README file for

"The Effects of Incarceration on Mortality"

By Samuel Norris, Matthew Pecenco, and Jeffrey Weaver

This readme file contains instructions on how to replicate the results of our paper. The underlying data used for the project is not publicly shareable, and so we cannot post that data or share it with other researchers. However, we provide details on accessing all of the data sets used in the project on the next page. We also provide all of the cleaning code that can be used to transform the raw data for analysis as well as all of our analysis code.

Project Folders

The following diagram shows the levels of the project tree structure at a high level.

Project_Folder/

- README.pdf
- Code/
    - ado/
    - analysis/
    - cleaning/
    - pathnames.do
- Data/
    - Raw/
    - Interim/
    - Clean/
- Exhibits/

Before running the code, it is necessary to update the global paths in the pathnames.do file for the relevant file structure on your computer. If you set up your directories as in the above diagram, then it is only necessary to update the global $PROJ with the location of the project folder.

Within each sub-directory in the cleaning directory, there is a single code file that calls all of the other do-files in that directory (if relevant). In order to run all of the cleaning code in order, use the do-file "master_cleaning.do" in the "code/cleaning" directory. In order to run all of the analysis code, use the do-file "replication_analysis.do" in the "code/analysis" directory. Our code also uses a number of custom ado files, which can be found in the "code/ado" directory. See the below tables for details on each of the individual do-files we use for cleaning and analysis. The analysis output is all placed in the Exhibits folder.

Software and Computational Requirements

This section describes the software and computational requirements for running the relevant analysis code, as well as the relevant packages that must be installed.

- Stata version 15
  - ftools (as of 2020-08-17)
  - reghdfe (as of 2020-08-17)
  - outreg (as of 2020-08-17)
  - texsave (as of 2020-08-17)
  - estout (as of 2020-08-17)
  - unique (as of 2020-08-17)

Memory and Runtime Requirements

The code was last run on an Intel ® Core ™ i5-5200U CPU with 8 GB of RAM on Windows 10. Total run time for the analysis code is approximately 20 hours. Total run time for the cleaning code is approximately 5 days.

Project data

This table describes the different data sets used in the projects as well as how to get access to them.

Name of data	Description of data	Instructions for access
Cleveland Municipal Court records	Court dockets containing defendant name, date of birth, accused crime, all recorded information on the case, and case outcome.	Details on this data are available from https://clevelandmunicipalcourt.org/public-access . To request access to the data, contact the Clerk of Courts (ClerkofCourts@cmcoh.org)
Cuyahoga County Common Pleas Court records	Court dockets containing defendant name, date of birth, accused crime, all recorded information on the case, and case outcome.	Details on this data are available from https://cpdocket.cp.cuyahogacounty.us/ and https://coc.cuyahogacounty.us/en-us/public-records-request.aspx . To request access to the data, contact the Case Record Requests (coccfr@cuyahogacounty.us).
Franklin County Municipal and Common Pleas court records	Court dockets containing defendant name, date of birth, accused crime, all recorded information on the case, and case outcome.	To request access to this data, contact the Public Records Request office using the request form here.
Hamilton County Municipal and Common Pleas court records	Court dockets containing defendant name, date of birth, accused crime, all recorded information on the case, and case outcome.	To request access to this data, contact the Public Records Request office using the request form here.
National Longitudinal Survey of Youth	National panel dataset of individuals who were aged 12-17 when first interviewed in 1997. Contains extensive panel information on risky behaviors, criminal activity, and incarceration history.	Individuals can create accounts to access the data here. We only use the public-access version of the data set.
Ohio Department of Health mortality records	Restricted access birth certificate data from the Ohio Department of Health. This contains name and date of birth of the child, name and age of mother and father, and residential address of the mother.	To request access to this data, submit a request to the Ohio Department of Health Institutional Review Board. Details and instructions can be found here.

Social Security Death Master file	This contains the universe of reported deaths in the United States prior to 2010s for individuals with social security numbers. It contains the name, social security number, and date of death.	For access to the public SSA death master file, see instructions here.
Assorted summary statistics data	This data provides summary statistics on trends in mortality in Ohio and other states.	Panel A of table 1 uses data from BJS, Deaths in Custody Reporting Program, 2001 and 2005-2014; National Prisoner Statistics, 2001 and 2005-2014; and Federal Bureau of Prisons, 2001 and 2005-2014. Panel (c) uses CDC WONDER to measure age-adjusted mortality rate in Ohio and the rest of the United States by cause of death.

