"Discrimination in Multi-Phase Systems: Evidence from Child Protection"

Baron, Doyle, Emanuel, Hull, Ryan

Replication README

1. Overview

The code in this replication package creates all figures and tables presented in our paper. The replication process is documented in "0_master.do" and proceeds in two parts:

The first part contains most of our analysis and has three steps:

1. We generate analysis samples at the child-by-call or child-by-investigation levels using administrative datasets from Michigan.
2. Using the files generated in Step 1, we generate our main "screener" and "investigator" samples.
3. We use the datasets created in the first two steps to generate the tables and figures in the paper, except for Table A12, Figure 4, and the right-hand side of Figure A2. Table A12 is estimated using Matlab and is discussed in Section 3.

The second part generates Figure 4 and the right-hand side Figure A2 and requires two steps.

1. We import National Child Abuse and Neglect Data System (NCANDS) data and clean it to be as similar to our main Michigan sample as possible. We then generate estimates of unwarranted disparity for all the states in the NCANDS data and make the right-hand side of Figure A2.
2. We use R to generate Figure 4.

The code in the replication file was last run using Stata 18. The code takes about 6 days to run in its entirety. Table A12 is created using Matlab R2022b. Figure 4 is made using R 4.1.2.

2. Data Availability and Provenance Statements

We generate two datasets. The first dataset aggregates administrative data from Michigan. The second uses NCANDS data.

Michigan Data. The administrative education, child welfare, and juvenile justice data used for the project require a data sharing agreement with the relevant agency in Michigan and cannot be shared publicly online. We are able to post all of the code used for data analysis, however, and we will assist with any reasonable replication attempts for two years following publication.

The administrative education records were provided through the Center for Educational Performance and Information (CEPI) and the Michigan Department of Education (MDE) (CEPI, 2017; MDE, 2017). These records are confidential but may be obtained with Data Use Agreements with MDE and CEPI. Researchers interested in access to the data may contact the recently created Michigan Education Data Center at meri@miedresearch.org. See also https://medc.miedresearch.org/research-application.

The administrative child welfare records were provided through MDHHS (MDHHS, 2017,2019). These records are confidential but may be obtained with Data Use Agreements with MDHHS. Researchers interested in access to the data may contact the Division of Continuous Quality Improvement, Children's Services Agency within MDHHS at MDHHS-DCQI-CQI@michigan.gov.

The administrative juvenile justice records were provided through the State Court Administrative Office (SCAO) (SCAO, 2015). These records are confidential but may be obtained with Data Use Agreements with SCAO. Researchers interested in access to the data may contact Laura Hutzel, Director of the Statistical Research Staff, at hutzell@courts.mi.gov.

We note that each of these three data sources has a separate data request process. Linking these data sources requires negotiating data sharing agreements with all three parties. Note that this particular project does not use any education or juvenile justice data in the analysis. Rather, these datasets were inputs into the construction of the investigator analysis sample, as detailed in Gross and Baron (2022). Therefore, strictly speaking, a replicator may not need to access all three administrative data systems to replicate our main results, as long as the main sample restrictions in Gross and Baron (2022) are followed. The investigator analysis sample in the current project expands the original Gross and Baron (2022) dataset, which spanned 2008 to 2017, to include 2018 and 2019. The screener analysis sample is constructed using only data from MDHHS.

NCANDS data. The data from the National Child Abuse and Neglect Data System is housed at the National Data Archive on Child Abuse and Neglect (NDACAN). Researchers may obtain the data via application.

3. Description of Programs

The programs listed in 0_master.do create all figures and tables in the paper. There are four main sections in this program:

1. Build the main analysis samples at either the child-by-call or child-by-investigation levels from raw MDHHS data.

2. Build the main investigator- and screener-level analysis samples using the microdata built the first step.

3. Generate all tables and figures in the paper, except for Figure 4, the right-hand side of Figure A2, and Table A12.

To generate Table A12, run the following programs. The first is a Stata program, which generates the sample moments needed for estimation. The last two are Matlab programs, which estimate the hierarchical marginal treatment effects model and generate the table.

