# Racial Discrimination in Child Protection*

E. Jason Baron†

Joseph J. Doyle, Jr.‡

Natalia Emanuel§

Peter Hull¶

Joseph Ryan‖

PRELIMINARY - PLEASE DO NOT CITE OR DISTRIBUTE

**Abstract**

Ten percent of Black children spend time in foster care—twice the rate of white children—with widespread concerns that this disparity is due to racial discrimination. We study the disparate impact of foster care placement decisions, as measured by racial disparities in placement rates among children with the same potential for future maltreatment. We account for the selective observability of future maltreatment potential by leveraging the quasi-random assignment of cases to investigators. Using administrative data from nearly 220,000 maltreatment investigations in Michigan between 2008 and 2016, we find that Black children are 1.7 percentage points (50%) more likely to be placed in foster care than white children with identical potential for future maltreatment. This result is robust to different measures of maltreatment potential and estimation strategies, and is not driven by observable case characteristics. Disparate impact is concentrated among cases with high risk of future maltreatment in the home, with Black children twice as likely to be placed in foster care compared to white children (12% vs. 6%). Disparate impact is significantly larger among investigators who are white, who see a lower share of cases involving Black children, who are more experienced, and who have relatively high placement rates.

*The findings do not necessarily represent the opinion of the Federal Reserve Bank of New York or the Federal Reserve System. All errors are the authors'.

†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 episodes of abuse and neglect and placing children in foster care when deemed necessary. In practice, CPS involvement is remarkably common: 37 percent of children in the U.S. experience a maltreatment investigation at some point in their childhood, and 5 percent spend time in foster care (Kim et al., 2017; Wildeman and Emanuel, 2014). Moreover, there are large racial disparities in CPS contact: the majority of Black children—53 percent—experience an investigation by the age of 18, compared to 28 percent of white children (Kim et al., 2017). Black children are also twice as likely to spend time in foster care: 10 percent, relative to 5 percent for white children (Wildeman and Emanuel, 2014).

There is enormous interest in these racial disparities, and the extent to which they reflect racial discrimination in the decisions of CPS investigators. This question has been studied for at least fifty years (Baron, Goldstein and Ryan, 2022; Billingsley and Giovannoni, 1972; Chibnall et al., 2003; Clifford and Silver-Greenberg, 2017; Drake et al., 2011; Font, Berger and Slack, 2012; Pryce et al., 2019; Reddy, Williams-Isom and Putnam-Hornstein, 2022), and has garnered recent popular and media attention.[1] The emphasis on these disparities reflects the fact that CPS decisions can have tremendous impact on the children involved and society writ large, and involve a difficult trade-off. Leaving children in high-risk situations may lead to subsequent maltreatment, which is associated with decreased educational attainment and future earnings (Currie and Spatz Widom, 2010), increased criminal activity (Currie and Tekin, 2012), and impaired physical and mental health (Lansford et al., 2006).[2] Yet, the decision to place a child in foster care is among the most far-reaching government interventions. Placement in foster care can have substantial effects on educational attainment, earnings, and criminal activity—both positive and negative, depending on the context (Bald et al., 2022*a*,*b*; Baron and Gross, 2022; Doyle, 2007, 2008; Gross and Baron, 2022).

Given the prevalence and high-stakes nature of investigators' contact with families, racial discrimination could lead to sub-optimal decisions and exacerbate racial inequalities in key long-term outcomes. Still, it is challenging to attribute racial disparities in CPS involvement to discrimination alone. A large literature explores whether Black children are more likely to experience CPS involvement, conditional on observable traits such as poverty (e.g.,

[1] See, for example, 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/ (November 23, 2022).

[2] Peterson, Florence and Klevens (2018) estimate the lifetime cost of new child abuse and neglect cases in the U.S. each year is as high as 10.9% of GDP.

Billingsley and Giovannoni, 1972; Chibnall et al., 2003; Courtney et al., 1996; Drake et al., 2011; Font, Berger and Slack, 2012; Paxson and Waldfogel, 1999, 2002; Putnam-Hornstein et al., 2013; Shaw et al., 2008; Wulczyn et al., 2013). However, such studies may suffer from omitted variable bias (OVB) if there are unobserved racial differences in risk. Conversely, they may suffer from included variable bias (IVB) if discrimination operates through the observable traits that an analyst conditions on.

This paper is the first to use a quasi-experimental design to address these potential biases and estimate the disparate impact of foster care placement decisions across white and Black families. Disparate impact is a broad legal term that identifies a practice as discriminatory if it leads to an adverse impact on a protected class without substantial legitimate justification, regardless of the decision-maker's intent. (Donohue, 2007). It incorporates, in particular, inequities from both racial bias (e.g. Becker, 1957) and statistical discrimination (e.g. Aigner and Cain (1977); Arrow (1973); Phelps (1972)) as well as systemic forms of discrimination that are often outside the scope of study in economics (e.g. Bohren, Hull and Imas, 2022).

An important question for such an analysis is how to define a "legitimate justification" for potential disparities in foster care. Following Arnold, Dobbie and Hull (2022), we interpret legitimate justification as objective potential outcomes, such as worker productivity in an employment context, profits in a lending context, or pretrial misconduct in bail decisions. Because the legal objective of CPS investigators is to allow most children to remain at home while minimizing the risk of subsequent maltreatment, a natural measure of disparate impact considers placement rates among white and Black children with the same potential for future maltreatment in the home. Our baseline analysis operationalizes this idea by defining subsequent maltreatment as being re-investigated for alleged child maltreatment within six months of the focal investigation, though we show robustness to alternative proxies (such as whether the child was re-investigated within shorter periods of time, whether the child had a substantiated investigation within six months, or whether the child was re-investigated for physical abuse within 6 months). While discrimination may arise at any stage of the CPS process, we focus on the decision to place a child in foster care as it is perhaps the farthest-reaching CPS tool, has large impacts on key adult outcomes, and is the basis for most of the current policy interventions aimed at addressing racial disparities.

The key identification challenge in measuring such discrimination is the selective observability of a child's potential for future maltreatment in the home. This potential outcome is revealed by subsequent maltreatment outcomes among children left at home, but not among children placed into foster care. We address this challenge by leveraging variation across quasi-randomly assigned investigators, who vary in their tendency to place children in

foster care. We use this strategy to estimate disparate impact from nearly 220,000 child maltreatment investigations in Michigan between 2008 and 2016.

To build intuition for our identification strategy, consider a setting with a quasi-randomly assigned investigator whose placement rate is near zero (i.e., who leaves nearly all children in the home). By virtue of quasi-random assignment, the subsequent maltreatment rates among Black and white children left at home by this investigator are close to the average subsequent maltreatment rates among all Black and white children. These race-specific maltreatment rates capture the correlation between future maltreatment potential and race; consequently, they can be used to correct for OVB in simple placement disparities that do not control for future maltreatment potential. In practice, we estimate these average maltreatment rates by extrapolating observed subsequent maltreatment rates across quasi-randomly assigned investigators with low placement rates. Because the placement rate in our setting is low (roughly 3%), there is a large set of investigators with very low placement rates—allowing us to also construct informative bounds on the average maltreatment rates without any extrapolation. This strategy builds on prior work by Arnold, Dobbie and Hull (2022), who use quasi-random bail judge assignment to measure disparate impact in pretrial release decisions, and a broader literature on "identification at infinity" in sample selection models (Andrews and Schafgans, 1998; Chamberlain, 1986; Heckman, 1990).

Applying this strategy, we find that Black children in Michigan are 1.7 percentage points more likely to be placed in foster care than white children with identical potential for future maltreatment. This conditional disparity represents 50% of the baseline foster care placement rate (at 3.4 percentage points), and is 70% larger than the racial disparity in placement rates from an observational analysis that conditions on child and investigation traits (at 1.0 percentage points). Observed disparities *understate* the level of disparate impact because of negative OVB: we find that Black children are around 2 percentage points less likely to see future maltreatment when left at home, relative to white children.

These results are robust to alternative time frames for re-investigation, our interpretation of investigators' legal objective, and extrapolation methods. Non-parametric bounds show tight estimates of disparate impact, of around 1.4 and 1.8 percentage points. Interestingly, adjusting for child and investigation traits leads to virtually identical estimates of disparate impact—suggesting non-race characteristics play a limited role in the disparate impact.

Our baseline measure of disparate impact is a weighted average of two conditional disparities: placement disparities among the population with potential for subsequent maltreatment, and placement disparities among children without potential for subsequent maltreatment.

In other words, the average 1.7 percentage point disparity we find could arise primarily from racial differences in "true positives" (i.e. higher foster care placement rates among Black children at high risk of future maltreatment, relative to similarly risky white children) or racial differences in "false positives" (i.e. higher foster care placement rates among Black children at low risk of future maltreatment, relative to similarly risky white children). Disentangling these scenarios is important, as they could suggest different policy responses.

We find that disparate impact in foster care placement is concentrated among children with subsequent maltreatment potential in the home. Our preferred quasi-experimental estimates show that in this subpopulation the placement rate of Black children is more than twice the placement rate of white children (12% versus 6%). In contrast, the placement disparity among children without subsequent maltreatment potential is small and statistically insignificant. These findings add nuance to ongoing policy debates over reform to child welfare systems, which often focus on the possibility that Black children are "over-placed" into foster care relative to white children. Our estimates suggest disparate impact in placement decisions can arise primarily from the relative "under-placement" of white vs. Black children in potentially high-risk situations.

