rates among children assigned to the investigator.

Absent such an investigator, the key mean risk parameters can be estimated by extrapolating maltreatment rate variation across quasi-randomly assigned investigators. This approach is conceptually similar to how average potential outcomes at a treatment cutoff can be extrapolated from nearby observations in a regression discontinuity design. Here, potential maltreatment risk is extrapolated from quasi-randomly assigned investigators with low placement rates to the hypothetical investigator whose placement rate is zero. Mean risk estimates may, for example, come from the vertical intercept at zero of linear, quadratic, or local linear regressions of estimated at-home maltreatment rates $E[Y_i^* \mid D_{ij} = 0, R_i = r]$ on estimated placement rates $Pr(D_{ij} = 1 \mid R_i = r)$ across investigators j within each race r.[23]

We apply both strategies accounting for the fact that investigators in Michigan are only as-good-as-randomly assigned conditional on rotation fixed effects. We first use linear regression to estimate ϕ_{jr} and ψ_{jr} adjusting for these strata. Specifically, we estimate:

$$D_i = \sum_j \phi_{jw}(1 - B_i)Z_{ij} + \sum_j \phi_{jb} B_j Z_{ij} + X_i'\gamma + \varepsilon_i \tag{11}$$

among all screened-in cases, and:

$$Y_i = \sum_j \psi_{jw}(1 - B_i)Z_{ij} + \sum_j \psi_{jb} B_j Z_{ij} + X_i'\lambda + \nu_i \tag{12}$$

among screened-in cases where the child is not placed in foster care ($D_i = 0$). Here $D_i = \sum_j Z_{ij} D_{ij}$ indicates observed foster care placement, $Y_i = D_i Y_i^*$ indicates observed future maltreatment, and $B_i = \mathbf{1}[R_i = b]$ indicates a case involving a Black child. The $Z_{ij} \in \{0, 1\}$ dummies indicate assignment of case i to investigator j, and X_i is a vector of rotation fixed effects within which investigators are as-good-as-randomly assigned. We estimate both regressions with X_i de-meaned, such that the ϕ_{jw} and ϕ_{jb} coefficients capture strata-adjusted placement rates of investigator j among white and Black children, respectively. Similarly, the ψ_{jw} and ψ_{jb} coefficients capture strata-adjusted maltreatment rates among the white and Black children left at home by investigator j.[24] We use these estimates to bound or estimate the key mean risk parameters μ_w and μ_b, following the above formulas. Finally, we combine these two sets of estimates to estimate UDs, following equations (5) and (6).

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

[24] Regression adjustment is appropriate when placement and re-investigation rates are linear in the rotation strata for each judge and race, with constant coefficients (Arnold et al., 2022). We show below that our results are nearly identical if we do not adjust for strata fixed effects, however. This is because variation in placement rates is largely driven by variation within, rather than across, the strata (Baron and Gross, 2022).

IV.C Identifying Assumptions

The first assumption underlying our two strategies is the quasi-random assignment of investigators in Michigan, who vary in their tendency to place children in foster care. To verify this assumption, we construct a measure of investigator placement tendencies as a leave-one-out rate among all cases in the sample that were quasi-randomly assigned to the investigator. This measure is similar to instruments previously constructed to estimate the causal effects of foster care placement among screened-in cases (Doyle, 2007, 2008; Bald et al., 2022a; Baron and Gross, 2022; Gross and Baron, 2022). Specifically, we first regress the foster care placement indicator D_i on the rotation fixed effects X_i. We then calculate the leave-one-out average residual from this regression for each investigator.

Figure A1 summarizes the leave-one-out placement tendency measure. The measure ranges from -0.03 to 0.04, with a standard deviation of 0.023. Within a rotation, moving from the investigator with the lowest tendency to place to the one with the highest tendency is thus estimated to increase the probability that the child is placed in foster care by 7 percentage points, or over 200% relative to the mean placement rate in our sample. The figure further shows a strong, positive relationship between investigators' placement tendencies and the child's likelihood of being placed in foster care.

Table A1 confirms this first-stage relationship by reporting estimates of an OLS regression of placement decisions on the tendency measure, separately by race. Columns 1 and 3 include only rotation fixed effects, while Columns 2 and 4 add baseline controls. A one percentage point increase in an investigator's placement tendency increases the probability that the child is placed in foster care by 0.536 percentage points for Black children, and 0.591 percentage points for white children. The F-statistic for placement tendency ranges from 133 to 257, indicating a strong relationship.

Table A2 further probes an implication of as-good-as-random assignment: that observable child and case characteristics are uncorrelated with the placement tendencies of the assigned investigator. Separately by race, each column reports point estimates from an OLS regression of an indicator equal to one if the child was placed (Columns 1 and 3) and the investigator's placement tendency (Columns 2 and 4) on all child and investigation characteristics and rotation fixed effects. As expected due to the rotational assignment of child welfare investigators, a rich set of characteristics are not jointly predictive of the instrument ($p = 0.201$ for Black children and $p = 0.183$ for white children) despite being very predictive of placement. Moreover, Table A3 shows that investigators of a given race/gender are not differentially likely to be assigned to same-race/same-gender cases.

The second assumption underlying our two strategies is an implicit exclusion restriction: that investigators' placement tendencies only impact subsequent maltreatment outcomes through the decision to place a child into foster care. This assumption is inherently untestable, and—even with rich administrative data—we could not rule out all possible channels through which investigators with higher or lower tendencies to place may influence children's outcomes. Nevertheless, we believe exclusion is reasonable in this context.

One possible concern for the exclusion restriction is that investigators could influence children's experiences while in foster care. CPS investigators, however, do not remain in contact with children after placement. Once a child is placed in foster care, their case is transferred to a different agency staff member who works in a separate Foster Care Department. These workers are often also assigned to cases according to a rotation, but this rotation is distinct from the one for the initial investigators as they include different types of workers.

Another concern for the exclusion restriction in our setting is investigator discretion over the decision to assign families to prevention services. Specifically, investigators could assign families to community-based or targeted services, which range from referrals to food pantries to substance abuse and parenting classes. If these preventative services impact the probability of subsequent abuse, and these decisions are systematically correlated with investigator placement tendencies, then these decisions would violate the exclusion restriction. However, Gross and Baron (2022) and Baron and Gross (2022) show that the effects of these other decisions on children's outcomes (including subsequent maltreatment) are small and statistically insignificant in Michigan.

Importantly, although our approach can be understood as leveraging quasi-random investigator assignment as an instrument 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 investigators have a common ranking of cases by their appropriateness for foster care placement.[25] Intuitively, the non-parametric bounds on μ_r do not come from comparisons across investigators, while the extrapolated estimates of μ_r are valid as long as the *average* relationship between subsequent maltreatment rates and placement rates across investigators can be reliably estimated (at least for low placement rates).[26] In

[25]See Mueller-Smith (2015); Norris (2019); Mogstad et al. (2021) and Frandsen et al. (2023) for critiques of conventional first-stage monotonicity in similar settings.

[26]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. Further assume $E[Y_i^*|\nu_{ij}, \kappa_j, \lambda_j] = \mu + \lambda_j(\nu_{ij} - \frac{1}{2})$. 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.

practice, the large number of foster care investigators with low placement rates in our setting are likely to both make the non-parametric bounds informative and the extrapolations reliable.

V Unwarranted Disparity in Investigator Decisions

V.A Main Estimates

Bounding with Observed Subsequent Maltreatment. We first compute a range of possible mean maltreatment risk parameters, given by the overall average rates of placement and future maltreatment in our analysis sample. Applying equations (7) and (8) to the rates in Table 1, we estimate Black mean risk bounds of $\mu_b \in [0.144, 0.187]$ and white mean risk bounds of $\mu_w \in [0.168, 0.199]$. We then estimate the range of system-wide UDs given all combinations of (μ_b, μ_w) in these bounds, using equations (5) and (6) along with the investigator-specific estimates from equations (11) and (12).

Figure 2 shows robust evidence of system-wide UD in foster care placement decisions. The range of possible system-wide UDs, as measured in the case-weighted average Δ_j, is estimated to be from around 1.2 percentage points to around 2.4 percentage points. This range is thus above the observed racial disparity in placement rates (around 1.2 percentage points). Intuitively a higher average Δ_j reflects the fact that the range of white mean risk is generally higher than the range of Black mean risk, consistent with the selected outcome rates in Panel D of Table 1. By standard OVB logic, if investigators are more likely to place a child in foster care when the maltreatment potential is high, and maltreatment potential is lower among Black cases, then adjusting for maltreatment potential Y_i^* in our UD measure should increase the estimated disparity.[27]

Tighter bounds on system-wide UD can be obtained by using the placement and future misconduct rates of investigators with low placement rates, as shown in equation (9). Column 1 of Appendix Table A4 shows that restricting our focus to investigators with an estimated (strata-adjusted) placement rate of 0.03 or lower yields mean risk bounds of $\mu_b \in [0.146, 0.176]$ and $\mu_w \in [0.167, 0.197]$. These race-specific bounds allow us to construct a tighter bound on system-wide UD, of $[0.014, 0.020]$. Columns 2 and 3 show the bounds are tightened further by

[27] Formally, $E[D_{ij} \mid R_i = b] - E[D_{ij} \mid R_i = w] - \Delta_j = [(\delta_{jw1} - \delta_{jw0})p_b + (\delta_{jb1} - \delta_{jb0})p_w](\mu_b - \mu_w)$ when investigators are randomly assigned, where $\delta_{jry} = E[D_{ij} \mid R_i = r, Y_i^* = y]$ gives race- and investigator-specific placement rates among cases with or without future maltreatment potential and $p_r = Pr(R_i = r)$ gives the share of Black or white cases (see Section II.B of Arnold et al. (2022) for a derivation). Since investigators are likely to respond to maltreatment potential ($\delta_{jr1} > \delta_{jr0}$), this formula shows adjusting for maltreatment potential differences is likely to increase placement rate disparities when maltreatment potential is higher among white cases than Black cases ($\mu_w > \mu_b$).

focusing on investigators with estimated placement rates below 0.02 and 0.01. We estimate system-wide UD to be in $[0.016, 0.018]$ with the narrowest mean risk bounds.

Extrapolating Maltreatment Variation Across Investigators. To obtain point estimates of UDs, we extrapolate variation in the race-specific placement and subsequent maltreatment rates across the quasi-randomly assigned investigators. Figure 3 plots this variation, with a binned scatterplot of the estimates from equations (11) and (12).[28]

A large number of investigators have a placement rate close to zero, suggesting plausible grounds for extrapolation. We find that rates of future maltreatment among non-placed children tend to decrease with investigator placement tendencies for both white and Black children. This is consistent with a decision-model where, at the margin, investigators with greater tendencies to leave a child at home leave children with higher risk in the home; in such a model the similarity of the slopes by race suggests similar investigator skill at predicting maltreatment potential among white and Black children.[29] The strong negative relationship between investigators' placement tendencies and subsequent maltreatment is found with three specifications: linear, quadratic, and local linear regression.[30] The intercepts of each specification yield estimates of the key mean risk parameters.

Panel A of Table 3 reports our three sets of mean risk point estimates, which are very similar across specifications. As depicted by the black lines in Figure 2, our estimates fall on the low end of the non-parametric bounds. The most flexible local linear extrapolation indicates that the average future maltreatment rate is 0.177 (SE=0.002) in the population of white children and 0.154 (SE=0.003) in the population of Black children. Again, this two percentage point gap suggests that observational disparities in placement rates are likely to understate the extent of true UD in Michigan (negative OVB).

Panel B of Table 3 reports corresponding estimates of system-wide UD. All three extrapolations yield virtually identical estimates of 0.017 (SE=0.002), suggesting that, on average, Black children are placed in foster care at a 1.7 percentage point higher rate than white children with identical potential for future maltreatment. This represents a disparity of roughly half of the overall placement rate of 3.4 percentage points, and is nearly 90% larger than the observational disparity of 0.9 percentage points.

[28] Figure A2 shows that the extrapolation is nearly identical if we do not adjust for strata fixed effects.

[29] We formalize these observations with a parametric model of investigator decision-making in Online Appendix B. The slopes in Figure 3 can also be related to IV coefficients from race-specific regressions of maltreatment outcomes on placement decisions with investigator instruments. A negative slope suggests an IV coefficient finding foster care placement reduces future maltreatment, which is again consistent with investigators acting on predictions of maltreatment potential.

[30] Each specification is weighted inversely by the variance of estimation error in each investigator's re-investigation rates.

V.B Robustness

Table A5 considers robustness to alternative time frames for re-investigation, finding that the specific period considered has little impact. 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; this time horizon is somewhat arbitrary.[31] In Columns 2 through 5 of the table we instead use re-investigation within two, three, four or five months.[32] Reassuringly, we find very similar estimates of UD across all horizons.

Another 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 UD in our context by inflating the risk of a subsequent investigation for Black children left at home.