1. d_gen_moments_hmte.do

2. d_estimate_qje.m

3. d_make_table_qje.m

4. To generate the right-hand side of Figure A2 and Figure 4, run the following programs. The first is a Stata program that generates the NCANDS sample. The second is an R program that creates the maps.

1. ncands.do
2. make_maps.R

4. Computational requirements

Stata (code was last run with version 18). The first section of "0_master.do" will install all dependencies locally and should be run once.

Matlab (code was last run with release R2022b). All subroutines required for the estimation of the hierarchical marginal treatment effect model are included in the replication package.

The replication takes roughly 6 days to run. The reason it takes this long is the creation of the screener- and investigator-level datasets, which use microdata and a linear regression adjustment to estimate decision-maker specific treatment and outcome rates.

5. Instructions to replicators

A replicator should be able to submit a proposal to the relevant Michigan agencies and enter in a data use agreement with them. As described above, the NCANDS data can be accessed via an application to NDACAN. With all the relevant files, a replicator should run the programs in "0_master.do" after updating all the file paths in the individual programs. Note that we removed the directory names from each program after running the replication file for data security concerns.

Details about each raw data source from MDHHS and NDACAN required to generate the analysis samples:

Dataset	Description	Source	Provided
allegations.dta	Contains the universe of allegations for child maltreatment in Michigan from 2008-2017	MDHHS	No
alleg_match.dta	Allegation-match file for the universe of allegations for child maltreatment in Michigan from 2008-2017	MDHHS	No
child_level.dta	Child-level file including the universe of children who were part of a child maltreatment investigation in Michigan from 2008-2017	MDHHS	No
case.dta	Case-level file including the universe of child maltreatment investigations in Michigan from 2008-2017	MDHHS	No
old_case_raw.dta	Case-level file including the universe of child maltreatment investigations in Michigan from 1996-2008	MDHHS	No
living_arrangements.dta	Case-level file including the living arrangements while in foster care of children in Michigan who were removed from their homes following a child maltreatment investigation from 2008-2017	MDHHS	No
workers.dta	Case-worker assignment file for the universe of child maltreatment investigations in Michigan from 2008-2017	MDHHS	No
inv_insample.dta	List of primary CPS investigators in the state's human resource files from 2008 to 2017	MDHHS	No

final_prediction_race_gender.dta	Names of investigators and predicted race and gender, for investigators appearing in inv_insample.dta	MDHHS	No
2017_19_removals.dta	Contains the caseworker assignment and a removal indicator for the universe of child maltreatment investigations in Michigan from 2017 to 2019	MDHHS	No
screened_out_calls.csv	Contains the universe of screened-out hotline calls alleging child maltreatment in Michigan from 2017-2019	MDHHS	No
screened_in_calls.csv	Contains the universe of screened-in hotline calls alleging child maltreatment in Michigan from 2017-2019	MDHHS	No
CF2008v6.dta CF2009v7.dta CF2010v6.dta CF2011v6.dta CF2012v6.dta CF2013v6.dta CF2014v5.dta CF2015v5.dta CF2016v4.dta CF2017v4.dta CF2018v4.dta CF2019v5.dta CF2020v3.dta CF2021v2.dta	Contains the universe of reports of child maltreatment that have been sent by states to the US Department of Health and Human Services' Children's Bureau	NDACAN	No

References

CEPI (2017). "Individual K-12 Student Records." Michigan Consortium for Education Research. Accessed 2017-09-20.

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.

MDE (2017). "Individual K-12 Student Records." Michigan Consortium for Education Research. Accessed 2017-09-20.

MDHHS (2017). “Individual Child Maltreatment Complaints and Placement Records.” Child & Adolescent Data Lab. Accessed 2017-09-20.

MDHHS (2019). “Individual Child Maltreatment Complaints and Placement Records.” Child & Adolescent Data Lab. Accessed 2019-03-23.

SCAO (2015). “Individual Juvenile Delinquency Petitions.” Child & Adolescent Data Lab. Accessed 2019-08-24.