

DISCRIMINATION IN MULTI-PHASE SYSTEMS:
EVIDENCE FROM CHILD PROTECTION

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

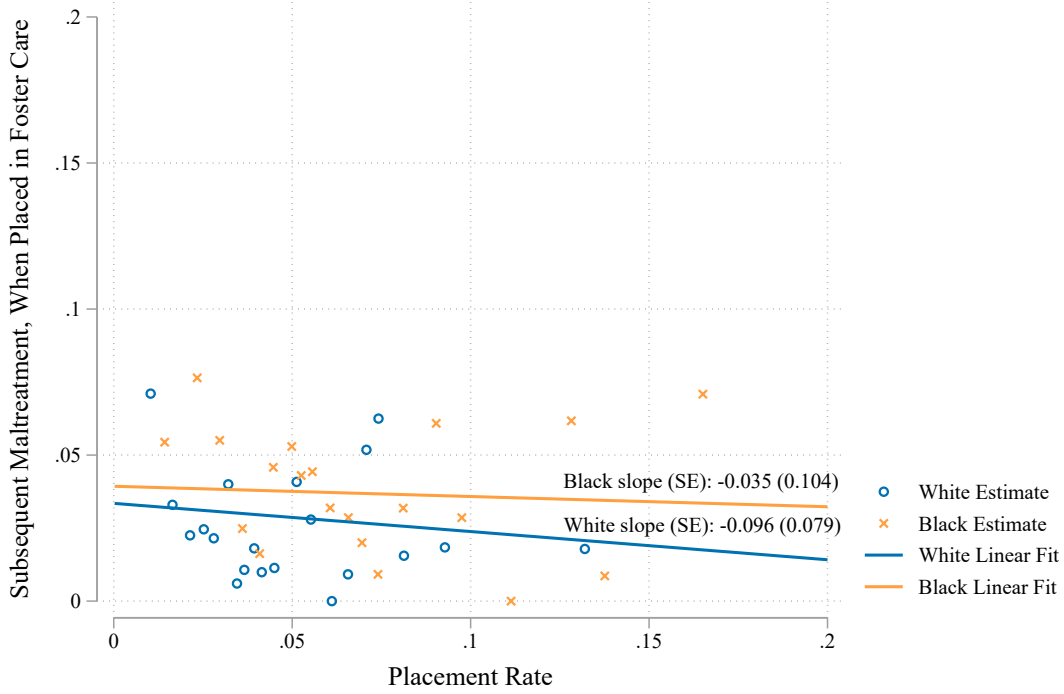
Online Appendix

Table of Contents

A Supplemental Figures and Tables	2
B Exclusion Restriction Tests and Extensions	20
C Impacts on Welfare-Relevant Outcomes	24
D Potential Drivers of Investigator UD	26
E Nationwide Estimates From NCANDS Data	30