We further decompose the overall level of discrimination by studying heterogeneity across investigators. We find that disparate impact is significantly higher for white investigators, more experienced investigators, investigators who handle a relatively low share of cases involving Black children, and investigators with relatively higher placement rates. Policies targeting these investigators may thus be most effective at reducing disparate impact.

These findings contribute to a growing literature in economics examining the effectiveness of foster care systems. Much of this literature leverages as-good-as-random investigator assignment to estimate the causal effects of foster care placement on children (e.g. Doyle (2007), Doyle (2008), Doyle (2013), Gross and Baron (2022), Baron and Gross (2022), and Bald et al. (2022*a*)) and parents (e.g. Grimon (2020) and Baron and Gross (2022)).[3] Ours is the first quasi-experimental study of disparate impact in foster care placement decisions.

Methodologically, we add to a growing literature leveraging the quasi-random assignment of decision-makers to estimate bias and discrimination in different decisions, such as pretrial release (e.g. Arnold, Dobbie and Yang (2018), Arnold, Dobbie and Hull (2022), and Hull (2021)), traffic stops (e.g. Feigenberg and Miller (2022)) and lending (e.g. Dobbie et al. (2021)). Our analysis of foster care decisions features a large number of decision-makers with very low placement rates, allowing for both precise non-parametric inferences on

[3]See Bald et al. (2022*b*) for a richer summary of this literature.

overall disparate impact and the statistical power to distinguish discrimination among high- or low-risk home situations. Our data also includes rich information on decision-maker characteristics, allowing a more detailed study of heterogeneity in disparate impact than in other settings.

The remainder of the paper is organized as follows. Section II describes the CPS setting and summarizes the doctrine of disparate impact. Section III describes our analysis sample and presents motivating results. Section IV develops our disparate impact measures and quasi-experimental approach. Section V presents our main findings, while Section VI presents findings on heterogeneity and potential mechanisms. Section VII concludes and discusses possible extensions for future drafts.

# II Setting

## II.A Child Protective Services and Foster Care Placement

The CPS system aims to protect children from maltreatment in their home environment; Figure 1 summarizes the CPS 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., food deprivation). Anyone can make a report to the hotline, though the most common reports are from education and law enforcement personnel.

Calls to the hotline are answered by screeners who assess whether the allegation of maltreatment conforms with state law and with guidance from Michigan's Department of Health and Human Services. 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 an exact day by shift, incoming calls are randomly assigned to screeners, with no exceptions. Incoming calls enter a queue, and the hotline system routs 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 play no other role in the investigation: 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.

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. Below, 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 crucial decisions that determine the intensity of child welfare's involvement. 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.

Second, if the investigation is substantiated, the investigator makes a judgement on how much risk the child may be in by continuing to live 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 keep the child with their family otherwise. Investigators have immense discretion over foster care placement. While there is technically a standardized system with a 22-question risk assessment form that helps determine whether placement is appropriate, many of the questions are inherently subjective and investigators often manipulate responses to match their priors (Bosk, 2015; Gillingham and Humphreys, 2010).

If the investigator determines that the child is in imminent risk of maltreatment, she requests that her supervisor submit a court petition to remove the child from the home. Anecdotally, it is rare for either the supervisor or the judge to disagree with the investigator's recommendation. Regardless of the decision to place a child in foster care, investigators can recommend preventative services to support the family. These services range from food banks to substance abuse or parenting classes. Previous work has found minimal impact of these decisions on a child's later outcomes (Baron and Gross, 2022; Gross and Baron, 2022).

Around three percent of cases result in a child being placed in foster case. In these cases the child is placed with either an unrelated foster family, relatives, or in a group home, while their custodial parents receive services to support family reunification. Children spend around 17

months in foster care on average, after which most are reunified with their parents.

## II.B Disparate Impact

We study the disparate impact of foster care placement decisions, as measured by disparities in placement rates among children with the same potential for future maltreatment. Disparate impact is one of two main legal doctrines of discrimination in United States case law, concerning the discriminatory effects of a policy or practice (rather than a decision-maker's intent).[4] Under the doctrine, a policy or practice is discriminatory if it leads to an adverse impact on a protected class and either the decision-maker cannot offer a substantial legitimate justification or if it can be shown that such a justification could be reasonably achieved by less disparate means. For example, in the landmark US Supreme Court Case of *Griggs v. Duke Power Co.*, disparate impact was found in an employment policy at the Duke Power Company which required employees to have a high school diploma to be considered for promotion. The court found that having a high school diploma had little to no relationship to a worker's productivity (the "legitimate justification" in this setting), yet drastically limited the eligibility of Black employees for employment.

Under investigator guidelines, the legitimate justification for foster care placement is a potential for subsequent maltreatment.[5] An ideal measure of disparate impact in this setting would thus compare the placement rates of different protected groups with identical potential for future maltreatment. The finding of such conditional disparities would likely be necessary (but perhaps not sufficient) evidence to win a disparate impact case—depending on, for example, whether or not it can be shown there exist alternative placement policies that yield less of a conditional disparity while achieving similar or better outcomes.[6] In Section IV, we formalize such a measure and discuss its identification.

[4]The other main doctrine of discrimination, disparate treatment, contrasts with disparate impact by prohibiting policies or practices motivated by a "discriminatory purpose" and thus generally requiring proof of intent. See Section I.A of Arnold, Dobbie and Hull (2022) for more background on the two doctrines and discussion of relevant case law.

[5]For example, according to Michigan's Department of Health and Human Services' *Children's Protective Services Policy Manuals*, "placement of children out of their homes should occur only if their well-being cannot be safeguarded with their families" (p.3). The manual further instructs investigators to "file a petition with the Family Division of Circuit Court 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).

[6]The disparate impact standard only applies in certain contexts, since it stems from statutory rules rather than constitutional law. Examples include employment via Title VII of the 1964 Civil Rights Act and housing via the Fair Housing Act of 1968. The disparate impact standard may also apply to all programs and activities receiving federal financial assistance via Title VI of the 1964 Civil Rights Act, which includes the child welfare agencies considered in our analysis (*United States v. Maricopa County*).

Importantly, disparate impact can arise from both classic forms of "direct" discrimination historically studied in economics, and from indirect or "systemic" discrimination traditionally considered by other fields (Bohren, Hull and Imas, 2022). The former channel includes direct discrimination from biased preferences and beliefs (e.g. Becker, 1957), as well as statistical discrimination (e.g. Aigner and Cain (1977); Arrow (1973); Phelps (1972)). The latter channel arises from non-race characteristics that embed discrimination from the past or other domains. We explore the role of non-race characteristics in some of our analyses below.

# III Data

## III.A Data Sources and Analysis Sample

Our primary analysis is based on data from the Michigan Department of Health and Human Services (MDHHS), consisting of the universe of maltreatment investigations in Michigan between January 2008 and June 2017. The data include details of each screened-in investigation, such as the allegation report date, allegation types as coded by the investigator, 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. We use the investigators' names to predict their race, ethnicity and gender.[7]

MDHHS data also contain the numeric identifier of the screener who screened-in the initial call, the exact time and date of the call, and data on all screened-out calls. This dataset could allow for an examination of disparate impact at two distinct points of CPS involvement—the screening decision and the foster care placement determination—as well as the degree to which the two may interact in a systemic way. While we do not present any results from this analysis in the current draft, we plan to incorporate it in future versions.

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 (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

[7]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 that a given first name is of a particular gender.

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.[8] 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. Panel A of Table 1 summarizes this sample. Overall, 70 percent of children in our sample are white and 48 percent are female. The average child is nearly seven years old. A large share of children—46 percent—have experienced an investigation prior to the focal one with around one previous investigation on average.

Panel B of Table 1 shows that 43 percent of investigations include a physical or substance abuse allegation and 10 percent of investigations include a domestic violence allegation. These rates are slightly higher for white children than Black children, while the rate of drug residence allegations (around 3 percent, on average) is slightly higher for Black children than white children.[9] Around 91 percent of alleged perpetrators are a parent or stepparent, with 5 percent of alleged perpetrators being a non-parent relative.

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, 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 percent of non-placed children are re-investigated for child maltreatment. This rate rises steadily with time, with around 17 percent 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 outcomes is consistent with investigator manuals instructing workers to place a child in foster care if they believe the child is in "imminent risk" and no other prevention services will protect the child from subsequent maltreatment in the home.[10]

Our primary maltreatment measure considers whether the child was re-investigated, which

[8] Our final sample includes 699 investigators, each assigned to 312 cases and 6,644 rotations on average.

[9] An investigation can include numerous allegations, so the listed categories are not mutually exclusive.

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

entails both a report to CPS and a decision by a hotline screener to begin an investigation. As mentioned above, only around 60% of all hotline calls are assigned for investigation. 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, or whether the child had an investigation for physical abuse (as opposed to neglect). Importantly, a subsequent substantiated investigation is not an ideal outcome, since repeat cases within a short period of time could be assigned to the same investigator, and thus the outcome may itself rely on decisions made by the initial investigator. In contrast, neither the decision to report nor to screen-in a case—both of which are necessary for a re-investigation—involve the prior investigator.[11] Our preferred estimates can be seen to isolate policy-relevant measures of disparate impact, 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" in the rest of the paper.

