Replication File for "The Politics of Policy: The Initial Mass Political Effects of Medicaid Expansion in the States"

Joshua D. Clinton and Michael W. Sances

January 14, 2022

We thank Drew Stommes, Peter Aronow and Fredrik Sävje for bringing our attention to bugs in the replication file we originally posted in December 2017. This updated replication file corrects those errors. The final section of this document explains the errors in the original files, which are still available on Dataverse as a previous version.

Contents

1 Steps to Replicate **2**

2 Description of Included Files **2**

3 Restricted Data **4**

4 Software **4**

5 Errors in the Original Version **5**

1 Steps to Replicate

1. Download all files in this replication file to a new directory, e.g. *Downloads/aca/*
2. Copy *_code/DCdensity.ado* to your *ado/personal* directory and copy *_code/cgmreg.ado* to your *ado/plus/c* directory. (See Section 4.)
3. Ensure you also have the packages listed under Section 4 installed.
4. Open Stata, set working directory to the directory chosen in step 1.
5. Following the instructions in Section 3, obtain the restricted data and run */turnout/setup turnout.do*.
6. Comment out lines 213 to 218 of *_code/setup.do.*
7. Run *_code/setup.do* to create the main data set *dataset.dta* with actual turnout and registration included.
8. To replicate all figures and tables, run *_code/replicate.do.* To replicate a particular figure/table, run *_code/figure X.do* or *_code/table X.do*

Note that the .do files *replicate.do*, *figure X.do*, and *table X.do* will at last partially execute if you just complete step 1 above, and will fully execute if you complete steps 1, 2, 3, and 4. However, without performing steps 5, 6, and 7, the .do files will make calls to the included *dataset.dta* file, which includes simulated values for the turnout and participation variables. See lines 213 to 218 of *_code/setup.do* and Section 2. In this case, your results will not match those in the article.

2 Description of Included Files

- *dataset.dta* – Required by all figure/table .do files. The version included with the replication package includes simulated data for the turnout and participation variables. See lines 213 to 218 of *_code/setup.do* and Section 3.
- *dataset_coords.dta* – Required by *figure 1.do*. Contains GIS data for states for the purposes of mapping.
- *dataset_distanceforplacebo.dta* – Required by *figure 6.do*. Contains raw distance between each county and state border. *figure 6.do* randomly reassigns expansion status to states, and then re-generates the distance between a county and states with a different expansion status than the county's home state.
- *_code/* – Contains .do files for replication.
  - → *cgmreg.ado* – Source code needed for the *cgmwildboot* command used in multiple .do files. This file must be copied to your *ado/plus/c* directory. (See Section 4).

→ *DCdensity.ado* – Source code for the density test conducted for Figure A5. This file must be copied to your *ado/personal* directory. (See Section 4).

→ *figure 1.do*, ..., *figure A18 A19.do* – Reproduces each figure (or pair of figures) from the main text and appendix.

→ *replicate.do* – Loops over the individual figure and table .do files and runs each. It also first sets the graphics scheme.

→ *setup.do* – Creates the *dataset.dta* file from the source data sets. Will not run if *turnout/turnout.dta* is missing. Lines 213-218 replace the actual turnout and registration variables with simulated data as we are prohibited from sharing the real data. See Section 3.

→ *table 1.do, ..., table A20 A21.do* – Reproduces each table (or pair of tables) from the main text and appendix. Note there is no file for Table A13, as this table is manually entered into the appendix .tex file.

- *_output/* – Destination for graphical and tabular outputs created by figure/table .do files. Empty by default.
- *demographics/* – Source files for county demographics.
- *demvoteshare/* – Source files for county Democratic vote.
- *distance/* – Source files for distance between county and state borders.
- *electionsheld/* – Sources files for elections held in states in different years.
- *eligibility/* – Source files for Medicaid eligibility limits by state.
- *exchanges/* – Source files for characteristics of state private health insurance exchanges.
- *expansion/* – Source files for state Medicaid expansion status.
- *gis/* – Source files for state and county GIS data.
- *migration/* – Source files for migration data.
- *turnout/* – Source files for turnout and registration data. Directory does not include any data by default; see Section 3.
- *uninsured/* – Source files for the share with health insurance by county.
- *vap/* – Source files for voting age population by county.

3 Restricted Data

We purchased data on voter turnout and voter registration from Dave Leip's Atlas of U.S. Elections (see `https://uselectionatlas.org/`). According to the terms of use, we are unable to share these data. Users wishing to replicate our results must purchase the following county-level data sets from Dave Leip's web site:

- *TurnoutData2004.xlsx*
- *TurnoutData2006.xlsx*
- *TurnoutData2008.xlsx*
- *TurnoutData2010.xlsx*
- *TurnoutData2012.xlsx*
- *TurnoutData2014.xlsx*
- *TurnoutData2016.xlsx*

After obtaining these files, save all of them to the *turnout* directory in our replication package. You can then run *turnout/setup turnout.do*, which will create the file *turnout/turnout.dta*. This will allow you to run *_code/setup.do* and create the *dataset.dta* file called by all the figure and table .do files.