Nevertheless, we next examine robustness of our proxy for subsequent maltreatment potential. Columns 6, 7, and 8 of Table A5 show that we obtain very similar results when we instead consider the potential for a subsequent *substantiated* investigation, the potential for a subsequent investigation that includes physical abuse, or the potential for foster care placement within six months if left at home. 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.

We also explore this concern using a subset of data (discussed in Section VI) that contain the category of the initial maltreatment reporter. When we estimate UDs separately for children referred by mandated (social workers, educational, medical, and law enforcement personnel) and non-mandated reporters (neighbors or other family members), we find similar estimates of UD across these two reporter types: we find a point estimate of 0.018 for mandated reporters (SE=0.003) and 0.015 for non-mandated reporters (SE=0.002). These results suggest that our main estimates are not driven by biases from a particular reporter type.

[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] As mentioned above, we omit the first month after the start of an investigation since investigators have 30 days to complete the focal investigation, and we do not observe a disposition date in the data.

V.C Potential Drivers

As noted in Section IV, UDs can arise from direct discrimination on the basis of race 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 prior investigations) 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 adjust the estimated investigator- and race-specific placement and subsequent maltreatment rates by the child and investigation characteristics in Column 3 of Table 2. We then recompute mean maltreatment risk and system-wide UD with these adjusted rates.

Table A6 suggests limited scope for indirect discrimination on non-race characteristics. Adjusting for child- and investigation-specific traits leads to very similar estimates of system-wide UD relative to our baseline estimates in Table 3 (1.6 and 1.7 percentage points, respectively). This finding suggests UD is similar across these non-race characteristics; indeed, Table A7 shows limited heterogeneity across separate analyses conducted for particular child subgroups. While estimates of UD are more positive for female children relative to male children, younger children relative to older children, and investigations involving neglect as opposed to physical abuse, all estimates are 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 B, we estimate a structural model of investigator decision-making akin to that in Arnold et al. (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 of identical maltreatment potential.

V.D Heterogeneity and Policy Implications

As noted above, we find minimal heterogeneity in unwarranted disparity by observable child and investigation characteristics. To further unpack the policy implications of unwarranted disparity in child protection, this section considers heterogeneity along two other key dimensions: by case severity and by investigator traits.

Heterogeneity by Maltreatment Potential

Table 4 shows significant heterogeneity in the UDs by the unobserved level of maltreatment potential that we condition on. Recall that our baseline UD measure Δ_j is a weighted average of two conditional disparities: the disparity among children with potential for subsequent maltreatment (Δ_{j1}) and the disparity among children with no potential for subsequent maltreatment (Δ_{j0}). The table shows that our average 1.7 percentage point disparity is concentrated in the former at-risk population, with a system-wide average of Δ_{j1} ranging from 5.7 percentage points to 6.1 percentage points, depending on the extrapolation. Panel B shows that white children with maltreatment potential are placed in foster care at a rate of around 5.8%, and Black children are about twice as likely to be placed in foster care in this subpopulation.[33] In contrast, there is little disparity among children without maltreatment potential, with a statistically insignificant average of Δ_{j0} of around 0.8 percentage points.

Heterogeneity by Investigator Characteristics

We next examine heterogeneity in the UDs across the observable characteristics of investigators. Since UD primarily arises among children with subsequent maltreatment potential, we focus on heterogeneity in the investigator-specific Δ_{j1}'s, though results are similar for heterogeneity in the overall Δ_j's. Specifically, Table 5 reports estimates from OLS regressions of the estimated Δ_{j1} on several investigator characteristics.

Column 1 of Table 5 documents significant racial *concordance* effects in investigator decisions: 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.[34] 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. However, because the vast majority of investigators in Michigan are white, the result is that, on average, Black children have higher placement rates in cases with maltreatment potential.

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 also have lower levels of UD in this subpopulation.

[33] Panel B of Table 4 also shows that investigators do distinguish between children in homes with and without subsequent maltreatment potential. For both Black and white children, placement rates are much lower in the population without subsequent maltreatment potential ($Y^* = 0$) than in the population with subsequent maltreatment potential ($Y^* = 1$), reinforcing that investigator decision-making distinguishes between the two types of home circumstances.

[34] See Dee (2005); Alsan et al. (2019); Ba et al. (2021); Edmonds (2022); Gershenson et al. (2022) for evidence of such effects in other high-stakes settings.

Policy Implications

The estimates thus far show that, while Black children in Michigan have higher foster care placement rates than white children conditional on subsequent maltreatment risk, the disparity is entirely concentrated among high-risk cases. These results provide a better understanding of racial discrimination in child protection, which may help guide the appropriate policy response. In particular, 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 children well-being, reducing subsequent maltreatment and improving longer-term outcomes. First, the quasi-experimental variation in Figure 3 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 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 are found precisely in these high-risk cases ($Y^* = 1$).[35]

These results 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

[35] While the causal effects of foster care for children on the margin may vary by context (e.g., see Doyle (2007, 2008)), the effects are likely to be most positive precisely in cases where there is subsequent maltreatment potential in the home.

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. This conclusion is bolstered by the structural model analysis in Online Appendix B, which suggests marginal white children face significantly higher rates of future maltreatment when left at home. Our findings of racial concordance in investigator decisions suggest that the leniency afforded by white investigators to white parents may, perhaps counterintuitively, lead to worse outcomes for their children.

These findings add nuance to ongoing policy debates over the possible "over-placement" of Black children in foster care.[36] Consider, for example, a policy that lowers the placement rate of Black children to equalize white and Black placement rates among children with identical maltreatment potential. Because the placement rate disparity is entirely driven by disparities in the population of children with $Y_i^* = 1$, the policy would only affect placement rates in this subpopulation. A back-of-the-envelope calculation based on our estimates suggests such a policy would increase the number of Black children who are subsequently maltreated by 7% (from around 8,800 to around 9,400 in our sample period).[37]

V.E External Validity

Because CPS is administered at the state or local level and the decision-making process of CPS agents may vary across states, a natural concern is to what extent our key findings are generalizable outside of Michigan CPS. To examine this question, we use more limited but nationwide data from the National Child Abuse and Neglect Data System (NCANDS, 2023) from 2008 to 2016 (matching our main sample period). This dataset contains information on child maltreatment investigations in most states in the U.S.[38] The data contain 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 in other states), we leverage the fact that the

[36] See, for example, the Minnesota African American Family Preservation Act, as well as policy recommendations in the New York State Bar Association's "Resolution addressing systemic racism in the child welfare system of the State of New York."

[37] There are 64,489 investigations of Black children in our sample (Table 1). Combining this with our estimate of $\mu_b = 0.155$ (Table 3) suggests roughly 10,000 children will be subsequently maltreated if not placed in foster care. In the status quo, 12% (or 1,200) of such children are placed into foster care. In a counterfactual where this share is lowered to 6% to equalize the placement rate of white and Black children with maltreatment potential (Table 4), 600 Black children are placed instead. Thus the number of Black children who are subsequently maltreated when left at home would increase from 8,800 to 9,400.

[38] While not publicly available, these data can be obtained by researchers free of charge upon request.

non-parametric bounds introduced in Section IV can be applied to state-level statistics to estimate state-specific UDs. As mentioned above, these bounds tend to be informative in the CPS context, given low treatment rates in most states. Online Appendix C describes the data and approach in more detail, and shows that we can replicate our key findings for Michigan CPS using this more limited dataset.[39]

Non-parametric bounds for each state in the NCANDS data show that our primary findings are generalizable to most U.S. states. Overall, we estimate nationwide UD bounds of $[0.002, 0.017]$. That is, the average state in the NCANDS data places Black children in foster care at higher rates than white children conditional on future maltreatment potential. Figure 4 further plots the average estimate in state-specific UD bounds, separately for cases with and without maltreatment potential. As in Michigan, UD in cases without maltreatment potential is near zero in most states while UD in cases with maltreatment potential tends to be significantly larger. We estimate a nationwide average UD among cases without maltreatment potential of 0.5 percentage points, while the average UD among cases with maltreatment potential is estimated at 3.6 percentage points.

VI A Systems-Based Analysis of Discrimination

By design, our estimates so far capture unwarranted disparity in investigator placement decisions holding fixed any decisions prior to the start of investigations. This focus follows much of the academic literature and recent policy discussions, and is natural since investigators directly make placement recommendations. A more complete analysis of discrimination in child protection would, however, take into account the possibility that earlier decisions—namely, the decision to screen-in calls for investigation—could affect and interact with subsequent foster care placement decisions.[40] This kind of "systems-based" analysis can reveal whether there is discrimination in the screening stage, and whether discrimination in this initial phase is mitigated or amplified by decisions made in the subsequent phase.

In contrast to investigators, screeners have received little attention in the literature and in

[39] As explained in Online Appendix C, our approach with these data does not leverage quasi-random investigator assignment (which is not possible, since the dataset does not contain investigator identifiers and investigators may not be quasi-randomly assigned to cases in all states). The non-parametric UD bounds are derived from aggregate statistics and, unlike in our primary analysis, do not adjust for rotation fixed effects. Reassuringly, this matters little for our Michigan replication.

[40] The analysis focuses on discrimination at the two stages under the purview of CPS. Potential external sources of discrimination include policies affecting maltreatment reporting, family reunification, and maltreatment itself. Recall that foster care supervision and reunification are handled by a separate department within the Department of Health and Human Services. Thus, screeners and investigators—the decision-makers we study—are the only decision-makers within CPS.

the popular press, perhaps due to a lack of data on screeners' actions and the fact that the screening decision is less salient. Nevertheless, there are reasons to suspect that the screening decision matters for racial disparities. Screeners screen-out 40% of initial calls; these calls tend to be short (about 15 minutes), and quick decisions may be particularly susceptible to bias (Devine, 1989; Dovidio et al., 1997; Kahneman, 2011; Agan et al., 2023). Screeners are told to screen-in calls in a way that minimizes the likelihood of subsequent maltreatment if the call is screened out, and have substantial discretion in how to implement this charge.

A systems-based analysis requires quasi-random assignment of cases at each stage in the process. Importantly, in our context, screeners are quasi-randomly assigned conditional on a day by shift fixed effect, as described in Section II. We use a more recent subset of the MDHHS data (2017–2019) that contains information on both the hotline screeners assigned to each call and the outcomes of screened-out calls. In this subset, we can thus follow an initial hotline call through both the screening and investigation decisions.[41]

We use this novel variation to measure *total* UDs in placement among all calls, and to decompose this total across the screening and investigation phases. With additional structure, this decomposition, which builds on the framework of Bohren et al. (2022), reveals the share of racial discrimination in eventual placement decisions that is attributable to screener decisions versus investigator decisions. Online Appendix D describes the sample for this analysis in more detail and validates the identifying assumptions required for our approach.

VI.A Decomposing Total Discrimination Across Two Phases

To formalize our approach, suppose i now indexes incoming calls rather than screened-in cases. Omitting screener and investigator indexing for simplicity, let $S_i \in \{0,1\}$ equal one if incoming call i would be screened-in and let $D_i \in \{0,1\}$ equal one if the call, when screened-in, would result in foster care placement. Placement decisions for each call i are thus given by $P_i = S_i D_i$. Previously, we implicitly conditioned on $S_i = 1$ (in which case $D_i = P_i$) to estimate UDs in investigator decisions. This UD measure, now denoted $\Delta^I = \Delta^I_0(1-\bar{\mu}) + \Delta^I_1\bar{\mu}$ for $\bar{\mu} = E[Y_i^* \mid S_i = 1]$, is composed of the conditional disparities:

$$\Delta^I_y \equiv E[D_i \mid R_i = b, Y_i^* = y, S_i = 1] - E[D_i \mid R_i = w, Y_i^* = y, S_i = 1]. \tag{13}$$

Now, without conditioning on screening, we can also consider a measure of "total" UD composed of conditional-on-Y_i^* disparities in foster care placement across the full population

[41] As we show below, estimates of UDs in investigator placement decisions are nearly identical in this later sample as those in Section V.

of calls:

$$\Delta_y^T \equiv E[P_i \mid R_i = b, Y_i^* = y] - E[P_i \mid R_i = w, Y_i^* = y]. \tag{14}$$

Namely, we consider $\Delta^T = \Delta_0^T(1 - \bar{\mu}^{full}) + \Delta_1^T \bar{\mu}^{full}$ for $\bar{\mu}^{full} = E[Y_i^*]$.