Panel D of Table 1 shows that white children are more likely to see subsequent maltreatment when left at home. Within six months of the focal investigation, 17% of white children see subsequent maltreatment, compared to 14% of Black children. Taken at face value, this outcome disparity may suggest white children have a higher level of "need" for foster care placement—and therefore that 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 virtually no 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 child and investigator characteristics. Specifically, we estimate regressions of the form:

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

[11] Of course, as in the bail setting where pretrial misconduct may be the result of discrimination in other parts of the criminal justice system, a re-investigation may be due to discrimination in other parts of the child protection system.

by ordinary least squares (OLS), 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 $\beta$ thus capture placement disparities among white and Black children in observably similar cases.

Column 1 of Table 2 presents estimates from a simple bivariate regression of $D_i$ on $B_i$; 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 fixed effects. Column 3 further adds controls for the child characteristics and focal investigation characteristics in Panels A and B of Table 1. The disparity in this specification shrinks slightly, to 1 percentage point.

A significant disparity in placement rates thus remains among observably similar children and cases. At 29 percent of the overall placement rate, this 1 percentage point disparity is economically meaningful. But the implications of this controlled disparity for disparate impact are at this point unclear: we cannot adjust in 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 omitted variables bias (OVB) and either over- or under-state the true level of disparate impact 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 disparate impact. We next discuss the quasi-experimental variation we will use to overcome both challenges.

## III.C Quasi-Experimental Variation

We leverage variation in the placement decisions of quasi-randomly assigned investigators in Michigan, who vary in their tendency to place children in foster care. To characterize this variation, 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 (Bald et al., 2022*a*; Baron and Gross, 2022; Doyle, 2007, 2008; Gross and Baron, 2022). To construct this measure, we first regress an indicator for whether the child was placed in care following the investigation on an exhaustive set of rotation fixed effects, the level at which investigators are quasi-randomly assigned to cases.[12] We then calculate the leave-one-out average residual from this regression for each investigator.

[12]Residualizing ensures that the leave-one-out measure captures the tendency of an investigator relative to other investigators in the local office at a given point in time.

Figure A1 summarizes the leave-one-out placement rate 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 also shows a strong, positive relationship between investigators' placement tendencies and the child's likelihood of being placed in foster care. Table A1 reports point estimates of an OLS regression of placement decisions on the 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 0.55 percentage points for Black children, and 0.6 percentage points for white children. The F-statistic for this regression ranges from 131 to 257, indicating a strong relationship.

Table A2 probes an implication of quasi-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 a host of child and investigation characteristics, age fixed effects, 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 despite being highly predictive of the probability of being placed ($p = 0.138$ for Black children and $p = 0.231$ for white children).

# IV Methods

## IV.A Disparate Impact Measure

We formalize the setting by considering a population of 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 in 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 racial disparities in these potential foster care placement decisions, leaving the process by which children get matched to investigators unspecified for now.

Specifically, we measure disparate impact in this setting by the average placement rate disparity between Black and white children with identical potential for future maltreatment. To build up to up this measure, we first define, for each investigator $j$, the placement rate disparity among Black and white children without future maltreatment potential:

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

Similarly, we define the placement rate disparity among Black and white children with future maltreatment potential:

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

Each $\Delta_{jy}$ parameter can be understood as capturing racial differences in the tendency of investigator $j$ to correctly and incorrectly classify children by their future maltreatment potential, with $\Delta_{j0}$ capturing disparities in "false positives" and $\Delta_{j1}$ capturing disparities in "true positives." We measure overall disparate impact by averaging these measures:

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

with weights given by the average future maltreatment risk in the population, $\bar{\mu} = E[Y_i^*]$. This weighted average captures the expected level of discrimination when investigator $j$ encounters a representative pool of children, with unknown future maltreatment potential. We measure the system-wide level of discrimination by the case-weighted average of $\Delta_j$ across investigators $j = 1, \ldots, J$.

We say that investigator $j$ discriminates against Black children when $\Delta_j > 0$, that she discriminates against White children when $\Delta_j < 0$, and that she does not discriminate against either group when $\Delta_j = 0$. By holding the potential population of assigned children fixed, estimates of $\Delta_j$ can be used to calculate both the average level of disparate impact in Michigan as well as variation in the level of disparate impact across investigators.

In some analyses we separately study the two components of disparate impact, $\Delta_{j0}$ and $\Delta_{j1}$, as this can lead to a more nuanced analysis of disparate impact. A finding of $\Delta_{j0} > 0$, for example, would show that Black children without future maltreatment potential are over-placed relative to white children without future maltreatment potential, while a finding of $\Delta_{j1} > 0$ could be interpreted as under-placement of white children with future maltreatment potential relative to Black children with future maltreatment potential.

The fundamental challenge in estimating $\Delta_j$ (and its components) is the selective

observability of future maltreatment potential. Among children who are not placed in foster care, $Y_i^*$ is directly observed by future maltreatment outcomes $Y_i$. But future maltreatment outcomes are censored among children who are placed in foster care, with $Y_i = 0$ by construction for virtually all children.[13] Thus, we cannot directly estimate equation (4).

Following Arnold, Dobbie and Hull (2022), we address this selection challenge by first re-writing the components of equations (2) and (3) in terms of a set of directly estimable moments and two unknown parameters: the average future maltreatment risk 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, we have for each race $r$:

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

and

$$E[D_{ij} \mid R_i = r, Y_i^* = 0] = 1 - E[1 - D_{ij} \mid R_i = r, Y_i^* = 0] = 1 - \frac{E[(1-D_{ij})(1-Y_i^*) | R_i = r]}{1-\mu_r}. \quad (6)$$

Since future maltreatment outcomes $Y_i$ directly reveal future maltreatment potential $Y_i^*$ among children not placed in foster care (i.e., those with $(1 - D_{ij}) = 1$), the numerators of equations (5) and (6) are not affected by selection:

$$E[(1 - D_{ij})Y_i^* \mid R_i] = E[(1 - D_{ij})Y_i \mid R_i] \quad (7)$$

and

$$E[(1 - D_{ij})(1 - Y_i^*) \mid R_i] = E[(1 - D_{ij})(1 - Y_i) \mid R_i] \quad (8)$$

When investigators are randomly assigned, these expressions are identified.[14] Thus, in quasi-experimental data the challenge of identifying $\Delta_j$ reduces to the challenge of identifying the two mean risk parameters, $\mu_b$ and $\mu_w$, which are the remaining moments in equations (5) and (6).

[13] Since the average stay in foster care in our data is 17 months long and virtually no children return home within six months, it is usually impossible to see if a child who has been placed in foster care is maltreated at home within six months.

[14] Formally, $E[(1-D_{ij})Y_i \mid R_i] = E[(1-D_i)Y_i \mid Z_{ij} = 1, R_i]$ and $E[(1-D_{ij})(1-Y_i) \mid R_i] = E[(1-D_i)(1-Y_i) \mid Z_{ij} = 1, R_i]$ where $Z_{ij}$ indicates the assignment of case $i$ to investigator $j$ and $D_i = \sum_j D_{ij}Z_{ij}$ is case $i$'s foster care placement status. Here we assume simple random assignment, such that $Z_{ij}$ is independent of $(D_{ij}, Y_i^*, R_i)$. We discuss how we adapt this approach to conditional random assignment below.

## IV.B Identification Strategies

We consider two strategies for estimating the two key mean risk parameters, $\mu_b$ and $\mu_w$, each of which yields estimates of disparate impact $\Delta_j$.

**Bounding with Observed Subsequent Maltreatment.** The first strategy bounds each $\mu_r$ by the observed subsequent maltreatment rate of race $r \in \{b, w\}$. To formalize this approach, let $D_i \in \{0, 1\}$ indicate realized placement into foster care. Since $Y_i^*$ is binary, and since $Y_i^* = Y_i$ when $D_i = 0$ (such that $(1 - D_i)Y_i = (1 - D_i)Y_i^*$), a lower bound on each mean risk parameter is given by:

$$\mu_r^L \equiv E[(1 - D_i)Y_i \mid R_i = r] = E[(1 - D_i)Y_i^* \mid R_i = r] \leq E[Y_i^* \mid R_i = r] = \mu_r. \tag{9}$$

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

$$\begin{aligned} \mu_r^U \equiv 1 - E[(1 - D_i)(1 - Y_i) \mid R_i = r] &= 1 - E[(1 - D_i)(1 - Y_i^*) \mid R_i = r] \\ &\geq 1 - E[1 - Y_i^* \mid R_i = r] = \mu_r. \end{aligned} \tag{10}$$

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

The $[\mu_r^L, \mu_r^U]$ bounds can be estimated by the overall rates of foster care placement and future maltreatment by race, absent any further assumptions. The width of the bounds for each $\mu_r$ is simply the placement rate of race $r$,

$$\mu_r^U - \mu_r^L = 1 - E[(1 - D_i)(1 - Y_i) \mid R_i = r] - E[(1 - D_i)Y_i \mid R_i = r] = E[D_i \mid R_i = r]. \tag{11}$$

Since placement in foster care is relatively rare, $E[D_i \mid R_i = r]$ is small such that the 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 Observed Variation Across Investigators.** Our second strategy for estimating the two mean risk parameters follows Arnold, Dobbie and Hull (2022) by extrapolating observed variation in future maltreatment rates across as-good-as-randomly assigned investigators. To build intuition for this approach, suppose investigators are completely randomly assigned, as in a randomized trial, and consider the possibility of an

investigator $j^*$ who places virtually no children of either race in foster care: $Pr(D_{ij^*} = 1 \mid R_i) \approx 0$. This investigator's race-specific future maltreatment rate would be close to the overall future maltreatment risk by race:

$$E[Y_i^* \mid D_{ij^*} = 0, R_i = r] \approx E[Y^* \mid R_i = r] = \mu_r \tag{12}$$

Moreover, by random assignment, this investigator $j^*$'s race-specific placement rate $Pr(D_{ij^*} = 1 \mid R_i)$ and future maltreatment rate $E[Y_i^* \mid D_{ij^*} = 0, R_i = r]$ would be identified by the placement rate and maltreatment rates among the children assigned to the investigator.