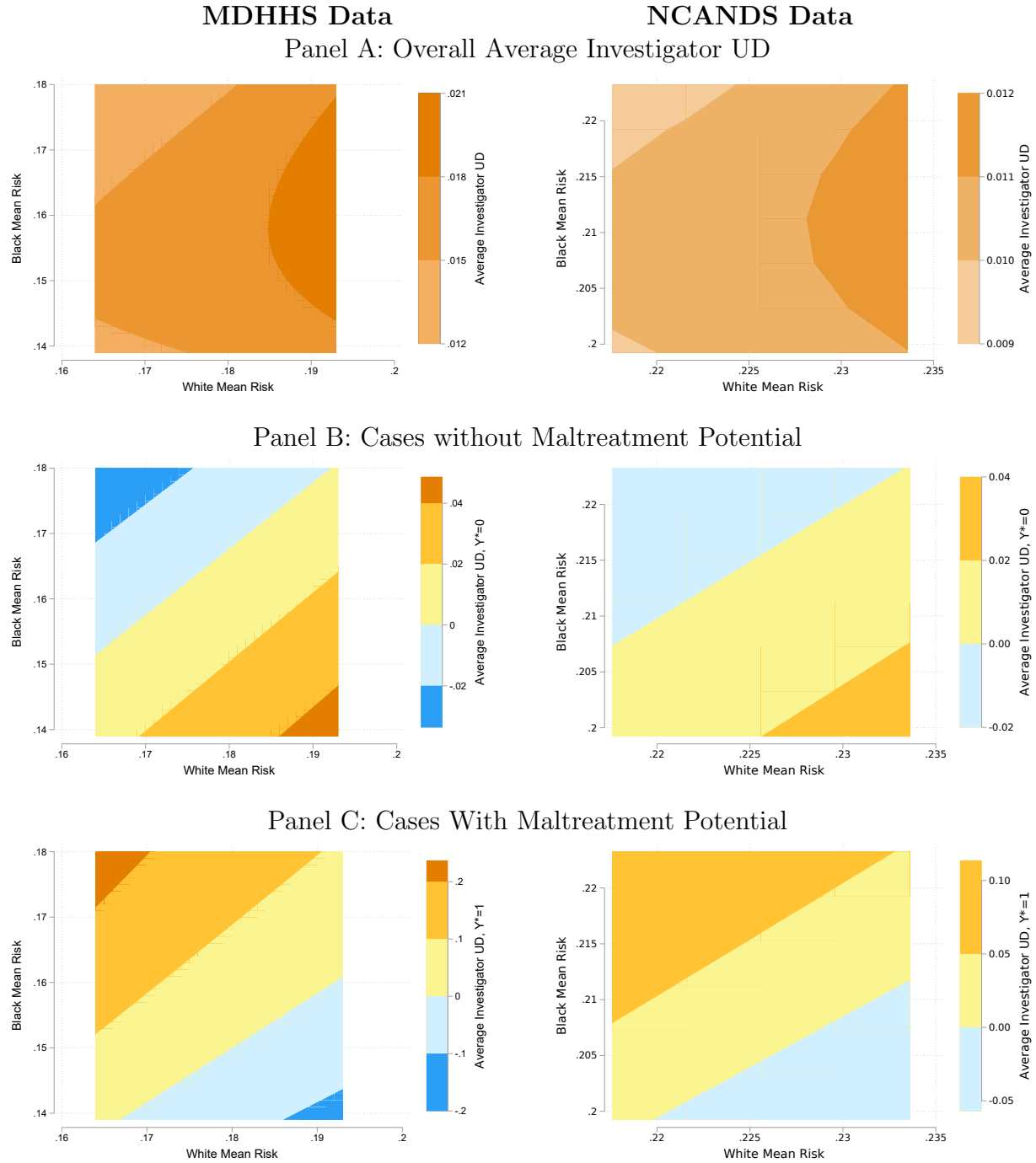
A SUPPLEMENTAL FIGURES AND TABLES

Figure A.I: Investigator Placement and Subsequent Maltreatment Rates in Foster Care



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

Figure A.II: Non-Parametric Bounds of Average Investigator UD (MDHHS vs. NCANDS)



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

Table A.I: Screener Balance Tests

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

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

Table A.II: Investigators Balance Tests

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

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

Table A.III: Additional Balance Tests for Investigator Assignment

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

	(1) Weighted average	(2) With maltreatment potential	(3) Without maltreatment potential
<i>Panel A: Conditional screen-in rates</i>			
Black children	0.632 (0.001)	0.634 (0.004)	0.632 (0.001)
White children	0.582 (0.001)	0.595 (0.003)	0.579 (0.001)
<i>Panel B: Placement rates among screened-in calls</i>			
Black children	0.048 (0.002)	0.134 (0.018)	0.029 (0.004)
White children	0.031 (0.001)	0.076 (0.012)	0.021 (0.003)
<i>Panel C: Placement rates among all calls</i>			
Black children	0.029 (0.001)	0.085 (0.012)	0.018 (0.003)
White children	0.018 (0.001)	0.045 (0.007)	0.012 (0.002)
Number of screeners	162	162	162
Number of investigators	814	814	814

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

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

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

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

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

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

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

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

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

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

Table A.XII: Hierarchical MTE Model Estimates

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

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

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

	(1) Baseline	(2) No Strata Adjustment	(3) No Weighting	(4) Empirical Bayes Shrinkage
<i>Panel A: Subsequent maltreatment risk</i>				
Black children	0.155 (0.003)	0.146 (0.003)	0.147 (0.004)	0.158 (0.004)
White children	0.175 (0.003)	0.172 (0.002)	0.166 (0.003)	0.183 (0.003)
<i>Panel B: Investigator UD</i>				
Mean across investigators	0.017 (0.002)	0.017 (0.002)	0.015 (0.002)	0.016 (0.002)
Number of investigators	814	814	814	814

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

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

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

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

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

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

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

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

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

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

B EXCLUSION RESTRICTION TESTS AND EXTENSIONS

Bounding Screening-In Effects

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

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

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

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

$$\begin{aligned} (1 - P_i)Y_i^* &= (1 - P_i)(1 - S_i)Y_i^{*,S=0} + (1 - P_i)S_iY_i^{*,S=1} \\ &= (1 - S_i)Y_i^{*,S=0} + S_i[(1 - D_i)Y_i^{*,S=1}], \end{aligned} \quad (17)$$

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

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

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

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

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

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

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

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

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

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

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

The upper bound can moreover be written:

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

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

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

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

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

Exclusion-Free Investigator UD Bounds

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

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

and

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

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

$$E[Y_i^* \mid R_i = r] \in [E[(1 - D_i)Y_i^* \mid R_i = r], 1 - E[(1 - D_i)(1 - Y_i^*) \mid R_i = r]], \quad (26)$$

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

$$\begin{aligned} \overline{\Delta}^D &= (E[D_i \mid R_i = b, Y_i^* = 0] - E[D_i \mid R_i = w, Y_i^* = 0]) (1 - E[Y_i^*]) \\ &\quad + (E[D_i \mid R_i = b, Y_i^* = 1] - E[D_i \mid R_i = w, Y_i^* = 1]) E[Y_i^*]. \end{aligned} \quad (27)$$

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

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

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

$$\begin{aligned} \bar{\Delta}^{D,T=0} = & (E[D_i \mid R_i = b, Y_i^*(0) = 0] - E[D_i \mid R_i = w, Y_i^*(0) = 0]) (1 - E[Y_i^*(0)]) \\ & + (E[D_i \mid R_i = b, Y_i^*(0) = 1] - E[D_i \mid R_i = w, Y_i^*(0) = 1]) E[Y_i^*(0)]. \end{aligned} \quad (28)$$

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

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

and

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

C IMPACTS ON WELFARE-RELEVANT OUTCOMES

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

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

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

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

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

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

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

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

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

D POTENTIAL DRIVERS OF INVESTIGATOR UD

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

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

Model Setup

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

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

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

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

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

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

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

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

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

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

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

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

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

SMD Estimator

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

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

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

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

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

Results

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

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

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

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

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

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

E NATIONWIDE ESTIMATES FROM NCANDS DATA

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

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

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

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

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

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

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

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

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

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

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