Cleaning do-files

The below table describes the do-files that are used for the cleaning in the paper. To run all of these, use master_cleaning.do to call all of the cleaning code in order.

	Code	Code description
	master_cleaning.do	Runs all of the cleaning files in the correct order
	cuyahoga/ master_clean_cuyahoga.do	Runs all of the files to clean the Cuyahoga Common Pleas and Cleveland municipal courts data
	cuyahoga/arraigned_indicator.do	Generates an indicator for whether the case was ever arraigned
	cuyahoga/case_costs_mun.do	Takes the raw municipal court data and creates outcomes of costs and fines for the case
	cuyahoga/case_dates_mun.do	Takes the raw municipal court data and cleans trial dates and docket assignments
	cuyahoga/case_dates.do	Takes the raw common pleas court data and cleans trial dates
	cuyahoga/clean_detainers.do	Takes the raw common pleas court data and cleans information on whether the defendant was already in jail
	cuyahoga/clean_judge.do	Takes the raw common pleas court data and cleans judge names and assignment order
	cuyahoga/ combine_clean_combined.do	Merges all of the interim data files created by the previous do files to create the analysis data set for Cuyahoga county
	cuyahoga/ combine_docket_entries_mun.d o	Merges interim municipal data files on outcomes and parole violations
	cuyahoga/ combine_docket_entries.do	Merges interim common pleas data files on outcomes and parole violations
	cuyahoga/date_returned.do	Formats data on date returned from parsing.
	cuyahoga/ def_characteristics_mun.do	Takes raw municipal court data and cleans defendant-level information
	cuyahoga/def_characteristics.do	Takes raw common pleas court data and cleans defendant-level information
	cuyahoga/def_charges_mun.do	Takes the raw municipal docket data and creates a file with the set of charges against the defendant in the case
	cuyahoga/def_charges.do	Takes the raw common pleas docket data and creates a file with the set of charges against the defendant in the case
	cuyahoga/ def_disposition_mun.do	Takes raw data and cleans case disposition and case transfer information
	cuyahoga/ def_docket_fulloutcomes_mun.R	Takes raw municipal docket data and cleans punishment and parole information
	cuyahoga/ def_docket_fulloutcomes.R	Takes raw common pleas docket information and cleans punishment and parole information for the 2000s

	cuyahoga/ def_docket_limitedoutcomes.R	Takes raw common pleas docket information and cleans punishment and parole information for the 1990s and early 2000s
	cuyahoga/drop_guilty.do	Takes raw common pleas data and identifies cases where the plea and arraignment occur on the same day
	cuyahoga/drop_same_plea.do	Takes raw common pleas data and identifies cases where a plea occurs on the same day as sentencing in another case
	cuyahoga/dropped_cases.do	Takes raw common pleas data and identifies cases where formal charges were dropped
	cuyahoga/extract_docket_mun.R	Reshapes municipal court data to be at the correct level for data cleaning
	cuyahoga/extract_docket.R	Reshapes common pleas court data to be at the correct level for data cleaning
	cuyahoga/find_pre_judge.do	Takes raw data and extracts first judge assigned for early cases in study period
	cuyahoga/ geocode_cuyahoga_mun.do	Merges geocoded courts records with information from the ACS about the neighborhood
	cuyahoga/judge_mun.do	Takes raw municipal court data and cleans judge names and assignment order
	cuyahoga/judge_transfers.do	Takes in judge assignment data and determines whether the case was transferred between judges
	cuyahoga/make_id.do	Generates unique identifiers for individuals
	cuyahoga/ pred_black_cuyahoga.do	Predicts defendant likelihood of being black, using voter roll information
	franklin/master_clean_franklin.do	Runs all cleaning files for Franklin county
	franklin/clean_bdays.do	Clean file of defendant birthdates
	franklin/create_geocoded_addresses.do	Output addresses to be geocoded and matched to census block groups.
	franklin/clean_closed.do	Get date when case was closed.
	franklin/clean_entries.do	Clean docket of events that happened during case.
	franklin/clean_charges.do	Clean and standardize charges.
	franklin/clean_courtrooms.do	Clean courtroom number.
	franklin/clean_codefendants.do	Identify groups of codefendants.
	franklin/clean_txt.do	Clean non-docket information from cases and merge in information about outcomes.
	franklin/clean_bdates.do	Adjusts birthdates to account for transposition mistakes that make the same individual tried multiple times appear to be a different individual.
	franklin/combine_bdates.do	Combine cleaned birthdates and rest of the data.
	franklin/clean_txt_mun.do	Clean municipal court records.
	franklin/make_id.do	Make a combined identifier to identify the same individual across Common Pleas and Municipal Court.
	franklin/combine_clean.do	Combine cleaned municipal and CP data. Construct instruments and define analysis sample.