In the absence of such an investigator, the key mean risk parameters can be estimated using statistical extrapolations of placement and 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 0. Mean risk estimates may, for example, come from the vertical intercept at zero of linear, quadratic, or local linear regressions of estimated maltreatment rates $E[Y_i^* \mid D_{ij} = 0, R_i = r]$ on estimated placement rates $E[D_{ij} \mid R_i = r]$ across investigators $j$ within each race $r$. Such an approach builds on a long literature on "identification at infinity" in sample selection models (Andrews and Schafgans, 1998; Chamberlain, 1986; d'Haultfoeuille and Maurel, 2013; Heckman, 1990) and a recent literature that estimates average treatment effects (ATEs) and marginal treatment effects (MTEs) with multiple discrete instruments (Brinch, Mogstad and Wiswall, 2017; Hull, 2020; Kowalski, 2016; Mogstad, Santos and Torgovitsky, 2018).

In practice, we adopt both of these intuitive strategies while accounting for the fact that investigators in Michigan are only as-good-as-randomly assigned within rotation fixed effects. We first use linear regression to estimate each investigator's race-specific foster care placement rate $E[D_{ij} \mid R_i = r]$ and future maltreatment rate $E[Y_i^* \mid D_{ij} = 0, R_i = r]$ accounting for these strata. Specifically, we estimate regressions of the form

$$D_i = \sum_j \alpha_j (1 - B_i) Z_{ij} + \sum_j \beta_j B_j Z_{ij} + X_i' \gamma + \varepsilon_i \tag{13}$$

among all cases, and

$$Y_i = \sum_j \phi_j (1 - B_i) Z_{ij} + \sum_j \psi_j B_j Z_{ij} + X_i' \lambda + \nu_i \tag{14}$$

among cases where the child is not placed in foster care. Here again $D_i$ indicates foster care placement, $Y_i$ indicates future maltreatment, and $B_i$ indicates a case involving a Black child. The $Z_{ij} \in \{0, 1\}$ 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 $\alpha_j$ and $\beta_j$ coefficients capture strata-adjusted placement rates of investigator $j$ among white and Black children, respectively. Similarly, the $\phi_j$ and $\psi_j$ coefficients capture strata-adjusted maltreatment rates among the white and Black children left at home by investigator $j$.[15] We use estimated coefficients to complete the formulas for $\Delta_{j0}$, $\Delta_{j1}$, and $\Delta_j$, given bounds or estimates of the mean risk parameters.

## IV.C Identifying Assumptions

Both of our identification strategies leverage the as-good-as-random assignment of investigators to cases, which we discuss and empirically validate in Section III above. Implicitly, our framework also imposes an exclusion restriction: that investigators' placement tendencies only impact subsequent outcomes through the decision to place in 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 the exclusion restriction is reasonable in this context.

One possible concern, for example, is that investigators could influence children's experiences while in foster care, which would violate the exclusion restriction. A similar concern arises when estimating the causal impact of incarceration because the judge who decides whether to incarcerate also plays a role in the sentencing decision. However, unlike the criminal justice setting, investigators do not remain in contact with children while they are in foster care. If 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 specific concern regarding the exclusion restriction in our setting is investigator

[15] Regression adjustment is appropriate when placement and re-investigation rates are linear in the rotation strata for each judge and race, with constant coefficients. See Section IV.A of Arnold, Dobbie and Hull (2022).

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, Baron and Gross (2022) show that, in Michigan, the effects of these other decisions on children's outcomes are small and statistically insignificant, and that our instrument operates primarily through the foster care decision.

Importantly, although our approach can be understood as leveraging quasi-random investigator assignment as an instrument for placement, it does not require a conventional assumption of first-stage monotonicity: i.e., that investigators have a common ranking of cases by their appropriateness for foster care placement.[16] Intuitively, the non-parametric bounds on $\mu_r$ do not come from comparisons across investigators, while the extrapolated estimates of $\mu_r$ 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).[17] In 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 Main Estimates

## V.A Discrimination in Foster Care Placement Decisions

**Bounding with Observed Subsequent Maltreatment.** We first compute the range of possible mean re-investigation risk parameters, given the average re-investigation rates in our analysis sample. Applying equations (9) and (10) to the estimates in Table 1, we estimate Black mean risk bounds of $\mu_b \in [0.143, 0.186]$ and white mean risk bounds of $\mu_w \in [0.168, 0.199]$. We then estimate the range of system-wide disparate impact given all combinations of $(\mu_b, \mu_w)$ in these bounds, using equations (5) and (6) along with the

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

[17] 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 $(\kappa_j, \lambda_j)$ 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 violates conventional first-stage monotonicity, since investigators differ both in their ordering of individuals by the appropriateness of placement ($\nu_{ij}$) and their relative skill at predicting subsequent maltreatment ($\lambda_j$). Nevertheless, when $E[\lambda_j|\kappa_j]$ is constant (linear) in $\kappa_j$, average future maltreatment rates are linear (quadratic) in placement rates, such that simple parametric extrapolations identify mean risk $\mu$.

investigator-specific estimates from equations (13) and (14).

Figure 2 shows robust evidence of system-wide disparate impact in foster care placement decisions. The range of possible system-wide disparate impact, as measured in the case-weighted average $\Delta_j$, is estimated to be from around 1.2 percentage points to around 2.4 percentage points. Most of this range is above the observed racial disparity in placement rates (around 1.2 percentage points). Intuitively a positive measure summarises that the range of white mean risk is overall higher than the range of Black mean risk, consistent with the selected outcomes 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 disparate impact measure should increase the estimated disparity.[18]

Tighter bounds on system-wide disparate impact can be obtained by using the placement and future misconduct rates of investigators with low placement rates, as shown in equation (11). Column 1 of Appendix Table A3 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 bounds lead to a tighter bound on system-wide disparate impact, of $[0.014, 0.020]$. Columns 2 and 3 show the bounds are tightened further as we focus on investigators with estimated (strata-adjusted) placement rates below 0.02 and 0.01. Our tightest bounds of system-wide disparate impact are precisely estimated as $[0.016, 0.018]$.

**Extrapolating Observed Variation Across Investigators.** To obtain point estimates of disparate impact, we extrapolate variation in the race-specific maltreatment and placement rates across the quasi-randomly assigned investigators. Figure 3 plots this variation, with a binned scatterplot of the estimates from equations (13) and (14).

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. The

[18] 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, Dobbie and Hull (2022)). 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 ($\mu_w > \mu_b$).

strong negative relationship between investigators' placement tendencies and subsequent maltreatment is found with three specifications: linear, quadratic, and local linear.[19] 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. All three extrapolations yield very similar estimates. As depicted by the black lines in Figure 2, our estimates fall on the low end of our earlier mean risk 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 disparate impact in Michigan.

Panel B of Table 3 reports corresponding estimates of system-wide disparate impact, constructed similarly as before. All three extrapolations yield virtually identical estimates of 0.017 (SE=0.2), or 1.7 percentage points. This represents a disparity of roughly half of the overall placement rate of 3.4 percentage points.

## V.B Robustness

We consider 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.[20] We consider whether the child was investigated within two, three, four or five months (Table A4, Columns 2 through 5).[21] Our estimates of system-wide disparate impact are similar, regardless of the time frame.

We also examine robustness to our interpretation of investigators' legal objective. If the legal objective is misspecified, our estimates may suffer from "omitted payoff bias" (Kleinberg et al., 2018). We have assumed that investigators try to minimize the probability of a subsequent investigation. When we consider that investigators may seek to minimize the likelihood of a subsequent *substantiated* investigation or seek to minimize the probability of a subsequent investigation that includes physical abuse, we find that estimates of system-wide

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

[20] As noted in Section II.A, 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.

[21] 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.

disparate impact are slightly smaller but similar to our baseline estimates (Table A4, Columns 6 and 7).

# VI Heterogeneity and Potential Mechanisms

## VI.A The Role of Non-Race Characteristics

Estimates of disparate impact can arise from direct discrimination—through the conscious or unconscious use of the child/family race—as well as "indirect" or "systemic" discrimination—that is, via conscious or unconscious use of non-race characteristics that may be correlated with race (Arnold, Dobbie and Hull, 2022; Bohren, Hull and Imas, 2022). We consider whether accounting for child-specific traits (such as age and gender and prior investigations) as well as investigation-specific traits (such as the nature of the allegations and relationship to the alleged perpetrator) changes our estimates. Specifically, we adjust the estimated investigator-specific placement and subsequent maltreatment rates by the child and investigation characteristics in Column 3 of Table 2.

Adjusting for child- and investigation-specific traits leads to virtually identical estimates of disparate impact in Table A5 relative to our baseline estimates in Table 3 (1.5 and 1.7 percentage points, respectively). This similarity suggests that non-race characteristics play a limited role in our context.

Indeed, our estimates of disparate impact are positive across various child- and investigation subgroups, as shown in Table 4. While the estimates are more positive for female children (relative to male children), young children (relative to older children), and investigations involving neglect (as opposed to physical abuse), all estimates are positive and statistically significant. Together, this reinforces that racial disparities show up in all types of cases.