To relate the investigator UDs studied previously to these new total UD measures, we use the fact that:

$$E[P_i \mid R_i, Y_i^*] = E[D_i \mid R_i, Y_i^*, S_i = 1] Pr(S_i = 1 \mid R_i, Y_i^*). \tag{15}$$

Hence, we can decompose equation (14) and relate it to equation (13) as:

$$\Delta_y^T = \Delta_y^I Pr(S_i = 1 \mid R_i = b, Y_i^* = y) + \Delta_y^S Pr(D_i = 1 \mid R_i = w, Y_i^* = y, S_i = 1) \tag{16}$$

where

$$\Delta_y^S \equiv E[S_i \mid R_i = b, Y_i^* = y] - E[S_i \mid R_i = w, Y_i^* = y]. \tag{17}$$

Equation (16) decomposes total UDs in the full sample of calls, Δ_y^T, into two terms. The first term involves UDs by investigators among screened-in calls, Δ_y^I. The second term involves UDs by screeners in the full sample of calls, Δ_y^S. In equation (16), these two disparities are weighted by the screening rate of Black children with $Y_i^* = y$ and the conditional-on-screening placement rate of white children with $Y_i^* = y$, respectively. This is akin to a Kitagawa–Oaxaca-Blinder (KOB) decomposition, derived by adding and subtracting $E[D_i \mid R_i = w, Y_i^* = y, S_i = 1] Pr(S_i = 1 \mid R_i = b, Y_i^* = y)$ and rearranging terms. Bohren et al. (2022) propose similar KOB decompositions for studying how discrimination evolves over systems comprised of multiple decision-makers.[42]

As usual with KOB decompositions, an alternative version of equation (16) comes from changing the "order" of the decomposition across the two phases. Namely, we can write:

$$\Delta_y^T = \Delta_y^I Pr(S_i = 1 \mid R_i = w, Y_i^* = y) + \Delta_y^S Pr(D_i = 1 \mid R_i = b, Y_i^* = y, S_i = 1), \tag{18}$$

by swapping the race of the children used to weight the Δ_y^I and Δ_y^S terms. We estimate both decompositions below.

[42] In the Bohren et al. (2022) framework, the first term of equation (16) can be seen as isolating discrimination from "signal inflation": i.e., from racial disparities in the generation of a "screened-in" signal S_i which constrains investigator actions. The second term of equation (16) potentially incorporates indirect discrimination from other signals, though as noted above we find minimal role of such non-race characteristics in driving investigator UDs.

As before, the challenge of estimating total UD is the selective observability of Y_i^*. We follow a similar approach as above to address this challenge. Specifically, we leverage the quasi-experimental assignment of screeners to estimate mean risk by race in the full population of calls $\mu_r^{full} = E[Y_i^* \mid R_i = r]$, and we then use these estimates to compute system-wide UD in this population. These mean risk estimates also yield estimates of screener UDs and the corresponding conditional screen-in rates $Pr(S_i = 1 \mid R_i = r, Y_i^* = y)$, which enter the decompositions. To complete the decompositions, we estimate system-wide UD in the screened-in subpopulation. We use the quasi-experimental assignment of investigators to estimate μ_r as before, but in the later sample of years where we observe both screener and investigator decisions. As we show below, mean risk and UD estimates in this sample are nearly identical to those in Table 3. This step also yields estimates of the conditional placement rates $Pr(D_i = 1 \mid R_i = r, Y_i^* = y, S_i = 1)$, which enter the decompositions.

As in our main analysis, leveraging the quasi-random assignment of screeners for identification relies on an implicit exclusion restriction: that screener tendencies impact subsequent outcomes solely through the decision to screen-in calls. This assumption is inherently untestable, but seems plausible given the limited channels by which screeners can impact cases after their short calls. However, this analysis also relies on an additional assumption: that the act of investigation itself does not change the potential for future maltreatment in the home (except through foster care placement). We provide evidence for this assumption in Online Appendix D. Specifically, we derive bounds for such direct effects using the two-stage quasi-randomization of screeners and investigators and show that, at least among marginal cases, these bounds are tight and straddle zero. Nevertheless, this is an additional assumption that is not required in the analyses in Section V.

The precision of these bounds and the estimation of μ_r^{full} are assisted by relatively low placement rates (since foster care placement censors observations of Y_i^*). In the full sample of calls, placement rates are even lower than among screened-in-calls (2%, as shown in Table A9), since $P_i = D_i S_i$ is zero whenever S_i or D_i are zero. We follow our baseline local linear extrapolation strategy to point identify μ_r^{full}—however, given the low placement rates, non-parametric bounds on μ_r^{full} and the other objects of interest are quite narrow.[43]

VI.B Results

We first show that the main patterns of investigator UDs shown in Section V hold up in the full population of calls: there is significant evidence of UDs, and these are concentrated among

[43]Figure A4 displays the extrapolations in the full population of calls, while Figure A5 displays non-parametric bounds instead.

calls with subsequent maltreatment potential. Panel A of Table 6 shows precise estimates of $\mu_b^{full} = 0.130$ and $\mu_w^{full} = 0.138$, while Panel B shows that total UD (Δ^T) is 1.1 percentage points (55% of the placement rate in the full sample of calls). This gap is larger than the unconditional disparity of 0.08 percentage points (shown in Table A9), again showing evidence of negative OVB via the difference in mean risk parameters. Panel B also shows that total UD is driven by disparities in the population with maltreatment potential. Specifically, the Black-white disparity in placement rates conditional on maltreatment potential (Δ_1^T) is nearly four percentage points, while the disparity in the population with no maltreatment potential (Δ_0^T) is significantly smaller (0.6 percentage points) and statistically insignificant.

We next document UDs in screener decisions. Panel B of Table 6 shows that calls involving Black children are screened-in at a rate 4.9 percentage points higher than those involving white children with identical maltreatment potential. This represents a disparity of roughly 8%, relative to a screen-in rate of 60% (see Table A9). In contrast to investigator decisions, this disparity is driven by both calls with and without subsequent maltreatment potential (4.5 and 5 percentage points, respectively). One possible explanation for this result is that the speed at which screeners make decisions makes it hard to infer maltreatment potential.

Taken together, these findings reveal an interesting pattern: Ultimate UDs in foster care placement decisions are entirely driven by disparities in the population of calls with subsequent maltreatment, despite UDs in screener decisions in both subpopulations ($Y^* = 1$ and $Y^* = 0$). We use the KOB decomposition to reconcile these results. We first replicate our main estimates of investigator UDs (in the population of screened-in calls), but in this more recent subsample. Panel B of Table 6 shows an overall disparity of 1.6 percentage points which is again driven by children with subsequent maltreatment potential.[44] The pattern of investigator UDs suggest that, even though screeners' decisions yield UDs in the subpopulation of calls with $Y^* = 0$, investigators, who make the ultimate placement recommendations, do not discriminate in this subpopulation. Indeed, as mentioned above, investigators are skilled at inferring risk and place a very small number of children with $Y^* = 0$ (Table 4, Panel B), thereby mitigating initial UDs at the screening phase in this subpopulation.

In contrast, the estimates in Panel B of Table 6 show that investigators amplify initial screener UDs in the population with $Y^* = 1$, despite observing effectively the same information and deliberating over a much longer time frame. This suggests that both screeners and investigators contribute to the eventual "under-placement" of white children. To quantify how much of the total "under-placement" is driven by screeners versus investigators, we obtain estimates of the two KOB decompositions. We focus on the subpopulation with $Y^* = 1$, since

[44]These estimates are nearly identical to those in Section V, suggesting stable patterns of UD over time.

total UD in the subpopulation with $Y^* = 0$ is virtually zero. Estimates in Panel C of Table 6 show that screener decisions account for 8-14% of eventual disparities in this subpopulation.

The fact that call screeners drive a significant share of eventual UD in foster care placement is 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. At the same time, the fact that investigators are the main contributors to total UDs underscores the importance of our primary analyses that focus on their decisions.

VII Conclusion

This study is the first to use quasi-experimental variation to measure racial discrimination in foster care placement decisions. We find substantial unwarranted disparity in the decisions of CPS investigators in Michigan, which exceeds observational racial disparities in placement rates because of negative OVB. Specifically, comparing Black and white children with the same potential for future maltreatment in the home, we find that Black children are nearly 2 percentage points (50%) more likely to be placed in foster care following a CPS investigation. This UD is concentrated in the population of children with subsequent maltreatment potential, is more pronounced among white investigators, and appears to be driven by direct discrimination in the form of racial bias. Together, these findings add nuance to ongoing policy recommendations focused on addressing UDs by raising the threshold to place Black children in foster care. Our findings show how UDs can arise from the relative under-placement of at-risk white children, perhaps through white investigators giving a broad benefit of the doubt to white parents. Using more limited nationwide data, we show that our primary findings generalize to most CPS systems in the U.S.

This study is also the first to use the quasi-random assignment of multiple sets of decision-makers to study how discrimination can perpetuate and compound across systems with multiple phases. Our estimates show that eliminating total UDs in foster care placement may require intervention at both phases of CPS involvement: policies that are able to reduce UDs in only the investigation or screening phase leave behind discrimination. Our KOB decomposition gives a simple empirical framework for crafting appropriate policy responses. Since the decomposition varies by future maltreatment potential, policymakers with scarce resources can structure policy responses as a function of both the relative costs of intervening at the screening versus investigation phases and how they weigh false positives and false negatives in placement decisions. Developing such policy responses given the findings in this paper is an important next step.

References

Agan, Amanda Y, Diag Davenport, Jens Ludwig, and Sendhil Mullainathan. 2023. "Automating Automaticity: How the Context of Human Choice Affects the Extent of Algorithmic Bias." National Bureau of Economic Research Working Paper 30981.

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

Alsan, Marcella, Owen Garrick, and Grant Graziani. 2019. "Does diversity matter for health? Experimental evidence from Oakland." *American Economic Review* 109 (12): 4071–4111.

Andrews, Donald WK, and Marcia MA Schafgans. 1998. "Semiparametric Estimation of the Intercept of a Sample Selection Model." *The 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.

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

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 S 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." *The 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.

Ba, Bocar A, Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021. "The role of officer race and gender in police-civilian interactions in Chicago." *Science* 371 (6530): 696–702.

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

Bald, Anthony, Joseph Doyle Jr., Max Gross, and Brian Jacob. 2022b. "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.

Biehal, Nina. 2014. "Maltreatment in Foster Care: A Review of the Evidence." *Child Abuse Review* 23 (1): 48–60.

Billingsley, Andrew, and Jeanne M Giovannoni. 1972. *Children of the Storm: Black Children and American Child Welfare.* Harcourt, Brace, Jovanovich.

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." *The 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.

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

Chibnall, Susan, Nicole M. Dutch, Brenda Jones-Harden, Annie Brown, Ruby Gourdine, Jacqueline Smith, Anniglo Boone, and Shelita Snyder. 2003. *Children of Color in the Child Welfare System: Perspectives from the Child Welfare Community.* Department of Health and Human Services, Children's Bureau, Administration for Children and Families.

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.

Courtney, Mark E, Richard P Barth, Jill Duerr Berrick, Devon Brooks, Barbara Needell, Linda Park, and Richard Needell. 1996. "Race and Child Welfare Services: Past Research and Future Directions." *Child Welfare* 75 (2): 99–137.

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.

Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" *American Economic Review* 95 (2): 158–165.

Dettlaff, Alan J, and Reiko Boyd. 2020. "Racial disproportionality and disparities in the child welfare system: Why do they exist, and what can be done to address them?" *The ANNALS of the American Academy of Political and Social Science* 692 (1): 253–274.

Devine, Patricia G. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled Components." *Journal of Personality and Social Psychology* 56 (1): 5–18.

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.

Dovidio, John F, Kerry Kawakami, Craig Johnson, Brenda Johnson, and Adaiah Howard. 1997. "On the Nature of Prejudice: Automatic and Controlled Processes." *Journal of Experimental Social Psychology* 33 (5): 510–540.

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.

Edmonds, Lavar. 2022. "Role models revisited: HBCUs, Same-Race Teacher Effects, and Black Student Achievement." Working paper.

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.

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.

Font, Sarah A, and Kathryn Maguire-Jack. 2020. "It's Not "Just poverty": Educational, Social, and Economic Functioning Among Young Adults Exposed to Childhood Neglect, Abuse, and Poverty." *Child Abuse & Neglect* 101 104356.

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

Gershenson, Seth, Cassandra MD Hart, Joshua Hyman, Constance A Lindsay, and Nicholas W Papageorge. 2022. "The long-run impacts of same-race teachers." *American Economic Journal: Economic Policy* 14 (4): 300–342.

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. 2023. "Effects of the Child Protection System on Parents." Working paper.

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

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.

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

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

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Macmillan.

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.

Lansford, Jennifer E, Patrick S Malone, Kenneth A Dodge, Joseph C Crozier, Gregory S Pettit, and John E Bates. 2006. "A 12-Year Prospective Study of Patterns of Social Information Processing Problems and Externalizing Behaviors." *Journal of Abnormal Child Psychology* 34 (5): 709–718.

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.

Palmer, Lindsey, Sarah Font, Andrea Lane Eastman, Lillie Guo, and Emily Putnam-Hornstein. 2022. "What Does Child Protective Services Investigate as Neglect? A Population-Based Study." *Child Maltreatment*.

Paxson, Christina, and Jane Waldfogel. 1999. "Parental Resources and Child Abuse and Neglect." *American Economic Review* 89 (2): 239–244.

Paxson, Christina, and Jane Waldfogel. 2002. "Work, Welfare, and Child Maltreatment." *Journal of Labor Economics* 20 (3): 435–474.

Peterson, Cora, Curtis Florence, and Joanne Klevens. 2018. "The Economic Burden of Child Maltreatment in the United States, 2015." *Child abuse & neglect* 86 178–183.