	hamilton/ master_clean_hamilton.do	Runs all of the files to clean the Hamilton Common Pleas and Municipal courts data
	hamilton/clean_defendants.do	Takes the raw docket data and creates a file with the set of defendant characteristics in the case
	hamilton/clean_charges.do	Takes the raw docket data and creates a file with the set of charges against the defendant in the case
	hamilton/clean_outcomes.do	Takes the raw docket data and creates a file with the set of sentences against the defendant in the case
	hamilton/ create_clean_charges.do	Merges the data on charges and outcomes at the charge level to produce the outcome for each individual charge
	hamilton/ concord_ohio_charges.do	Matches the offense types in Hamilton to the rest of the county data
	hamilton/judge_assignment.do	Takes the raw docket data and creates a panel of the judges ever on the case. This also determines the randomly assigned judge on the case
	hamilton/clean_attorney.do	Takes the raw docket data and creates a dataset with the characteristics of the attorneys on the case.
	hamilton/clean_plea_fines.do	Takes the raw docket data and determines whether the case involved any pleas or fines
	hamilton/ create_geocoded_addresses.do	Takes the geocoded address for the defendant listed on the court filing and merges in ACS information about that neighborhood.
	hamilton/ create_probation_record.do	Uses the raw docket data to determine whether the defendant was incarcerated due to probation or parole violations
	hamilton/create_clean_cases.do	Merges all of the interim data files created by the previous do files to create the analysis data set for Hamilton county
	helper/ calc_prev_incar.do	Creates tempfile with previous incarceration history
	helper/ calc_while_incar.do	Creates tempfile with whether during the period of incarceration
	helper/ docket_helper.R	Set of functions to aid extracting docket information.
	helper/ make_pid.do	Creates individual specific unique IDs for defendants
	helper/ make_z.do	Creates standard leave-out judge instruments for incarceration
	helper/ predict_race.do	Predicts defendant race using their name
	matching/master_matching.do	Run all matching files
	matching/make_panel_court.do	Make panel dataset to study dynamic effects of incarceration
	matching/match_court_death.do	Match defendants to SSA death records.
	matching/match_court_death_oh.do	Match defendants to Ohio death records.
	matching/name_rarity_bc_yr.do	Helper file used to estimate likelihood name is unique among Ohio births in a year.

	panel/combine_panels.do	Combines the panel data sets from all three counties
	parsing/parse_cuya_fugitive.py	Parses fugitive status in Cuyahoga common pleas data.
	parsing/parse_cuya_returned.py	Parses dates of return in Cuyahoga common pleas data.

Analysis do-files

This table describes the do-files that are used for the analysis in the paper.

	Code	Code description
	analysis/replication_analysis.do	Creates all of the exhibits in the paper. To run all of the exhibits, set RunAll equal to 1. To run an individual exhibit, set the local at the top of the do-file equal to one for the relevant exhibit
	analysis/helper/add_locals.do	Adds local variables to the regression output
	analysis/helper/ add_locals_iv_incaranycase.do	Adds local variables to the IV regression output
	analysis/helper/ gen_label_characteristics.do	Labels covariates
	analysis/helper/ setup_courtcentered_panel.do	Sets up quarterly panel data for analysis
	analysis/helper/ setup_courtcentered_panel_sixmonths.do	Sets up 6 month panel data for analysis