## VI.B Disparate Impact Decomposition

Our baseline measure of disparate impact, $\Delta_j$, is a weighted average of two conditional disparities: $\Delta_{j1}$—placement disparities among the population with potential for subsequent maltreatment ($Y_i^* = 1$)—and $\Delta_{j0}$—placement disparities among children without potential for subsequent maltreatment ($Y_i^* = 0$). Thus, while estimates of $\Delta_j$ are informative and policy-relevant on their own, if disparate impact is concentrated in either subpopulation, our baseline estimates may mask policy-relevant heterogeneity. Specifically, disparate impact may arise primarily from differences by race in true positives—placing in foster care Black children in high-risk situations at higher rates than white children in high-risk situations—or

false positives—placing Black children at higher rates in situations with low risk. Each of these two scenarios would suggest drastically different policy interventions to mitigate disparate impact.

We thus decompose our disparate impact estimates to better understand their sources (Table 5). Panel A shows estimates of $\Delta_{j1}$ and $\Delta_{j0}$, while Panel B shows each of the $\hat{\delta}_{jry}$ components. We find that disparate impact is concentrated in the population of children with subsequent maltreatment potential ($Y_i^* = 1$). Estimates from linear extrapolations, for example, show that Black children are placed in foster care at a rate 6.1 percentage points (over 100%) higher relative to white children in the population with subsequent maltreatment potential.[22] Racial disparities in the population without subsequent maltreatment potential ($Y_i^* = 0$) are much smaller and not statistically significant.[23] 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—these estimates suggest that disparate impact may arise primarily from the under-placing of white children who are in potentially dangerous situations.

## VI.C Heterogeneity by Investigator Characteristics

We examine whether investigators' observable characteristics are associated with greater disparate impact. Table 6 reports estimates of OLS regressions of the $\hat{\Delta}_j$'s on investigator observables. We find that investigators' individual racial disparities are significantly lower for Black investigators relative to white investigators (1.8 and 0.7 percentage points, respectively). Racial disparities is also lower for new investigators, hired at some point during our sample, for investigators with a caseload featuring an above-median share of Black children, and for investigators with below-median placement rates.

Since the bulk of disparate impact arises among children with subsequent maltreatment potential, we also explore whether investigators' observable characteristics are associated with greater values of $\Delta_{j1}$, racial disparities among children with high subsequent

[22] Panel B of Table 5 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 is substantively different in the two types of home circumstances.

[23] As in Figure 2, Figure A2 displays logical bounds on possible disparities where there is and is not maltreatment potential (case-weighted average $\Delta_{j1}$'s in Panel A and $\Delta_{j0}$'s in Panel B, respectively). The figure shows that most of the range of estimates of racial disparities within cases with subsequent maltreatment potential is above zero, while the range of possible disparities within cases without subsequent maltreatment potential is estimated to be from around -4 percentage points to 4 percentage points.

maltreatment potential. We find that placement disparities among children with potential for subsequent maltreatment are driven by investigators with less exposure to Black children (Table A6), consistent with the findings in Table 6. Specifically, the gaps are driven by white investigators, male investigators, investigators outside of urban counties, and investigators with caseloads with a below-median share of Black children.

Interestingly, the estimates in Column 1 suggest that white investigators tend to place in foster care white children in high-risk situations at lower rates than Black children in high-risk situations; conversely Black investigators tend to do the opposite, placing Black children in high-risk situations at lower rates than white children in high-risk situations. These estimates suggest that investigators of a given race tend to 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, white children tend to be under-placed relative to Black children in high-risk situations.

We also analyze the level of racial disparity among high-risk cases ($\Delta_{j1}$) over an investigator's career. Specifically, we compute separate estimates of $\Delta_{j1}$ in the first, second, and third tercile of investigations that an investigator handles in our analysis sample. For this exercise, we focus only on the set of 457 "new" investigators in our sample so as to make terciles of cases as comparable as possible among investigators. Figure A3 shows that estimates of $\Delta_{j1}$ increase significantly over time. That is, racial disparities among children in high-risk situations increases with investigators' experience.

# VII Conclusion

This study is the first to use a quasi-experimental design to test for racial discrimination in foster care placement decisions. 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. Importantly, we show that disparate impact is concentrated in the population of children with subsequent maltreatment potential in the home. Disparate impact is significantly larger among investigators who are white, who are assigned a lower share of cases involving Black families, who are more experienced, and who have relatively high foster care placement rates.

There are several questions raised by this analysis that we hope to address in future drafts. First, we hope to disentangle the drivers of disparate impact in foster care decisions. Our primary analysis suggests non-race characteristics play a minimal role in this setting, leaving on the table classic models of race-based bias and statistical discrimination. Imposing

additional structure on the quasi-experimental variation, as in Arnold, Dobbie and Hull (2022) and Chan, Gentzkow and Yu (2022), may allow us to distinguish between these potential drivers and allow for a more nuanced analysis of appropriate policy responses.

Second, we hope to study how discrimination in foster care decisions may be affected by earlier stages of the CPS process. We focus on disparate impact in foster care decisions since these are perhaps the farthest-reaching CPS tool, previous work has found that they can have substantial impacts on key long-term outcomes, and they are the basis for most current policy interventions aimed at addressing racial disparities in child welfare systems. However, discrimination at other points of CPS involvement (e.g., at the screening decision) may exist and interact with subsequent foster care decisions. As discussed above, MDHHS data contain information on the hotline screeners and the screened-out calls. Screeners are themselves quasi-randomly assigned and have substantial discretion on whether to screen-in a call. We hope use this information to examine racial disparities in CPS interventions through a "systems-based" approach (Bohren, Hull and Imas, 2022; Darity Jr and Mason, 1998; Darity et al., 2005; Pincus, 1996; Powell, 2007), in which we analyze disparities as the outcome of discrimination across different stages of CPS involvement.

Finally, we hope to broaden our analysis to different CPS systems. We currently examine only one state where we are certain of the quasi-random assignment of investigators. However, given that underlying contexts likely vary considerably and that the child maltreatment system is locally administered, it would be valuable to know how similar other states may be. In future versions of this study, we hope to use national data to contextualize our state-level findings.

# References


**Aigner, Dennis J, and Glen G Cain.** 1977. "Statistical theories of discrimination in labor markets." *Ilr Review*, 30(2): 175–187.

**Andrews, Donald WK, and Marcia MA Schafgans.** 1998. "Semiparametric estimation of the intercept of a sample selection model." *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(1).

**Arnold, David, Will Dobbie, and Crystal S Yang.** 2018. "Racial bias in bail decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

**Arnold, David, Will S Dobbie, and Peter Hull.** 2022. "Measuring racial discrimination in bail decisions." *American Economic Review*, Forthcoming.

**Arrow, Kenneth J.** 1973. "The Theory of Discrimination." in O. Ashenfelter and A. Rees (eds.).

**Bald, Anthony, Eric Chyn, Justine S. Hastings, and Margarita Machelett.** 2022*a*. "The causal impact of removing children from abusive and neglectful homes." *Journal of Political Economy*, Forthcoming.

**Bald, Anthony, Joseph Doyle Jr., Max Gross, and Brian Jacob.** 2022*b*. "Economics of Foster Care." *Journal of Economic Perspectives*, Forthcoming.

**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.

**Baron, E Jason, Ezra G Goldstein, and Joseph Ryan.** 2022. "The Push for Racial Equity in Child Welfare: Can Blind Removals Reduce Disproportionality?" *Journal of Policy Analysis and Management*.

**Becker, Gary S.** 1957. *The economics of discrimination*. University of Chicago press.

**Billingsley, Andrew, and Jeanne M Giovannoni.** 1972. "Children of the storm: Black children and American child welfare."

**Bohren, J Aislinn, Peter Hull, and Alex Imas.** 2022. "Systemic discrimination: Theory and measurement." National Bureau of Economic Research.

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

**Brinch, Christian N, Magne Mogstad, and Matthew Wiswall.** 2017. "Beyond LATE with a discrete instrument." *Journal of Political Economy*, 125(4): 985–1039.

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

**Chan, David C, Matthew Gentzkow, and Chuan Yu.** 2022. "Selection with variation in diagnostic skill: Evidence from radiologists." *The Quarterly Journal of Economics*, 137(2): 729–783.

**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.

**Clifford, Stephanie, and Jessica Silver-Greenberg.** 2017. "Foster care as punishment: The new reality of 'Jane Crow'." *The New York Times*. https://www.nytimes.com/2017/07/21/nyregion/foster-care-nyc-jane-crow.html.

**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*, 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.

**Darity Jr, William A, and Patrick L Mason.** 1998. "Evidence on discrimination in employment: Codes of color, codes of gender." *Journal of Economic Perspectives*, 12(2): 63–90.

**Darity, William, et al.** 2005. "Stratification economics: the role of intergroup inequality." *Journal of Economics and finance*, 29(2): 144.

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

**Donohue, John J.** 2007. "Antidiscrimination Law." In *Handbook of Law and Economics.* , ed. A. Mitchell Polinsky and Steven Shavell. New York:Elsevier.

**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, Jr, Joseph J.** 2013. "Causal effects of foster care: An instrumental-variables approach." *Children and Youth Services Review*, 35(7): 1143–1151.