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.

Pryce, Jessica, Wonhyung Lee, Elizabeth Crowe, Daejun Park, Mary McCarthy, and Greg Owens. 2019. "A Case Study in Public Child Welfare: County-Level Practices that Address Racial Disparity in Foster Care Placement." *Journal of Public Child Welfare* 13 (1): 35–59.

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.

Raz, Mical, and Vivek Sankaran. 2019. "Opposing Family Separation Policies for the Welfare of Children." *American Journal of Public Health* 109 (11): 1529.

Rebbe, Rebecca. 2018. "What is Neglect? State Legal Definitions in the United States." *Child Maltreatment* 23 (3): 303–315.

Reddy, Julia, Anne Williams-Isom, and Emily Putnam-Hornstein. 2022. "Racial Sensitivity Training: An Inadequate Solution to Systemic Racial Disparities in Child Protection Systems." *Child Abuse & Neglect* 128 105584.

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, Dorothy. 2009. *Shattered Bonds: The Color of Child Welfare*. Hachette UK.

Roberts, Kelsey V. 2019. "Foster Care and Child Welfare." Working paper.

Shaw, Terry V, Emily Putnam-Hornstein, Joseph Magruder, and Barbara Needell. 2008. "Measuring Racial Disparity in Child Welfare." *Child Welfare* 87 (2): 23–36.

Skiena, Steven, and Charles B Ward. 2014. *Who's Bigger?: Where Historical Figures Really Rank*. Cambridge University Press.

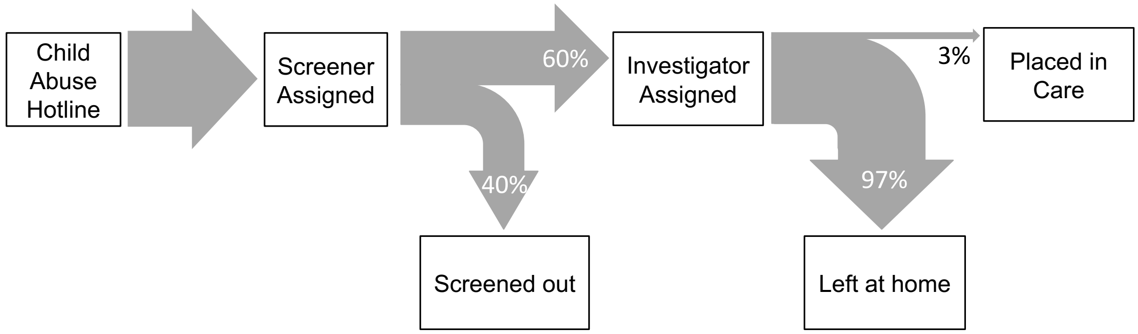
Soares, Sara, Vania Rocha, Michelle Kelly-Irving, Silvia Stringhini, and Silvia Fraga. 2021. "Adverse Childhood Events and Health Biomarkers: A Systematic Review." *Frontiers in Public Health* 9 649825.

USDHHS. 2017. "Child Welfare Outcomes Report Data." United States Department of Health and Human Services. https://cwoutcomes.acf.hhs.gov/cwodatasite/recurrence/index.

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.

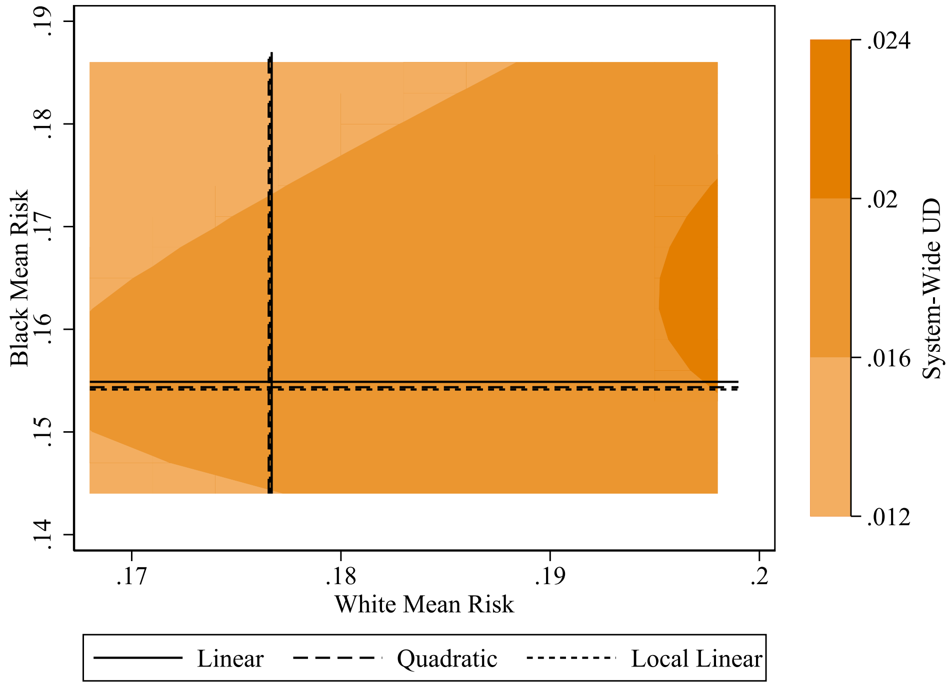
Wulczyn, Fred, Robert Gibbons, Lonnie Snowden, and Bridgette Lery. 2013. "Poverty, Social Disadvantage, and the Black/White Placement Gap." *Children and Youth Services Review* 35 (1): 65–74.

Figure 1: Child Protection in Michigan



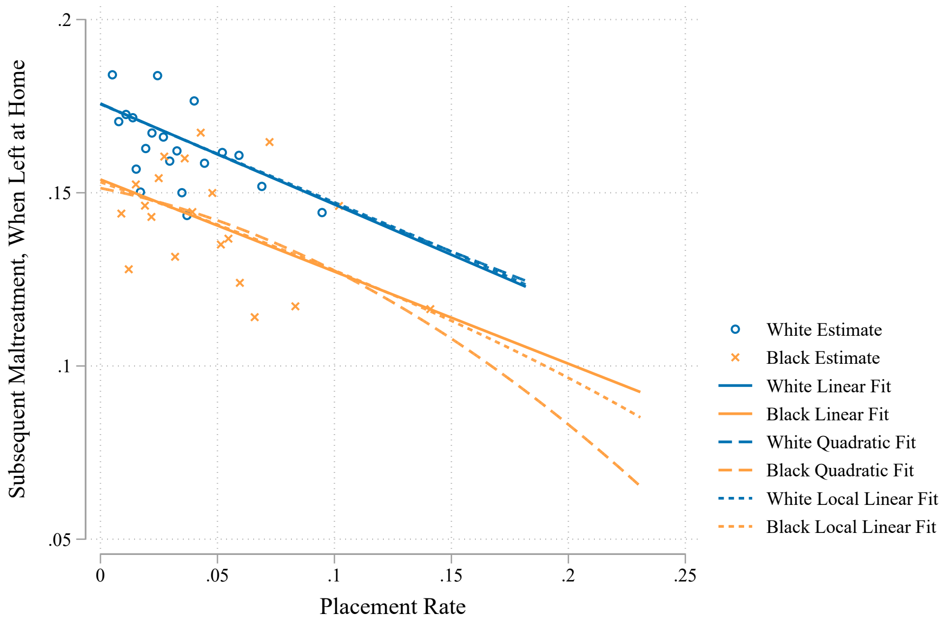
Notes: The figure describes the child protection process in Michigan. Both screeners and investigators are quasi-randomly assigned. The percentages on screening in and out refer to all calls received; percentages thereafter refer to investigated cases.

Figure 2: Robustness of System-Wide UD Estimates



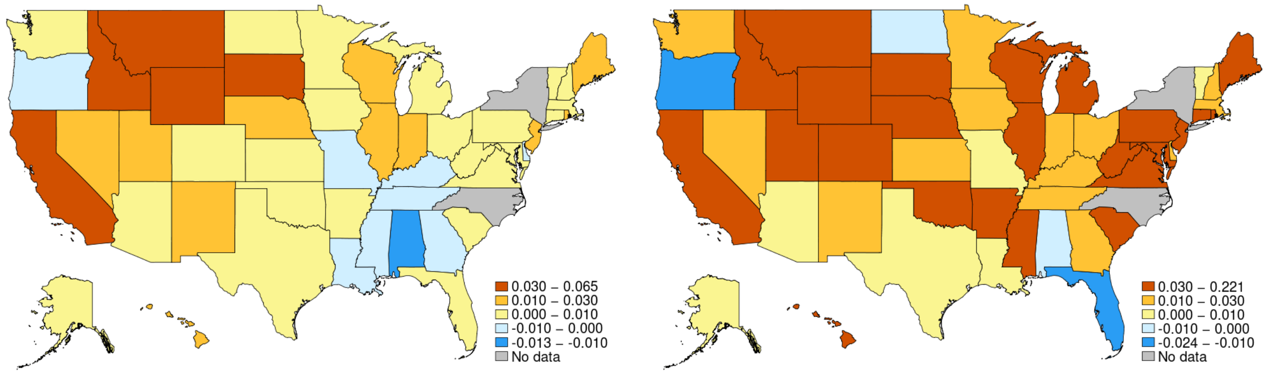
Notes. This figure shows how our estimates of system-wide UD change under different estimates of Black and white mean risk. The mean risk estimates obtained from the linear, quadratic, and local linear extrapolations in Panel A of Table 3 are indicated by solid, dashed, and dotted lines, respectively. The ranges of Black and white mean risk reflect the bounds implied by the average placement and subsequent maltreatment rates in the sample: $\mu_b \in [0.144, 0.187]$ and $\mu_w \in [0.168, 0.199]$.

Figure 3: Investigator Placement and Conditional Maltreatment Rates



Notes. The figure shows a binscatter (20 equal-sized bins) of placement rates, by race, for the 699 investigators in our sample against rates of a subsequent maltreatment investigation within 6 months for children left at home. All estimates adjust for ZIP code by investigation year fixed effects. The figure also plots race-specific linear, quadratic, and local linear curves of best fit, obtained from investigator-level regressions that inverse-weight by the variance of the estimated re-investigation rate among children left at home. The local linear regressions use a Gaussian kernel with a race-specific rule-of-thumb bandwidth (0.11).

Figure 4: Nationwide Unwarranted Disparity by Underlying Case Severity



(a) No maltreatment potential ($Y^* = 0$) (b) Maltreatment potential ($Y^* = 1$)

Notes. This figure uses national data from NCANDS to show the mean of state-specific UD bounds between 2008 and 2016. Panel (a) presents estimates for cases without maltreatment potential while Panel (b) presents estimates for cases with maltreatment potential. See Online Appendix C for additional details regarding the dataset and methodology.

Table 1: Summary Statistics

	(1) All Children	(2) Black Children	(3) White Children
	Panel A: Child Characteristics		
White	0.704	0.000	1.000
Female	0.482	0.483	0.481
Age at investigation	6.804	6.467	6.945
Child had a previous investigation	0.455	0.447	0.458
Number of previous investigations	1.030	0.947	1.065
	Panel B: Investigation characteristics		
Investigation included a...			
Domestic violence allegation	0.106	0.099	0.108
Drug residence allegation	0.031	0.034	0.030
Physical or substance abuse allegation	0.426	0.412	0.432
Physical neglect allegation	0.445	0.483	0.429
Improper supervision allegation	0.525	0.493	0.538
Threatened harm allegation	0.329	0.325	0.331
Alleged perpetrator includes the parent/step-parent	0.910	0.902	0.913
Alleged perpetrator includes a non-parent relative	0.052	0.063	0.048
Alleged perpetrator includes someone unrelated	0.103	0.095	0.106
	Panel C: Treatment		
Foster care placement rate	0.034	0.043	0.031
	Panel D: Outcomes, if not placed		
Re-investigated for child maltreatment within...			
2 months	0.065	0.060	0.067
3 months	0.094	0.085	0.097
4 months	0.120	0.108	0.125
5 months	0.144	0.129	0.150
6 months	0.166	0.150	0.173
Number of investigations	217,704	64,489	153,215
Number of children	181,928	54,873	127,055

Notes. This table summarizes the analysis sample. The sample consists of maltreatment investigations of children in MI (0-16 years old) between 2008 and 2016, assigned to investigators with at least 200 cases of both white and black children during this period. The sample excludes investigations of sexual abuse (about 1% of all investigations), since these cases are not quasi-randomly assigned. The final sample consists of 217,704 unique investigations of 181,928 children assigned to 699 investigators.

