README for:

"Safer in School? The Impact of Compulsory Schooling on Maltreatment and Associated Harms"

by Adam A. Dzulkipli, Nicole Black, David W. Johnston, and Leonie Segal

Overview

The code available in this replication package produces cleaned datasets for analysis and recreates all of the figures and table presented in our paper and in the accompanying Online Appendix. In particular, the replication package contains the following Stata do files:

- **"1_cleaning.do"** – Constructs the main analysis file from the raw datasets.
- **"2_main_figures.do"** – Reproduces figures presented in the main text.
- **"3_main_tables.do"** – Reproduces tables presented in the main text.
- **"4_appendix_figures.do"** – Reproduces figures presented in the Online Appendix.
- **"5_appendix_tables.do"** – Reproduces tables presented in the Online Appendix.
- **"6_intext.do"** – Reproduces additional analyses discussed in the main text.

Further details on the Stata programs are provided below. The data used, which comes from the Impacts of Child Abuse and Neglect (iCAN) study, is confidential and **not** publicly available.

Computational Requirements: The Stata code was last run using Stata MP 19.5 for Windows 11 on a Dell Precision 5570, with a total runtime of approximately ~8 hours across all do files. The computational time may vary depending on the specific machine or version of Stata used. Note that we have removed directory names from each do file to preserve individual privacy.

Data Availability

The de-identified administrative datasets used in this paper are provided by various government and semi-government agencies in South Australia (SA), including the SA Department for Education, the SA Department of Health and Wellbeing, and the SA Department for Child Protection. These datasets cover the population of children born in SA from 1986 to 2017. Data linkage across datasets was facilitated by SA-NT DataLink using probabilistic matching and clerical review to identify unique individuals across data sources.

The authors are not authorised to publicly circulate these proprietary datasets. Access to the data is possible in principle, but subject to undergoing an application review process, passing ethics clearance, and obtaining approval from the individual data custodians. Due to the sensitive nature of the data, any output generated from the data is required to be

submitted for feedback by the data custodians. For further information regarding data access, please contact author Leonie Segal (leonie.segal@unisa.edu.au).

Description of Programs

- **1_cleaning.do:**
  - This program cleans the raw data and constructs the main variables necessary for the analysis. The code produces intermediate datasets based on the various South Australian administrative datasets, and then merges these data to form the final dataset used for the analysis presented in the paper.
  - The raw iCAN data files called by the program are listed as follows:
    - **ican_cohall.dta** – Master iCAN cohort file, containing demographic characteristics and birth-related variables from the SA Births Registry and Perinatal Statistics Collection, for the cohort of all children born in SA from 1986 to 2017.
    - **educ.sav** – Data from the SA Department for Education, covering enrolments in public schools in SA from 2005 to 2017.
    - **Merged_CP_notifications_all_variables_FINAL.sav** – Data from the SA Department for Child Protection, covering all child protection notifications from 1986 to 2017.
    - **Merged_CP_OOHC_PSLK3_Final.sav** – Data from the SA Department for Child Protection, covering all out-of-home-care placements from 1990 to 2017.
    - **ican_ed_coh_cp.dta** – Data from the SA Department of Health and Wellbeing, covering all emergency department (ED) presentations between 2003 and 2017.
    - **PSLK to Absence details.sav** – Data from the SA Department for Education, covering recorded absences from school in the first two terms of the school year, between 2007 and 2017**.**
  - Runtime: ~5 minutes

- **2_main_figures.do:**
  - This program recreates all of the main figures in the main text of the paper. The figures are exported as individual PDF files, and can then be combined together manually using a typesetting software such as LaTeX or Microsoft Word.
  - Figures reproduced by this program:
    - Figure 1
    - Figure 2
    - Figure 3
  - Runtime: <1 minute

- **3_main_tables.do:**
  - This program recreates all of the main tables presented in the main text of the paper. Tables are exported as .txt files using the 'esttab' package in Stata. If a table has multiple panels, the panels are appended together using the 'append'

option in 'esttab'. The table will then require further manual editing in a typesetting software to get the table to the exact format as it is presented in the paper. The 'esttab' command can also be altered to export the tables as .tex files if so desired.

- Tables reproduced by this program:
  - Table 1
  - Table 2
  - Table 3
  - Table 4
- Runtime: <1 minute

- **4_appendix_figures.do:**
  - This program recreates all of the Appendix figures presented in the supplemental Online Appendix of the paper. The figures are exported as individual PDF files, and can then be combined together manually using a typesetting software such as LaTeX or Microsoft Word.
  - Figures reproduced by this program:
    - Appendix Figure B.1
    - Appendix Figure B.2
    - Appendix Figure B.3
    - Appendix Figure B.4
    - Appendix Figure B.5
    - Appendix Figure B.6
  - Runtime: ~1.5 hours

- **5_appendix_tables.do:**
  - This program recreates all of the Appendix tables presented in the supplemental Online Appendix of the paper. Tables are exported as .txt files using the 'esttab' package in Stata. If a table has multiple panels, the panels are appended together using the 'append' option in 'esttab'. The table will then require further manual editing in a typesetting software to get the table to the exact format as it is presented in the paper. The 'esttab' command can also be altered to export the tables as .tex files if so desired.
  - Tables reproduced by this program:
    - Appendix Table A.4
    - Appendix Table A.5
    - Appendix Table A.7
    - Appendix Table A.8
    - Appendix Table B.1
    - Appendix Table B.2
    - Appendix Table B.3
    - Appendix Table B.4
    - Appendix Table B.5
    - Appendix Table B.6
    - Appendix Table C.1

    - Runtime: ~2.5 hours

- **6_intext.do:**
    - This program recreates additional results/analyses which are discussed in the main text of the paper, but not formally presented in any table or figure.
    - Runtime: ~4 hours

Replication Instructions: The following Stata packages should first be installed as they are necessary to run all of the do files in the replication package successfully:

- ftools
- reghdfe
- estout
- rdrobust
- pzms
- rddisttestk (.ado file written by Brigham Frandsen available online from https://sites.google.com/view/brighamfrandsen/software?authuser=0)
- rddensity
- rdhonest

Code to install the above packages is provided in "**1_cleaning.do**" (commented out in case the packages are already installed). The only exception is the 'rddisttestk' package, where the ado file has to be downloaded manually from Professor Brigham Frandsen's website and placed in Stata's ado directory. Next, the relevant directories should be defined in the individual do files and the raw data placed in the designated 'raw' directory. The do files should then be run sequentially based on their numbering to build the analytical dataset and to reproduce all of the tables and figures, in both the main text and in the supplemental Online Appendix.