**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.

**d'Haultfoeuille, Xavier, and Arnaud Maurel.** 2013. "Another look at the identification at infinity of sample selection models." *Econometric Theory*, 29(1): 213–224.

**Feigenberg, Benjamin, and Conrad Miller.** 2022. "Would eliminating racial disparities in motor vehicle searches have efficiency costs?" *The 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.

**Frandsen, Brigham, Lars Lefgren, and Emily Leslie.** 2023. "Judging judge fixed effects." *American Economic Review*, 113(1): 253–77.

**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.

**Grimon, Marie-Pascale.** 2020. "Effects of the Child Protection System on Parents." Working 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.

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

**Hull, Peter.** 2020. "Estimating hospital quality with quasi-experimental data." *Working Paper*.

**Hull, Peter.** 2021. "What marginal outcome tests can tell us about racially biased decision-making." *NBER Working Paper #28503*.

**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. PMID: 27997240.

**Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan.** 2018. "Human decisions and machine predictions." *The quarterly journal of economics*, 133(1): 237–293.

**Kowalski, Amanda E.** 2016. "Doing more when you're running late: Applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments." *NBER Working Paper #w22363*.

**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): 3663–98.

**Mogstad, Magne, Andres Santos, and Alexander Torgovitsky.** 2018. "Using instrumental variables for inference about policy relevant treatment parameters." *Econometrica*, 86(5): 1589–1619.

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

**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.

**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." *The 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.** 2007. "Structural racism: building upon the insights of John Calmore." *NCL Rev.*, 86: 791.

**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, 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.

**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.

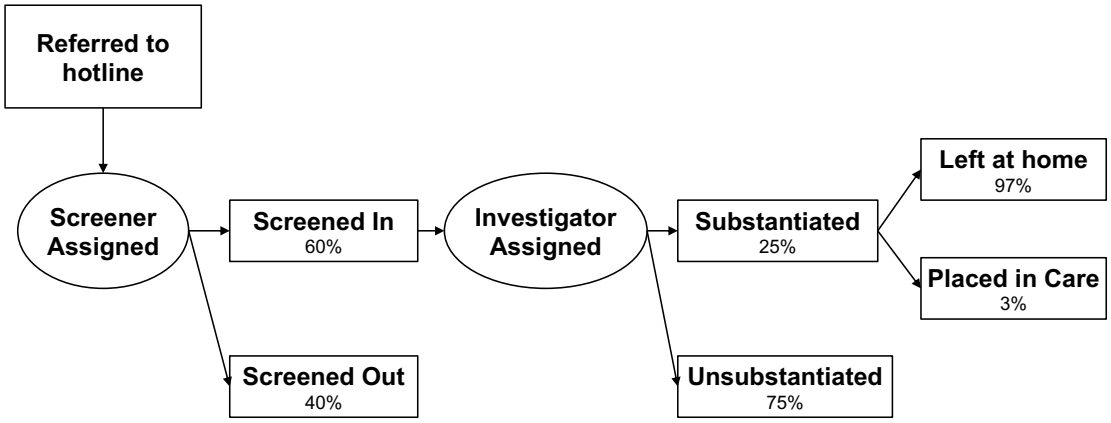
**Shaw, Terry V, Emily Putnam-Hornstein, Joseph Magruder, and Barbara Needell.** 2008. "Measuring racial disparity in child welfare." *Child welfare*, 87(2): 23–36.

**Skiena, Steven, and Charles B Ward.** 2014. *Who's bigger?: Where historical figures really rank.* Cambridge University Press.

**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.

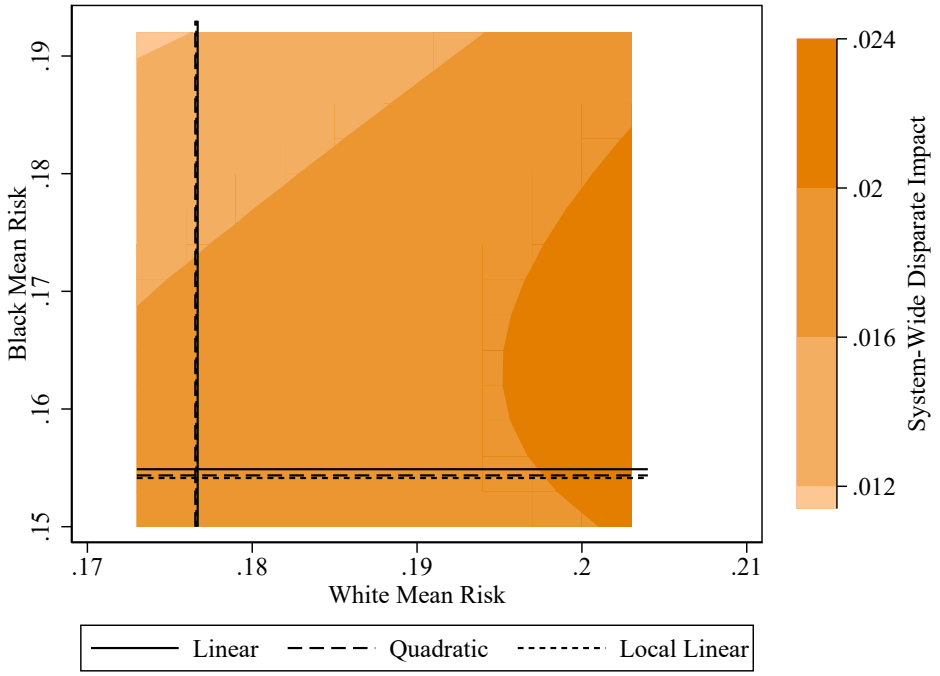
**Wulczyn, Fred, Robert Gibbons, Lonnie Snowden, and Bridgette Lery.** 2013. "Poverty, social disadvantage, and the black/white placement gap." *Children and Youth Services Review*, 35(1): 65–74.

Figure 1: Child Welfare Process in Michigan



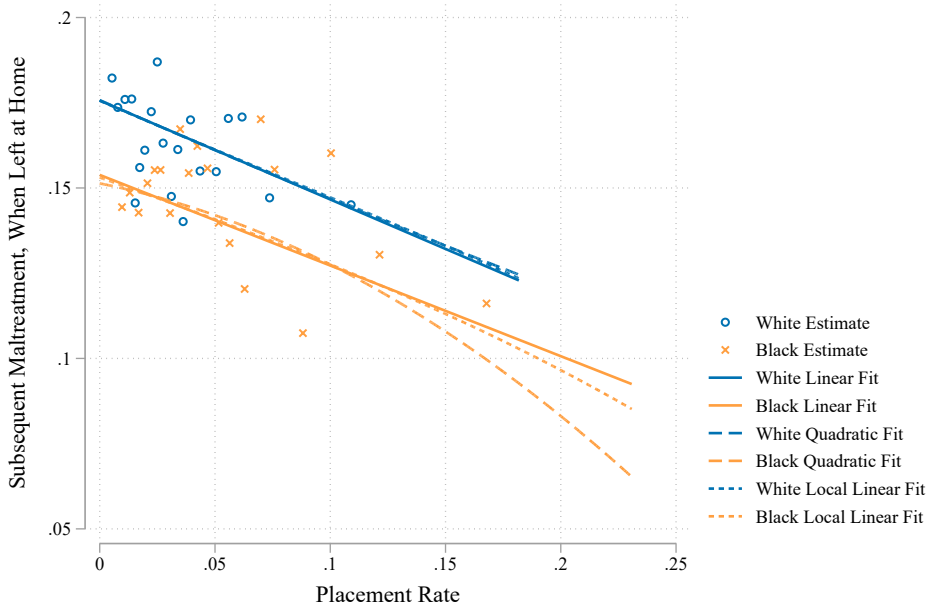

Notes: The figure describes the child maltreatment investigation process in Michigan. Both screeners and investigators are quasi-randomly assigned. The percentages on screening in and out refer to all calls received; percentages thereafter refer to investigated cases. "Substantiated" means that the investigator found enough evidence to support the allegation of abuse or neglect. Conditional on substantiation, in cases with the most intensive risk the child is also placed in foster care.

Figure 2: Robustness of Estimates of System-Wide Disparate Impact




Notes. This figure shows how our estimates of system-wide disparate impact 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: $\mu_b \in [0.150, 0.193]$ and $\mu_w \in [0.173, 0.204]$.

Figure 3: Investigator-Specific Placement and Conditional Re-Investigation 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).