Table 2: Descriptive Regressions of Foster Care Placement

	(1)	(2)	(3)
Black	0.012***	0.012***	0.009***
	(0.002)	(0.002)	(0.001)
Female			-0.001
			(0.001)
Child had a previous investigation			0.002*
			(0.001)
Number of previous investigations			0.002***
			(0.000)
Domestic violence allegation			-0.004*
			(0.002)
Drug residence allegation			0.045***
			(0.007)
Physical or substance abuse allegation			0.015***
			(0.001)
Physical neglect allegation			0.024***
			(0.002)
Improper supervision allegation			0.006***
			(0.001)
Threatened harm allegation			0.024***
			(0.002)
Alleged perpetrator includes the parent/step-parent			0.017***
			(0.005)
Alleged perpetrator includes a non-parent relative			-0.000
			(0.002)
Alleged perpetrator includes someone unrelated			0.020***
			(0.002)
Mean placement rate	0.034	0.034	0.034
Number of investigations	217,704	217,704	217,704
Age dummies	N	N	Y
Rotation Group FE	N	Y	Y

Notes. This table reports OLS estimates of regressions of an indicator equal to one if the child was placed in foster care on child and investigation characteristics. The regressions are estimated on the sample described in Table 1. Robust standard errors, two-way clustered at the child and investigator levels, are reported in parentheses. Age dummy variables are included in Column 3 but are excluded for ease of exposition.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Estimates of Mean Maltreatment Risk by Race and UD

	(1) **Linear extrapolation**	**(2)** **Quadratic extrapolation**	**(3)** **Local linear extrapolation**
	Panel A: Mean maltreatment risk by race		
Black children	0.155***	0.154***	0.154***
	(0.002)	(0.002)	(0.003)
White children	0.176***	0.177***	0.177***
	(0.002)	(0.002)	(0.002)
	Panel B: System-wide UD		
Mean across investigators	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)
Mean placement rate	0.034	0.034	0.034
Number of investigators	699	699	699

Notes. The table summarizes estimates of mean maltreatment risk and UD from different extrapolations of the variation in Figure 3. Panel A reports estimates of race-specific average re-investigation risk. Panel B reports estimates of system-wide UD (weighted by the number of investigations assigned to each investigator). To estimate mean risk, Column 1 uses a linear extrapolation of the variation in Figure 3, while Column 2 uses a quadratic extrapolation and Column 3 uses a local linear extrapolation with a Gaussian kernel and a rule-of-thumb bandwidth. Robust standard errors, two-way clustered at the child and investigator level, are obtained by a bootstrapping procedure (500 times) and appear in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Decomposition of UD by Subsequent Maltreatment Potential

	(1) Linear extrapolation	**(2) Quadratic extrapolation**	**(3) Local linear extrapolation**
	Panel A: Conditional UD		
With maltreatment potential	0.061***	0.058***	0.057***
	(0.021)	(0.021)	(0.021)
No maltreatment potential	0.008	0.008	0.009
	(0.006)	(0.006)	(0.006)
	Panel B: Conditional placement rates		
Black, with maltreatment potential	0.119	0.117	0.115
White, with maltreatment potential	0.059	0.058	0.058
Black, no maltreatment potential	0.034	0.034	0.034
White, no maltreatment potential	0.025	0.026	0.026
Number of investigators	699	699	699

Notes. This table summarizes estimates of racial disparities in true/false negative rates from different extrapolations of the variation in Figure 3. Δ_{j1} reflects racial disparities among children with potential for subsequent maltreatment ($Y_i^* = 1$) and Δ_{j0} captures racial disparities among children without potential for subsequent maltreatment ($Y_i^* = 0$). Panel B shows removal rates conditional on Y_i^* and race. Robust standard errors, two-way clustered at the child and investigator level, are obtained by a bootstrapping procedure (500 times) and appear in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Regressions of UD On Investigator Characteristics

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Black investigator	-0.141**						-0.081
	(0.063)						(0.066)
Other race	-0.224						-0.208
	(0.150)						(0.152)
Female investigator		-0.118**					-0.108**
		(0.052)					(0.055)
New investigator			-0.067				-0.060
			(0.049)				(0.054)
Urban county				-0.152***			-0.063
				(0.048)			(0.050)
High share of Black children					-0.158***		-0.102**
					(0.048)		(0.047)
Low placement rate						-0.009	-0.012
						(0.050)	(0.054)
Control Mean	0.081	0.141	0.086	0.132	0.136	0.064	0.059
Number of investigators	699	699	699	699	699	699	699

Notes. This table reports OLS estimates of regressions of estimates of Δ_{j1} (based on a local linear extrapolation) on investigator observable characteristics. 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 in the entire sample of investigators, as previously reported in Panel A, Column 3 of Table 4. 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 are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Systems-Based Analysis of Discrimination

	(1)	(2)	(3)
	Panel A: Mean maltreatment risk		
	Full sample		**Screened-in sample**
Black children	0.130***		0.132***
	(0.004)		(0.003)
White children	0.138***		0.147***
	(0.003)		(0.002)
	Panel B: System-wide UD		
	Total	**Screeners**	**Investigators**
Total	0.011***	0.049***	0.016***
	(0.001)	(0.011)	(0.002)
Maltreatment potential	0.038***	0.045***	0.055***
	(0.014)	(0.014)	(0.021)
No maltreatment potential	0.006	0.050***	0.008
	(0.005)	(0.011)	(0.006)
	Panel C: KOB decompositions (first/second)		
	Total	**Screeners**	**Investigators**
With maltreatment potential	1.00	0.080 / 0.140	0.920 / 0.860

Notes. The table summarizes estimates of mean risk of maltreatment within 6 months and system-wide UD in screener and investigation decisions, based on local linear extrapolations as described in the text. The sample includes 190,776 cases, 162 screeners, and 642 investigators in Michigan from 2017–2019. Panel A shows mean risk by race in the full population of calls received at the child abuse hotline as well as in the subsample of screened-in calls. Panel B shows estimates of total UD (Δ^T), as well as UDs in screener decisions (Δ^S) and investigator decisions (Δ^I), both among all children and by maltreatment potential. Panel C shows our two KOB decompositions of total UD among all children and by maltreatment potential. Standard errors are obtained by a bootstrapping procedure (500 times) and appear in parentheses. Standard errors in the first and second columns are clustered at the screener level, while standard errors in the third column are clustered at the investigator level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

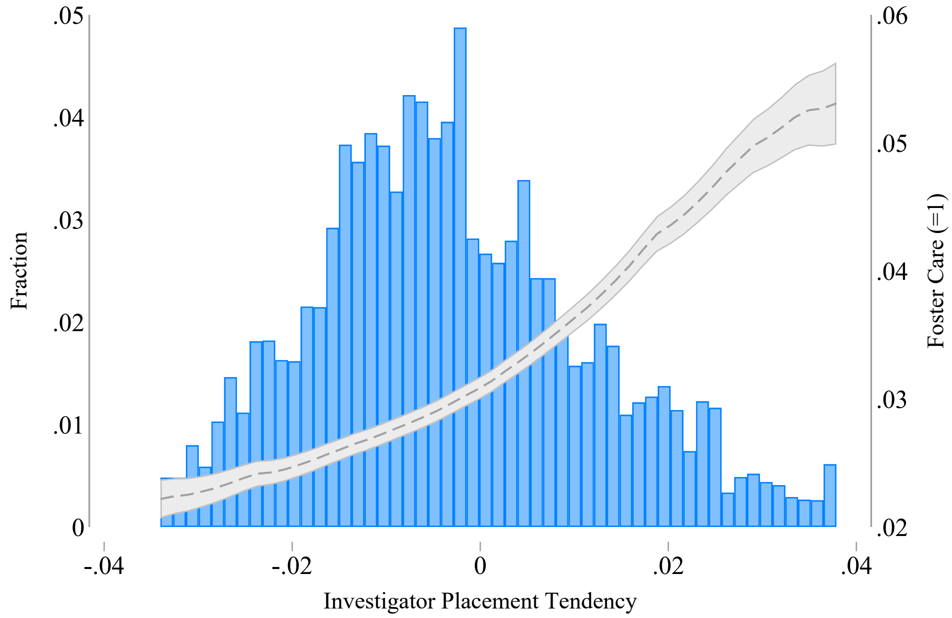
Racial Discrimination in Child Protection

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

Online Appendix

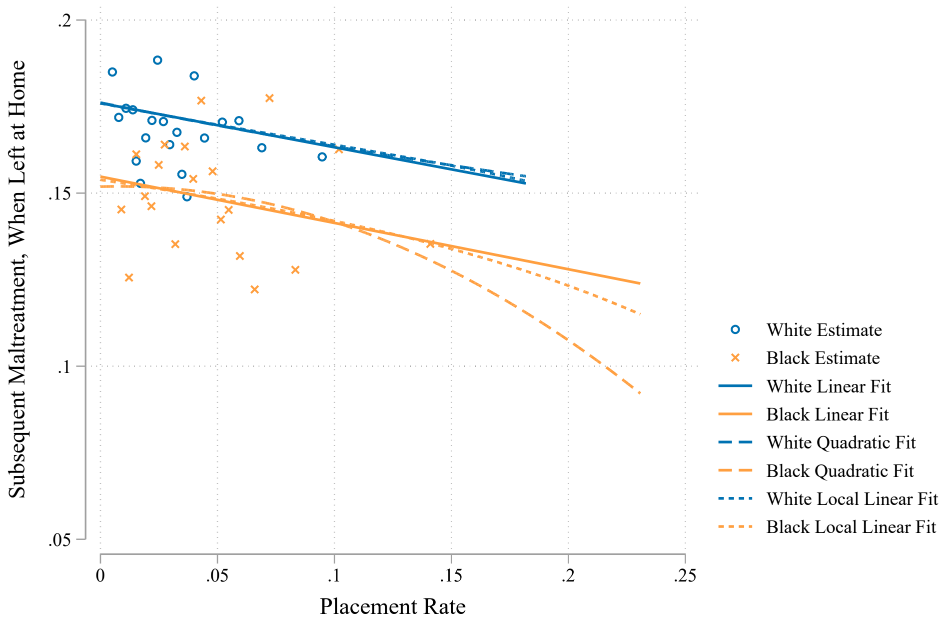
A Supplemental Figures and Tables

Figure A1: Distribution of Investigator Placement Tendency



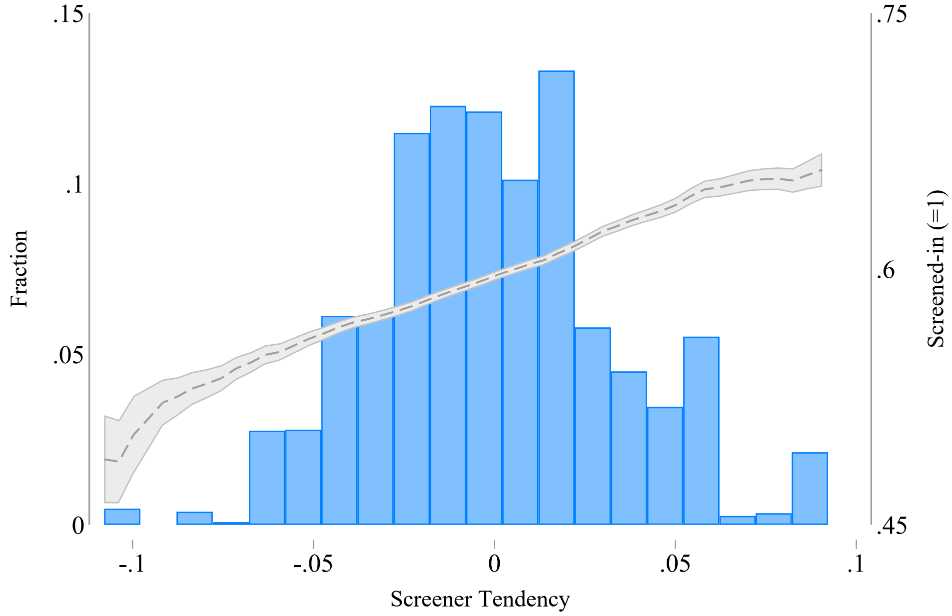
Notes. This figure shows the distribution of the investigator foster care placement tendency instrument net of ZIP code by investigation year fixed effects. The dashed line shows point estimates from a non-parametric regression of an indicator variable equal to one if the child was placed in foster care on Z and the shaded region shows the 95 percent confidence interval. Standard errors are clustered by child and investigator.

Figure A2: Investigator Placement and Conditional Maltreatment Rates Without Strata Adjustment



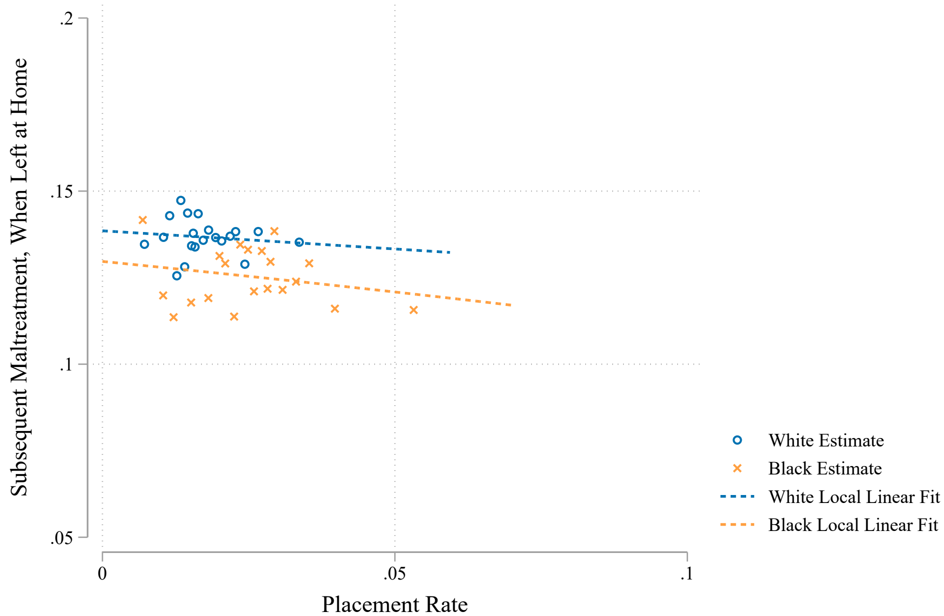
Notes. The figure shows a binscatter (20 equal-sized bins) of placement rates, by race, for the 699 investigators in our sample against rates of a subsequent maltreatment investigation within 6 months for children left at home. In contrast to Figure 3, these estimates do not adjust for ZIP code by investigation year fixed effects. The figure also plots race-specific linear, quadratic, and local linear curves of best fit, obtained from investigator-level regressions that inverse-weight by the variance of the estimated re-investigation rate among children left at home. The local linear regressions use a Gaussian kernel with a race-specific rule-of-thumb bandwidth.

Figure A3: Distribution of Screener Tendencies



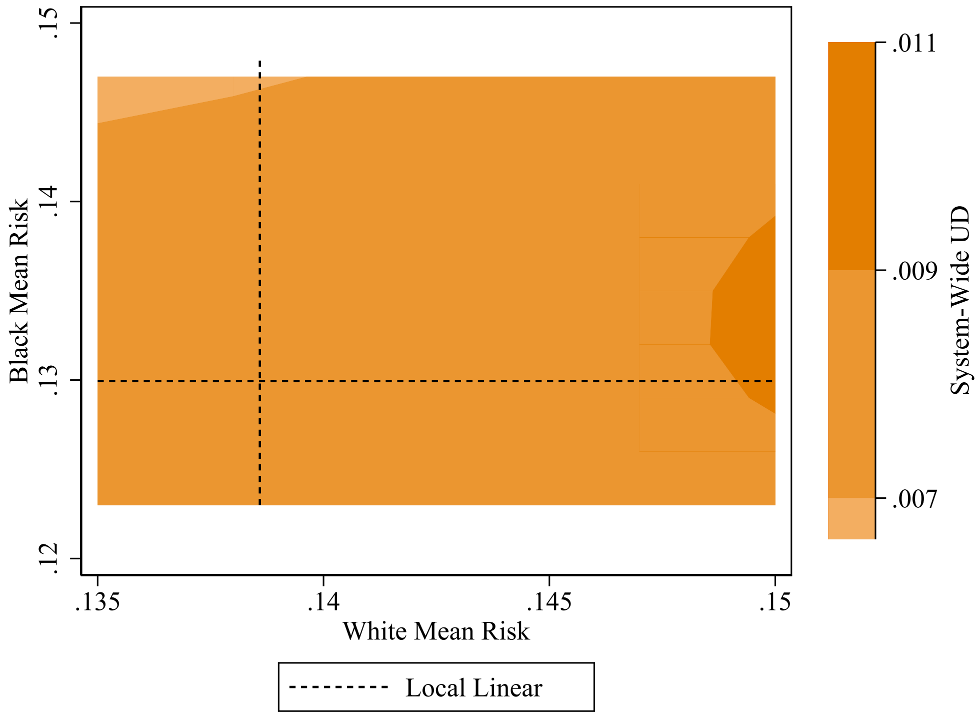
Notes. This figure shows the distribution of the screener tendency instrument net of day × shift fixed effects. The dashed line shows point estimates from a non-parametric regression of an indicator variable equal to one if the child was screened-in on Z and the shaded region shows the 95 percent confidence interval. Standard errors are clustered by screener.

Figure A4: Screener Placement and Conditional Maltreatment Rates



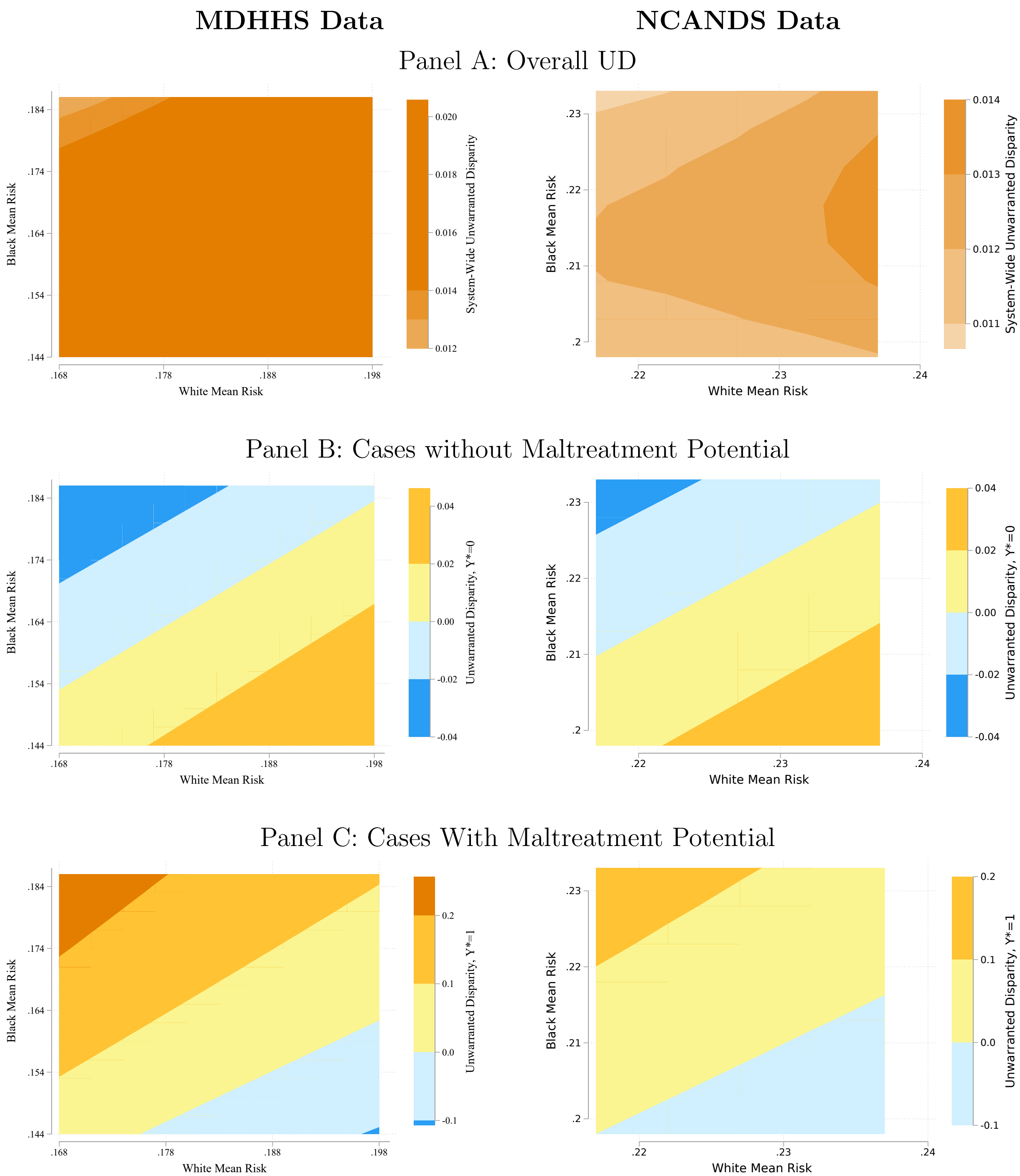
Notes. The figure shows a binscatter (20 equal-sized bins) of placement rates, by race, for the 162 screeners in our sample against rates of a subsequent maltreatment investigation within 6 months for children left at home. All estimates adjust for day by shift fixed effects. The figure also plots race-specific 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. The local linear regressions use a Gaussian kernel with a race-specific rule-of-thumb bandwidth.

Figure A5: Robustness of UD Estimates in the Full Population of Calls



Notes. This figure shows how our estimates of Δ^T change under different estimates of Black (μ_b^{full}) and white mean risk (μ_w^{full}) in the full population of calls. The mean risk estimates obtained from local linear extrapolations in Figure A4 are indicated by dashed lines. The ranges of Black and white mean risk reflect the bounds implied by the average placement and subsequent maltreatment rates in the sample: $\mu_b \in [0.123, 0.148]$ and $\mu_w \in [0.135, 0.152]$.

Figure A6: Non-Parametric Bounds of UD (MDHHS vs. NCANDS)



Notes: This figure shows non-parametric bounds of 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 UD change under different estimates of Black (μ_b) and white (μ_w) mean risk. The ranges of Black and white mean risk reflect the bounds implied by the average placement and subsequent maltreatment rates in the respective sample.

Table A1: First-Stage Effects of Investigator Removal Tendency

	(1)	(2)	(3)	(4)
	Black children		**White children**	
Placement tendency	0.548***	0.536***	0.625***	0.591***
	(0.048)	(0.046)	(0.039)	(0.038)
F-statistic	133	134	257	238
Observations	64,489	64,489	153,215	153,215
Rotation group FE	Y	Y	Y	Y
Baseline controls	N	Y	N	Y
Age fixed effects	N	Y	N	Y

Notes. This table reports OLS estimates of regressions of an indicator equal to one if the child was removed on investigator removal tendencies. The regressions are estimated on the sample described in Table 1. Removal tendency is estimated using data from other cases assigned to the investigator. All regressions control for zip code by investigation year fixed effects. Columns 2 and 4 include age fixed effects and the variables listed in Column 3 of Table 2. We 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 et al., 2019). Robust standard errors, two-way clustered at the child and investigator level, are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Balance Tests for Investigator Assignment

	(1)	(2)	(3)	(4)
	Black children		**White children**	
Dependent variable:	Placed	Placement Tendency	Placed	Placement Tendency
Female	-0.0011	0.0002	-0.0011	0.0001
	(0.0015)	(0.0002)	(0.0008)	(0.0001)
Child had a previous investigation	0.0051*	-0.0004	0.0027**	0.0001
	(0.0027)	(0.0004)	(0.0014)	(0.0001)
Number of previous investigations	0.0018*	-0.0000	0.0020***	0.0000
	(0.0011)	(0.0001)	(0.0005)	(0.0000)
Domestic violence allegation	-0.0107**	-0.0005	-0.0010	0.0002
	(0.0044)	(0.0005)	(0.0022)	(0.0002)
Drug residence allegation	0.0094	0.0014	0.0553***	-0.0000
	(0.0104)	(0.0009)	(0.0076)	(0.0003)
Physical or substance abuse allegation	0.0123***	0.0008**	0.0189***	0.0004***
	(0.0029)	(0.0004)	(0.0014)	(0.0001)
Physical neglect allegation	0.0205***	0.0004	0.0270***	0.0002
	(0.0026)	(0.0004)	(0.0016)	(0.0002)
Improper supervision allegation	-0.0004	0.0004	0.0107***	0.0000
	(0.0026)	(0.0004)	(0.0015)	(0.0001)
Threatened harm allegation	0.0364***	0.0005	0.0266***	0.0004**
	(0.0034)	(0.0004)	(0.0016)	(0.0002)
Alleged perpetrator includes the parent/step-parent	0.0060	0.0014	0.0198***	0.0006
	(0.0078)	(0.0010)	(0.0057)	(0.0005)
Alleged perpetrator includes a non-parent relative	-0.0032	0.0001	0.0023	-0.0000
	(0.0043)	(0.0006)	(0.0029)	(0.0003)
Alleged perpetrator includes someone unrelated	0.0261***	0.0001	0.0169***	0.0001
	(0.0045)	(0.0005)	(0.0024)	(0.0002)
Age dummies	Y	Y	Y	Y
F-Statistic from Joint Test	18.903	1.222	35.902	1.242
P-Value from Joint Test	0.000	0.201	0.000	0.183
Observations	64,489	64,489	153,215	153,215

Notes. This table reports OLS estimates of regressions of the dependent variable on child/investigation characteristics. The regressions are estimated on the sample described in Table 1. Removal tendency is estimated using data from other cases assigned to the investigator. All regressions control for zip code by investigation year fixed effects. 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 and individual age dummy variables, which we exclude from the table for ease of exposition. Robust standard errors, two-way clustered at the child and investigator levels, are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: 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 described in Table 1. 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Bounds on Estimates of Mean Maltreatment Risk and UD