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...* | | | |
| Physical or substance abuse allegation | 0.426 | 0.412 | 0.432 |
| Domestic violence allegation | 0.106 | 0.099 | 0.108 |
| Drug residence allegation | 0.031 | 0.034 | 0.030 |
| Alleged perpetrator is the parent/step-parent | 0.910 | 0.902 | 0.913 |
| Alleged perpetrator is a non-parent relative | 0.052 | 0.063 | 0.048 |
| *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: Benchmarking Regression

| **Dependent variable:** | **(1) Foster Care** | **(2) Foster Care** | **(3) Foster Care** |
|---|---|---|---|
| Black | 0.012*** | 0.012*** | 0.010*** |
| | (0.002) | (0.002) | (0.002) |
| Female | | | -0.001 |
| | | | (0.001) |
| Child had a previous investigation | | | 0.005*** |
| | | | (0.001) |
| Number of previous investigations | | | 0.003*** |
| | | | (0.000) |
| Physical or substance abuse allegation | | | 0.010*** |
| | | | (0.001) |
| Domestic violence allegation | | | 0.005** |
| | | | (0.002) |
| Drug residence allegation | | | 0.054*** |
| | | | (0.006) |
| Alleged perpetrator is a non-parent relative | | | 0.001 |
| | | | (0.002) |
| Alleged perpetrator is the parent/step-parent | | | 0.017*** |
| | | | (0.005) |
| Mean placement rate | 0.034 | 0.034 | 0.034 |
| Number of investigations | 217,704 | 217,704 | 217,704 |
| Rotation Group FE | N | Y | Y |
| Age FE | N | N | 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Estimates of Mean Risk by Race and Disparate Impact

| | (1) Linear extrapolation | (2) Quadratic extrapolation | (3) Local linear extrapolation |
|---|---|---|---|
| | *Panel A: Mean 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 disparate impact* | | |
| 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 risk and disparate impact 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 disparate impact (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: Heterogeneity in Estimates of $\Delta_j$ by Child and Investigation Characteristics

| | (1) All Children | (2) Male Children | (3) Female Children | (4) Young Children | (5) Old Children | (6) Phy Abuse | (7) Neglect |
|---|---|---|---|---|---|---|---|
| | *Panel A: Mean risk by race* | | | | | | |
| Black children | 0.155*** | 0.156*** | 0.150*** | 0.154*** | 0.148*** | 0.142*** | 0.155*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| White children | 0.177*** | 0.174*** | 0.176*** | 0.196*** | 0.155*** | 0.153*** | 0.182*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) |
| | *Panel B: Estimates of $\Delta_j$, by subgroup* | | | | | | |
| 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) | (0.002) | (0.002) |
| Number of investigators | 699 | 701 | 661 | 691 | 571 | 229 | 807 |

Notes. This table summarizes estimates of mean risk and disparate impact 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 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) disparate impact. 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 5: Decomposition of Disparate Impact by Subsequent Maltreatment Potential

| | (1) Linear extrapolation | (2) Quadratic extrapolation | (3) Local linear extrapolation |
|---|---|---|---|
| | *Panel A: Conditional Disparate Impact* | | |
| $\Delta_{j1}$ | 0.061*** | 0.058*** | 0.057*** |
| | (0.021) | (0.021) | (0.021) |
| $\Delta_{j0}$ | 0.008 | 0.008 | 0.009 |
| | (0.006) | (0.006) | (0.006) |
| | *Panel B: Conditional Removal Rates* | | |
| $E[D\|R = b, Y^* = 1]$ | 0.119 | 0.117 | 0.115 |
| $E[D\|R = w, Y^* = 1]$ | 0.059 | 0.058 | 0.058 |
| $E[D\|R = b, Y^* = 0]$ | 0.034 | 0.034 | 0.034 |
| $E[D\|R = w, Y^* = 0]$ | 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. $\Delta_{j1}$ reflects racial disparities among children with potential for subsequent maltreatment ($Y_i^* = 1$) and $\Delta_{j0}$ 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.
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

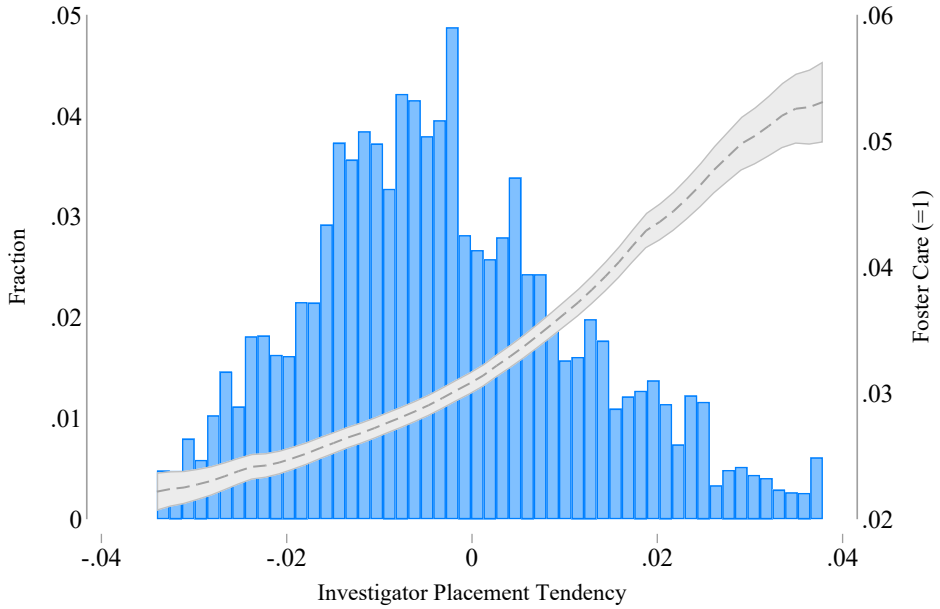
Table 6: Heterogeneity in Estimates of $\Delta_j$ by Investigator Characteristics

| **Dependent variable:** | **(1)** $\hat{\Delta}_j$ | **(2)** $\hat{\Delta}_j$ | **(3)** $\hat{\Delta}_j$ | **(4)** $\hat{\Delta}_j$ | **(5)** $\hat{\Delta}_j$ | **(6)** $\hat{\Delta}_j$ | **(7)** $\hat{\Delta}_j$ |
|---|---|---|---|---|---|---|---|
| Black investigator | -0.011** | | | | | | -0.007 |
| | (0.005) | | | | | | (0.005) |
| Other race | -0.004 | | | | | | -0.005 |
| | (0.010) | | | | | | (0.010) |
| Female investigator | | -0.008 | | | | | -0.010 |
| | | (0.006) | | | | | (0.006) |
| New investigator | | | -0.011** | | | | -0.008 |
| | | | (0.005) | | | | (0.005) |
| Urban county | | | | -0.005 | | | 0.006 |
| | | | | (0.005) | | | (0.004) |
| High share of black children | | | | | -0.014*** | | -0.016*** |
| | | | | | (0.005) | | (0.004) |
| Low placement rate | | | | | | -0.009* | -0.009 |
| | | | | | | (0.005) | (0.005) |
| Control Mean | 0.018 | 0.023 | 0.021 | 0.020 | 0.024 | 0.022 | 0.017 |
| Number of investigators | 699 | 699 | 699 | 699 | 699 | 699 | 699 |

Notes. This table reports OLS estimates of regressions of disparate impact estimates (based on a 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 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 B, Column 1 of Table 3. 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

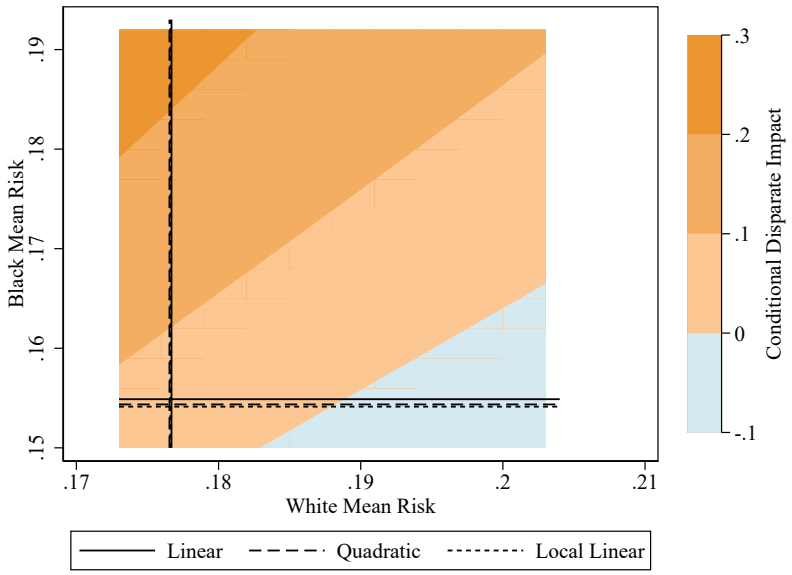
# Online Appendix: Supplemental Figures and Tables

Figure A1: Distribution of Investigator Placement Tendency

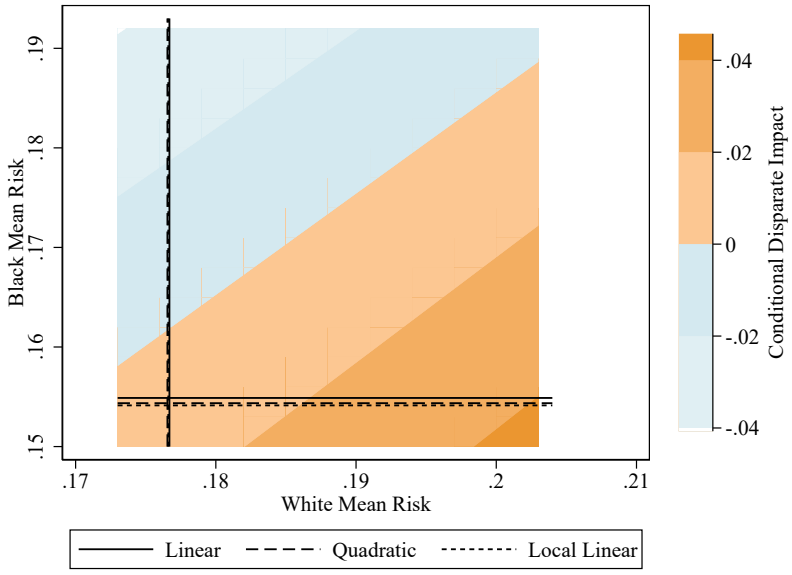



Notes. This figure shows the distribution of the investigator foster care placement tendency instrument net of zip code by investigation year fixed effects. The dashed line shows point estimates from a non-parametric regression of an indicator variable equal to one if the child was placed in foster care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

Figure A2: Robustness of Estimates of Disparate Impact by Subsequent Maltreatment Potential



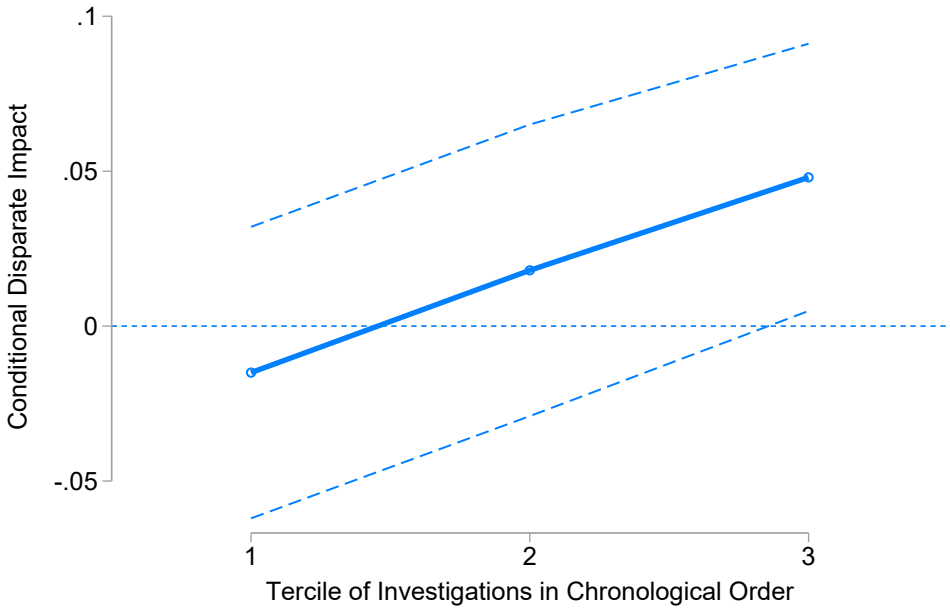

(a) $\Delta_{j1}$




(b) $\Delta_{j0}$

Notes. This figure shows how our estimates of $\Delta_{j1}$ (Panel A) and $\Delta_{j0}$ (Panel B) 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: $\mu_b \in [0.150, 0.193]$ and $\mu_w \in [0.173, 0.204]$.

Figure A3: Estimates of $\Delta_{j1}$ by Investigator Experience




Notes. This figure examines the evolution of disparate impact estimates in the population where $Y_i^* = 1$ and over an investigator's career. We compute separate estimates of $\Delta_{j1}$ in the first, second, and third tercile of investigations that an investigator handles in our analysis sample. We estimate $\Delta_{j1}$ via local linear extrapolations and weight the average in each tercile by the number of investigations assigned to each investigator in that tercile of cases. For this figure, we focus only on the set of 457 "new" investigators in our sample—those hired at some point during the estimation period—to make terciles of cases as comparable as possible among investigators. 95% confidence intervals are obtained by a bootstrapping procedure (500 times). Standard errors used in the construction of these intervals are clustered at the child and investigator levels.

Table A1: First-Stage Effects of Investigator Removal Tendency

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | **Black children** | | **White children** | |
| Removal Tendency | 0.548*** | 0.544*** | 0.625*** | 0.604*** |
| | (0.048) | (0.047) | (0.039) | (0.039) |
| F-statistic | 133 | 131 | 257 | 243 |
| Observations | 64,489 | 64,489 | 153,215 | 153,215 |
| Rotation group fixed effects | 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, Stock and Sun, 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

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Black children | | White children | |
| **Dependent variable:** | Foster Care | Removal Tendency | Foster Care | Removal Tendency |
| Female | -0.001 | 0.000 | -0.000 | 0.000 |
| | (0.002) | (0.000) | (0.001) | (0.000) |
| Child had a previous investigation | 0.007** | -0.000 | 0.005*** | 0.000 |
| | (0.003) | (0.000) | (0.001) | (0.000) |
| Number of previous investigations | 0.003** | -0.000 | 0.002*** | 0.000 |
| | (0.001) | (0.000) | (0.000) | (0.000) |
| Physical or substance abuse allegation | 0.007** | 0.001** | 0.012*** | 0.000*** |
| | (0.003) | (0.000) | (0.001) | (0.000) |
| Domestic violence allegation | 0.000 | -0.000 | 0.007*** | 0.000 |
| | (0.005) | (0.001) | (0.002) | (0.000) |
| Drug residence allegation | 0.023** | 0.002* | 0.068*** | 0.000 |
| | (0.010) | (0.001) | (0.008) | (0.000) |
| Alleged perpetrator is a non-parent relative | -0.004 | 0.000 | 0.004 | -0.000 |
| | (0.004) | (0.001) | (0.003) | (0.000) |
| Alleged perpetrator is the parent/step-parent | 0.007 | 0.001 | 0.021*** | 0.001 |
| | (0.008) | (0.001) | (0.006) | (0.001) |
| Number of investigations | 64,489 | 64,489 | 153,215 | 153,215 |
| F-Stat from joint test | 16.372 | 1.324 | 33.728 | 1.202 |
| P-Value from joint test | 0.000 | 0.138 | 0.000 | 0.231 |
| Rotation group fixed effects | Y | Y | Y | Y |
| Age fixed effects | Y | Y | Y | Y |

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 age 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 A3: Bounds on Estimates of Mean Risk and Disparate Impact

| | **(1) From 0.03 Removal Rate** | **(2) From 0.02 Removal Rate** | **(3) From 0.01 Removal Rate** |
|---|---|---|---|
| | *Panel A: Mean 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 disparate impact* | | |
| 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) |
| Mean placement rate | 0.034 | 0.034 | 0.034 |
| 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) disparate impact. To estimate bounds on mean risk, Column 1 uses a 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 disparate impact 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 A4: Estimates of Mean Risk by Race and Disparate Impact (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** |
|---|---|---|---|---|---|---|---|
| | | | | *Panel A: Mean risk by race* | | | |
| Black children | 0.155*** | 0.132*** | 0.110*** | 0.086*** | 0.060*** | 0.039*** | 0.017*** |
| | (0.003) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| White children | 0.177*** | 0.153*** | 0.129*** | 0.099*** | 0.067*** | 0.050*** | 0.020*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.001) | (0.002) | (0.002) |
| | | | | *Panel B: System-wide disparate impact* | | | |
| Mean across inv. | 0.017*** | 0.016*** | 0.016*** | 0.016*** | 0.016*** | 0.015*** | 0.016*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Mean placement rate | 0.034 | 0.034 | 0.034 | 0.034 | 0.034 | 0.034 | 0.034 |
| Number of investigators | 699 | 699 | 699 | 699 | 699 | 699 | 699 |

Notes. This table summarizes estimates of mean risk and disparate impact 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) disparate impact. Estimates in Panel A come from 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 A5: Estimates of Mean Risk by Race and Disparate Impact (Covariate Adjustment)

| | (1) Linear extrapolation | (2) Quadratic extrapolation | (3) Local linear extrapolation |
|---|---|---|---|
| | *Panel A: Mean risk by race* | | |
| Black children | 0.129*** | 0.135*** | 0.131*** |
| | (0.003) | (0.004) | (0.005) |
| White children | 0.158*** | 0.150*** | 0.157*** |
| | (0.004) | (0.003) | (0.004) |
| | *Panel B: System-wide disparate impact* | | |
| Mean across investigators | 0.015*** | 0.015*** | 0.015*** |
| | (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 risk and disparate impact 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 disparate impact (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 A6: Heterogeneity in Estimates of $\Delta_{j1}$ by Investigator Characteristics

| **Dependent variable:** | **(1)** $\hat{\Delta}_{j1}$ | **(2)** $\hat{\Delta}_{j1}$ | **(3)** $\hat{\Delta}_{j1}$ | **(4)** $\hat{\Delta}_{j1}$ | **(5)** $\hat{\Delta}_{j1}$ | **(6)** $\hat{\Delta}_{j1}$ | **(7)** $\hat{\Delta}_{j1}$ |
|---|---|---|---|---|---|---|---|
| Black investigator | -0.141** | | | | | | -0.079 |
| | (0.063) | | | | | | (0.066) |
| Other race | -0.223 | | | | | | -0.206 |
| | (0.149) | | | | | | (0.152) |
| Female investigator | | -0.118** | | | | | -0.107** |
| | | (0.052) | | | | | (0.054) |
| New investigator | | | -0.066 | | | | -0.061 |
| | | | (0.049) | | | | (0.054) |
| Urban county | | | | -0.151*** | | | -0.062 |
| | | | | (0.048) | | | (0.050) |
| High share of black children | | | | | -0.160*** | | -0.104** |
| | | | | | (0.048) | | (0.047) |
| Low placement rate | | | | | | -0.005 | -0.006 |
| | | | | | | (0.050) | (0.054) |
| Control mean | 0.082 | 0.142 | 0.087 | 0.133 | 0.138 | 0.063 | 0.061 |
| Number of investigators | 699 | 699 | 699 | 699 | 699 | 699 | 699 |

Notes. This table reports OLS estimates of regressions of estimates of $\Delta_{j1}$ (based on a 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. 87% 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 1 of Table 5. 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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.