	(1) **From 0.03 Placement Rate**	**(2)** **From 0.02 Placement Rate**	**(3)** **From 0.01 Placement Rate**
	Panel A: Mean maltreatment risk by race		
Black children	[0.146***, 0.176***]	[0.148***, 0.168***]	[0.151***, 0.161***]
	(0.002,0.002)	(0.002,0.002)	(0.002,0.002)
White children	[0.167***, 0.197***]	[0.170***, 0.190***]	[0.173***, 0.183***]
	(0.001,0.001)	(0.001,0.001)	(0.001,0.001)
	Panel B: System-wide UD		
Mean across investigators	[0.014***, 0.020***]	[0.016***, 0.019***]	[0.016***, 0.018***]
	(0.002,0.002)	(0.002,0.002)	(0.002,0.002)
Number of investigators	699	699	699

Notes. Panel A reports bounds on race-specific average re-investigation risk, while Panel B reports bounds on system-wide (case-weighted) UD. To estimate bounds on mean risk, Column 1 uses a local linear fit of re-investigation rates among investigators who remove 3% of Black and white children. Columns 2 and 3 form bounds from investigators who remove 2% and 1% of 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 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 (500 times) and appear in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Estimates of Mean Maltreatment Risk by Race and UD (Alternative Outcomes)

	(1) Baseline	(2) Inv. within 5 months	(3) Inv. within 4 months	(4) Inv. within 3 months	(5) Inv. within 2 months	(6) Confirmed victim within 6 months	(7) Abuse inv. within 6 months	(8) Placed within 6 months
	Panel A: Mean maltreatment risk by race							
Black children	0.154***	0.131***	0.110***	0.086***	0.059***	0.039***	0.017***	0.004***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
White children	0.177***	0.153***	0.129***	0.099***	0.067***	0.049***	0.019***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
	Panel B: System-wide UD							
Mean across investigators	0.017***	0.016***	0.016***	0.016***	0.015***	0.015***	0.016***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of investigators	699	699	699	699	699	699	699	699

Notes. This table summarizes estimates of mean risk and UD for different outcome variables. Panel A reports estimates of race-specific average re-investigation risk, while Panel B reports estimates of system-wide (case-weighted) UD. Estimates in Panel A come from local linear extrapolations. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (500 times) and appear in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Estimates of Mean Risk by Race and UD (Covariate Adjustment)

	(1) Linear extrapolation	**(2) Quadratic extrapolation**	**(3) Local linear extrapolation**
	Panel A: Mean maltreatment risk by race		
Black children	0.140***	0.142***	0.140***
	(0.002)	(0.002)	(0.003)
White children	0.165***	0.153***	0.163***
	(0.002)	(0.002)	(0.003)
	Panel B: System-wide UD		
Mean across investigators	0.016***	0.016***	0.016***
	(0.001)	(0.001)	(0.001)
Number of investigators	699	699	699

Notes. The table summarizes estimates of mean risk and UD from different extrapolations of the variation in Figure 3 but where placement and re-investigation rates adjust for the covariates in Column 3 of Table 2 in addition to zip code by year fixed effects. Panel A reports estimates of race-specific average re-investigation risk. Panel B reports estimates of system-wide UD (weighted by the number of investigations assigned to each investigator). To estimate mean risk, Column 1 uses a linear extrapolation of the variation in Figure 3, while Column 2 uses a quadratic extrapolation and Column 3 uses a local linear extrapolation with a Gaussian kernel and a rule-of-thumb bandwidth. Robust standard errors, two-way clustered at the child and investigator level, are obtained by a bootstrapping procedure (500 times) and appear in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Heterogeneity in 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) Neglect**
	Panel A: Mean maltreatment risk by race						
Black children	0.154***	0.156***	0.150***	0.154***	0.148***	0.140***	0.155***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.002)
White children	0.177***	0.174***	0.176***	0.196***	0.155***	0.145***	0.182***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.002)
	Panel B: System-wide UD						
Mean across investigators	0.017***	0.013***	0.015***	0.018***	0.011***	0.010**	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)
Number of investigators	699	701	661	691	571	229	807

Notes. This table summarizes estimates of mean risk and 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 as "young" a child who is 7 years old or younger (the median age in our sample). Investigations coded as "neglect" are those that did not include a physical abuse allegation. Estimates come from a local linear extrapolation of the variation in Figure 3. However, unlike Figure 3, the extrapolations are done within the given characteristic. Panel A reports estimates of race-specific average re-investigation risk, while Panel B reports estimates of system-wide (case-weighted) UD. Robust standard errors, two-way clustered at the child and investigator levels, are obtained by a bootstrapping procedure (500 times) and appear in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: 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. Robust standard errors, two-way clustered at the child and the investigator level, are obtained by a bootstrapping procedure and appear in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Summary Statistics (Calls from 2017–2019)

	(1)	(2)	(3)	(4)	(5)	(6)
	All Hotline Calls			**Screened-in Calls**		
	All Children	**Black Children**	**White Children**	**All Children**	**Black Children**	**White Children**
	Panel A: Child characteristics					
White	0.647	0.000	1.000	0.627	0.000	1.000
Female	0.491	0.489	0.492	0.485	0.485	0.485
Age at investigation	8.407	7.879	8.695	7.930	7.442	8.220
	Panel B: Call characteristics					
Call included:						
Physical abuse	0.243	0.251	0.238	0.267	0.275	0.263
Physical neglect	0.305	0.341	0.285	0.330	0.365	0.308
Improper supervision	0.719	0.691	0.734	0.709	0.685	0.723
Threatened Harm	0.103	0.108	0.100	0.128	0.126	0.129
Reporter category:						
Education personnel	0.185	0.149	0.205	0.177	0.144	0.197
Law enforcement personnel	0.150	0.155	0.148	0.179	0.181	0.178
Family member	0.197	0.175	0.209	0.184	0.167	0.195
Medical personnel	0.123	0.153	0.107	0.129	0.153	0.115
Counselor/therapist	0.055	0.039	0.063	0.043	0.031	0.051
Other	0.290	0.329	0.268	0.287	0.324	0.265
	Panel C: Treatment					
Screen-in rate	0.596	0.630	0.578	1.000	1.000	1.000
Foster care placement rate	0.020	0.025	0.017	0.034	0.040	0.030
	Panel D: Outcomes, if not placed					
Re-investigated within 6 months	0.133	0.126	0.137	0.134	0.124	0.139
Number of calls	190,776	67,348	123,428	113,776	42,413	71,363

Notes. This table summarizes the analysis sample of hotline calls for alleged child maltreatment in Michigan from 2017–2019.

Table A10: First-Stage Effects of Screeners

	(1)	(2)	(3)	(4)
	Black children		**White children**	
Screen-in tendency	0.750***	0.758***	0.927***	0.938***
	(0.077)	(0.081)	(0.043)	(0.051)
F-statistic	95	88	467	339
Observations	67,267	67,267	123,404	123,404
Day-by-shift FE	Y	Y	Y	Y
Baseline controls	N	Y	N	Y
Age FE	N	Y	N	Y

Notes. This table reports OLS estimates of regressions of an indicator equal to one if the child was screened-in on the screener's tendencies to screen-in. All specifications include day $\times$ shift fixed effects. Screen-in tendency is estimated using data from other cases assigned to the screener. We 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 et al., 2019). Robust standard errors, clustered at the screener level, are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Balance Tests for Screener Assignment

	(1)	(2)	(3)	(4)
	Black children		**White children**	
Dependent variable:	Screened-in	Screen-in Tendency	Screened-in	Screen-in Tendency
Female	-0.007**	0.000	-0.015***	-0.000
	(0.004)	(0.000)	(0.003)	(0.000)
Physical neglect	0.070***	0.001	0.064***	0.001
	(0.007)	(0.001)	(0.006)	(0.001)
Improper supervision	-0.018**	0.001	-0.032***	0.000
	(0.007)	(0.001)	(0.005)	(0.001)
Threatened Harm	0.093***	-0.002	0.157***	-0.002
	(0.012)	(0.002)	(0.012)	(0.002)
Education personnel reporter	-0.046***	-0.000	-0.058***	-0.001
	(0.010)	(0.001)	(0.007)	(0.001)
Family member reporter	-0.087***	-0.001	-0.085***	-0.000
	(0.009)	(0.001)	(0.006)	(0.001)
Medical personnel reporter	-0.054***	-0.000	-0.028***	-0.000
	(0.011)	(0.001)	(0.008)	(0.001)
Counselor/Therapist reporter	-0.152***	0.001	-0.143***	-0.000
	(0.014)	(0.002)	(0.009)	(0.001)
Other reporter	-0.075***	-0.001	-0.054***	-0.001
	(0.008)	(0.001)	(0.006)	(0.001)
Age dummies	Y	Y	Y	Y
F-Statistic from Joint Test	40.987	1.299	99.894	1.430
P-Value from Joint Test	0.000	0.235	0.000	0.171
Observations	67,267	67,267	123,404	123,404

Notes. This table reports OLS estimates of regressions of the dependent variable on child/investigation characteristics. Screen-in tendency is estimated using data from other cases assigned to the screener. All regressions control for day $\times$ shift fixed effects. 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 and individual age dummy variables, which we exclude from the table for ease of exposition. Robust standard errors, clustered at the screener level, are reported in parentheses. The variables included in this table differ from those in Table A2 because the 2017–2019 subsample includes a slightly different set of variables. For example, it includes a smaller subset of allegation descriptions and it includes the reporter type of the initial allegation.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Bounds on $E[Y_{i1}^* - Y_{i0}^* | S_{i1} > S_{i0}]$

$E[(1-D_i)Y_{i1}^* \vert S_{i1} > S_{i0}]$	0.148***
	(0.017)
$E[D_i \vert S_{i1} > S_{i0}]$	0.022**
	(0.015)
$E[Y_{i1}^* \vert S_{i1} > S_{i0}] \in$	[0.148***, 0.170***]
	(0.017,0.019)
$E[Y_{i0}^* \vert S_{i1} > S_{i0}]$	0.154***
	(0.014)
$E[Y_{i1}^* - Y_{i0}^* \vert S_{i1} > S_{i0}] \in$	[-0.007, 0.016]
	(0.022,0.024)

Notes. The table summarizes estimates of each of the components needed to bound $E[Y_{i1}^* - Y_{i0}^* | S_{i1} > S_{i0}]$. Standard errors reported in parentheses are clustered at the screener level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Potential Drivers of Unwarranted Disparity

The main analysis shows unwarranted disparity in 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 et al., 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 UDs in foster care decisions. We follow the model and estimation approach in Arnold et al. (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 the corresponding moments 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 the benefit of leaving at home exceeds the posterior risk cost, yields the following decision rule:

$$D_{ij} = \mathbb{1}[\pi_{jR_i} \geq p_j(\nu_{ij}, R_i)] \tag{19}$$

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 in turn leads to UDs against the group with the lower subjective benefit, since the investigator generally makes different decisions for children with the same maltreatment potential Y_i^*.[45]

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 this 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 that, empirically, non-race characteristics are not a primary driver of UDs in our context.

To estimate the model, we first re-write Equation (19) 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:

$$\mu_{jr}(t) = E[Y_i^* \mid U_{ij} = t, R_i = r] \tag{20}$$

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 ($t \times 100$th) percentile of risk. $\Pi_{jr} = E[D_{ij} \mid R_i = r]$ parameterizes the leave-at-home rate of investigator j, and $\int_0^{\Pi_{jr}} \mu_{jr}(t)dt = E[Y_i^* \mid D_{ij} = 1, R_i = r]$.

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 et al., 2022, 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

[45] Inaccurate racial stereotyping tends to be observationally equivalent to racial animus, and can therefore similarly result in UDs (Arnold et al., 2018, 2022; Hull, 2021; Bohren et al., 2020). In particular, Arnold et al. (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.

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 et al. (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 et al. (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:

$$D_{ij} = \mathbb{1}[\pi_{jR_i} \geq p_j(\nu_{ij}, R_i)] = \mathbb{1}[\kappa_{jR_i} \geq Y_i^* + \eta_{ij}] \tag{21}$$

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 et al. (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 Figure 3. 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 the model-free analysis, we adjust for ZIP code by year fixed effects to ensure investigator assignment is as good as random.

Results

Table A8 reports SMD estimates of the race-specific moments: the mean misconduct risk, the first and second moments of marginal released 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 3, which range from 2.1 to 2.3 percentage points, depending on the extrapolation.

We also find that white children have higher mean marginal released outcomes relative to Black children, implying racial bias per the discussion above. Mean subsequent maltreatment risk is 0.93 (SE=0.018) for white children, compared to 0.875 (SE=0.019) for 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 3 are equal by race.[46] Because all three extrapolations are virtually identical, we focus on the linear extrapolation for this exercise. We fail to reject the null hypothesis that the slopes are equal by race ($p = 0.791$).

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.

C Nationwide Estimates From NCANDS Data

This section explores the generalizability of our key findings to other states. Section V showed that (i) Black children in Michigan are placed at higher rates than white children, conditional on maltreatment potential; and (ii) that this UD is primarily driven by cases with maltreatment potential. The second finding, in particular, adds nuance to ongoing policy discussions focused on the possibility that Black children are "over-placed" in foster care. A natural concern is to what extent these findings are unique to Michigan.

To explore generalizability, we use more limited data from the National Child Abuse and

[46] As discussed above, differences in signal quality by race would manifest in different slopes in Figure 3.

Neglect Data System Child files (NCANDS, 2023), sourced from the National Data Archive on Child Abuse and Neglect at Cornell University. This dataset, while not publicly available, can be accessed by researchers free of charge via a successful 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 2016, 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 UDs in each state. We instead leverage the fact that the non-parametric bounds in Figure 2 can be applied to state-level aggregate statistics in order to bound race-specific maltreatment potential.[47] Recall that a lower-bound (μ_{jr}^{L}) on μ_r can be derived by assuming that none of the children of race r who would be placed in foster care by investigator j (and for whom we cannot observe future maltreatment potential) would have had future maltreatment (equation 7). An upper bound (μ_{jr}^{U}) can be derived by assuming the opposite: that all children of race r who would have been placed in foster care by investigator j would have had future maltreatment (equation 8).

In the main analysis, we estimate average bounds among investigators with low placement rates, leveraging conditionally random investigator assignment (and using a linear adjustment to estimate investigator-specific placement and subsequent maltreatment rates to account for rotation fixed effects). We then apply these bounds on (μ_b, μ_w) to construct bounds on investigator-specific UDs, using equations (5) and (6) along with the investigator-specific estimates from equations (11) and (12). Finally, we estimate system-wide UD by case-weighted averages of the investigator-specific UDs.

Because the NCANDS data do not contain investigator identifiers or zip code information, we instead apply the non-parametric bounds to state-level aggregate statistics: the overall race-specific average placement rate, ϕ_r, and the overall average subsequent maltreatment rate (measured as a subsequent investigation within six months) among those not placed in foster care, ψ_r. Analogously to equations (7) and (8), these yield bounds on the mean risk

[47] We make the same general sample restrictions in the NCANDS data as in our primary dataset, 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.

parameters:

$$\mu_r^L \equiv (1-\phi_r)\psi_r = E[(1-D_i)Y_i^* \mid R_i = r] \leq E[Y_i^* \mid R_i = r] = \mu_r \tag{22}$$

and

$$\mu_r^U \equiv (1-\phi_r)\psi_r = 1 - E[(1-D_i)(1-Y_i^*) \mid R_i = r] \geq 1 - E[1-Y_i^* \mid R_i = r] = \mu_r, \tag{23}$$

where D_i again indicates placement. Applying these bounds directly to ϕ_r and ψ_r, analogously to equations (5) and (6), yields a range of estimates of system-wide UD for each state.

A subtlety to the interpretation of these aggregate-data 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 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 (Figure A2). But we are unable to directly test this in other states.[48]

Note that the width of the bounds of race-specific risk in a given state will be equal to the race-specific placement rate in that state. The bounds are therefore likely to be informative, since treatment rates tend to be low in CPS. For example, only five states in the NCANDS dataset have overall placement rates that are above 10% during this time period and only one state has a placement rate above 20% (Hawaii, at 21.5%).

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.[49] Figure A6 summarizes this comparison. The left-hand side of the figure shows non-parametric bounds using administrative data from Michigan, both for overall UD (as previously reported in Figure 2) 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, 0.024] while the range in the NCANDS data is [0.011, 0.014]. The UD range in cases without maltreatment potential is nearly identical across the two datasets, at [-0.04, 0.04], and is similar for cases with maltreatment potential: [-0.1,

[48]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., 2022b), and South Carolina (Roberts, 2019). But this is likely not the case in every CPS system.

[49]As mentioned above, because NCANDS data do not contain investigator identifiers, we are unable to restrict the NCANDS sample to cases assigned to investigators who were assigned to at least 200 cases (as in our main analysis).

0.3] in the MDHHS data and [-0.1, 0.2] in NCANDS data. The average estimate within the bounds in the MDHHS data is 0.4 percentage points in cases without maltreatment potential and 8 percentage points in cases with maltreatment potential. The average estimates within the bounds in the NCANDS data are 0.5 percentage points and 3.5 percentage points, respectively. Thus, the two datasets show that, in Michigan, overall UD in placement decisions is positive: for example, in NCANDS data, Black children are up to 1.4 percentage points more likely to be placed than white children, conditional on maltreatment potential. Moreover, the overall UD tends to be primarily driven by disparities in cases with maltreatment potential.

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 2016, 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 4. Section V.E discusses these results.

D Systems-Based Analysis: Sample and Assumptions

Data and Analysis Sample

Our primary analysis in Section VI is based on MDHHS data from January 2017–December 2019 consisting of the universe of hotline calls for child maltreatment in Michigan. The data include the details of each call, including the child's unique numeric identifier, race, and gender, as well as the report date, allegation types recorded by the screener, the child's relationship to the alleged perpetrator, and the reporter type. Conditional on the call being screened-in, the dataset includes the same variables as those in our primary analysis above (with the exception of investigators' names).

We construct our analysis sample as follows. We begin with the 681,090 unique hotline calls in Michigan between January 2017 and December 2019 that involved either white or Black children. We drop observations with a missing screener numeric identifier ($N = 13,885$). We then drop calls for which we cannot observe child welfare outcomes for at least six months after the focal call ($N = 122,656$), as this is the primary outcome of interest. 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 ($N = 15,909$), as these are handled by a non-random subset of investigators. We also drop screened-in cases with missing child ZIP code information ($N = 30,462$) since quasi-random assignment of investigators is conditional on geography, and we keep only the first call involving each child (which drops $N = 145,495$ calls) since repeat screened-in investigations tend to be assigned to the initial investigator.

The resulting analysis sample consists of 190,776 hotline calls, fielded by 162 unique screeners and shuttled to 642 unique investigators. Table A9 presents summary statistics for this sample. 65% of children in the sample are white and 49% are female. Common reporters of child maltreatment include education personnel (18.5%), law enforcement (15%), and family members (19.7%). Nearly 60% of calls were screened-in during this time period.

The foster care placement rate in the full population of calls is 2% (1.7% for white children and 2.5% for Black children). This incorporates a screening-in rate of 63% among Black cases and 58% for white cases; conditional on a call being screened-in, the placement rate is 3.4% (3% for white children and 4% for Black children). White children are more likely to experience subsequent maltreatment when left at home, both in the full set of calls and the subset of screened-in calls—though the difference in subsequent maltreatment rates is larger in the latter. In the full population of calls, 13.7% of white children experience subsequent maltreatment within six months of the focal call, compared to 12.6% of Black children. Among screened-in calls, the rates are 13.9% for white children and 12.4% for Black children.

Identifying Assumptions

Section VI leverages variation in the screening decisions of quasi-randomly assigned screeners who vary in their tendencies to screen-in calls. To characterize this variation, we construct a measure of screen-in tendencies as a leave-one-out rate among all calls in the sample that were quasi-randomly assigned to the screener. We regress an indicator for whether or not the call was screened-in on an exhaustive set of day by shift fixed effects and then calculate the leave-one-out average residual from this regression for each screener.

We next probe the identifying assumptions required for our approach. Figure A3 summarizes the distribution of the instrument. In a given day by shift, the range in the difference in screen-in rates across screeners is approximately 20 percentage points; the standard deviation is 3.4 percentage points. The figure also shows a strong, positive relationship between screener tendencies and the child's likelihood of being screened-in. Table A10 reports point estimates of an OLS regression of screening decisions on the instrument, separately by race. Columns 1 and 3 include only day by shift fixed effects, while Columns 2 and 4 add the baseline controls in Table A9. A one percentage point increase in the screener's tendency to screen-in increases the probability that the call is screened-in by 0.75 percentage points for Black children and

0.927 percentage points for white children. The F-statistic for this regression ranges from 88 to 467, indicating a strong relationship.

Table A11 probes an implication of screener quasi-random assignment: that observable child and call characteristics are uncorrelated with the screen-in tendencies of the assigned screener. Each column reports point estimates from an OLS regression of an indicator equal to one if the child was screened-in (Columns 1 and 3) and the screener's tendencies (Columns 2 and 4) on day by shift fixed effects and a range of child and call covariates. A rich set of characteristics are not predictive of the instrument, despite strongly predicting treatment ($p = 0.235$ for Black children and $p = 0.171$ for white children).

The exclusion restriction—that screener tendencies influence children's outcomes only through the screening decision—is plausible in this context. Screeners play no additional role in the investigation beyond the screening decision: if a call is screened-in, it is sent to the alleged victim's local child welfare office for formal investigation, and the investigation is then quasi-randomly assigned to an investigator. A screened-out call concludes MDHHS involvement. Moreover, the screener's role in the process is limited: according to MDHHS, calls typically last about 15 minutes.

One concern regarding the exclusion restriction is that maltreatment potential in the population of screened-out calls may not be similar to maltreatment potential in the population of screened-in calls that would not end up in foster care placement. To see this, note that subsequent maltreatment (Y_i) can be written:

$$Y_i = (1 - D_i)S_i Y_{i1}^* + (1 - S_i)Y_{i0}^* \tag{24}$$

where Y_i indicates foster care placement, Y_{i1}^* indicates foster care placement when the child is screened-in but not placed ($S_i = 1$ and $D_i = 0$), and Y_{i0}^* indicates foster care placement when the child is not screened-in ($S_i = 0$). When there are no investigation effects, i.e., when $Y_{i1}^* = Y_{i0}^* \equiv Y_i^*$, equation (24) reduces to:

$$Y_i = (1 - D_i S_i)Y_i^* \tag{25}$$

Thus, $Y_i = 0$ when the child is placed, but otherwise $Y_i = Y_i^*$ regardless of whether she is screened-out ($S_i = 0$) or screened-in but not placed ($D_i = 0$). The key assumption for the systems-based analysis is thus that equation (24) is a good approximation of equation (25); i.e., that direct effects of investigation $Y_{i1}^* - Y_{i0}^*$ are small on average.

To probe this assumption, we rely on the fact that our screener instrument allows us to bound the average of such effects for a set of marginal children. Consider a binary instrument Z_i which satisfies the typical IV independence, exclusion, and monotonicity assumptions at the screener stage. If we estimate a regression of Y_i on S_i, instrumenting for S_i via Z_i, we identify:

$$\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[S_i|Z_i=1]-E[S_i|Z_i=0]}=E[(1-D_i)Y^*_{i1}-Y^*_{i0}|S_{i1}>S_{i0}] \tag{26}$$

where S_{i1} and S_{i0} indicate potential screening statuses as a function of the instrument. When nobody is placed in foster care ($D_i = 0$), this expression captures the average effect of a screened-in investigation on subsequent maltreatment among compliers. Otherwise, we can bound this effect. To see this, first note that the average Y^*_{i0} for compliers is itself identified:

$$\frac{E[(1-S_i)Y_i|Z_i=1]-E[(1-S_i)Y_i|Z_i=0]}{E[1-S_i|Z_i=1]-E[1-S_i|Z_i=0]}=E[Y^*_{i0}|S_{i1}>S_{i0}] \tag{27}$$

Similarly:

$$\frac{E[S_iY_i|Z_i=1]-E[S_iY_i|Z_i=0]}{E[S_i|Z_i=1]-E[S_i|Z_i=0]}=E[(1-D_i)Y^*_{i1}|S_{i1}>S_{i0}] \tag{28}$$

These expressions can be used to bound $E[Y^*_{i1}|S_{i1}>S_{i0}]$, and therefore $E[Y^*_{i1}-Y^*_{i0}|S_{i1}>S_{i0}]$. To see this, observe that:

$$E[Y^*_{i1}|S_{i1}>S_{i0}]\in[E[(1-D_i)Y^*_{i1}|S_{i1}>S_{i0}],1-E[(1-D_i)(1-Y^*_{i1})|S_{i1}>S_{i0}]]. \tag{29}$$

The lower bound is directly given by equation (28), and the upper bound can be rewritten:

$$E[(1-D_i)(1-Y^*_{i1})|S_{i1}>S_{i0}]=1-E[D_i|S_{i1}>S_{i0}]-E[(1-D_i)Y^*_{i1}|S_{i1}>S_{i0}] \tag{30}$$

Finally, note that $E[D_i|S_{i1}>S_{i0}]$ is identified:

$$\frac{E[S_iD_i|Z_i=1]-E[S_iD_i|Z_i=0]}{E[S_i|Z_i=1]-E[S_i|Z_i=0]}=E[D_i|S_{i1}>S_{i0}]$$

We can thus use Z_i to obtain estimates of the lower and upper bounds of $E[Y^*_i|S_{i1}>S_{i0}]$ and then subtract $E[Y^*_0|S_{i1}>S_{i0}]$.

In practice, we use our continuous screener leniency instrument to form analogous bounds. Table A12 shows estimates of each component of the calculation. Bounds on average complier $Y^*_1-Y^*_0$ are tight and straddle zero. Specifically, we bound this average effect to be between -0.7 and 1.7 percentage points. The narrow bounds straddling zero provide confidence that the decomposition in Section VI is valid.