If you do not complete the steps described in this section, the version of *dataset.dta* will include simulated data for turnout and registration as placeholders.

1 to go back to Section 1.

4 Software

Stata was used to produce all figures and tables. We primarily used Version 15. We have successfully ran the replication files on Versions 14 and 17 as well.

The following additional packages are required to run the Stata .do files. Install these using *ssc install [package name]* unless otherwise noted with a * below.

- * *cgmreg* – Source code required by the *cgmwildboot* command which is used for regressions in the table .do files. Must be installed manually by coping the included file at *_code/cgmreg.ado* to your *ado/plus/c* directory.

- *clustse* – Package that includes the *cgmwildboot* command for the regressions used in the table .do files.
- * *DCdensity* – Source code for running the density test for Figure A5. Must be installed manually by copying the included file at *_code/DCdensity* to your *ado/personal* directory.
- *estout* – For outputting certain tables as .tex files.
- *listtex* – For outputting most of the regression tables as .tex files.
- *spmap* – Used by *figure 1.do* to create a map of border states.
- *sxpose* – Command for transposing string data (the built-in *xpose* command only works for numerical data) for tabular output of regression results.
- *unique* – Command for counting the number of unique observations. Required by *cgmwildboot*.

1 to go back to Section 1.

5 Errors in the Original Version

The original replication files were posted to `http://dx.doi.org/10.7910/DVN/X0ZPH4` in December, 2017. The names of the files posted were:

- *dataset_coords.dta*
- *dataset_dataverse.dta*
- *dataset_distanceforplacebo.dta*
- *figure 1 - map of expansion states.do*
- *figure 2 - change in insurance.do*
- *figure 3 - raw data.do*
- *figure 4 - rd registration by uninsured.do*
- *figure 5 - rd turnout by uninsured.do*
- *figure 6 - permutation test.do*
- *figure 7 - trends.do*
- *figure A1 - estimates by bandwidth.do*

- *figure A2 - change in insurance all counties.do*
- *figure A3 - rd registration by uninsured all counties.do*
- *figure A4 - rd turnout by uninsured all counties.do*
- *figure A5 - mccrary.do*
- *figure A6 - migration by state.do*
- *figure A7 - change in insurance no interaction.do*
- *figure A8 - rd registration no int.do*
- *figure A9 - rd turnout no int.do*
- *figure A10 - all placebos.do*
- *figure A11 - continuity in observables.do*
- *figure A12 - trends uninsurance.do*
- *figure A13 - state pair compare.do*
- *figure A14 A15 - change in insurance by party.do*
- *figure A16 A17 - rd registration by party.do*
- *figure A18 A19 - rd turnout by party.do*
- *replicate.do*
- *setup leip.do*
- *table 1 - rd insurance.do*
- *table 2 - rd registration.do*
- *table 3 - rd turnout.do*
- *table A1 - rd insurance all counties.do*
- *table A2 - rd registration all counties.do*
- *table A3 - rd turnout all counties.do*
- *table A4 - rd insurance cont.do*
- *table A5 - rd registration cont.do*
- *table A6 - rd turnout cont.do*
- *table A7 - robustness checks uninsured.do*

- *table A8 - robustness checks registration.do*
- *table A9 - robustness checks turnout.do*
- *table A10 - rd insurance no int.do*
- *table A11 - rd registration no int.do*
- *table A12 - rd turnout no int.do*
- *table A14 - within state analysis.do*
- *table A15 - within state analysis with fe.do*
- *table A18 A19 - rd insurance by party.do*
- *table A20 A21 - rd registration with interaction by party.do*
- *table A22 A23 - rd turnout with interaction by party.do*

Error #1: No documentation file.

The first error was that we did not include any documentation or "ReadMe" file explaining the contents of the replication package and/or how to use the contents to reproduce the figures and tables in the paper.

The intended purpose of each of the files was as follows:

- *dataset_coords.dta* – Same as described in Section 2.
- *dataset_dataverse.dta* – A version of the main analysis data set that omits the proprietary turnout and registration variables.
- *dataset_distanceforplacebo.dta* – Same as described in Section 2.
- *figure 1 - map of expansion states.do*, ..., *figure A18 A19 - rd turnout by party.do* – Same as described in Section 2.
- *replicate.do* – (1) Runs *setup leip.do* to create the proprietary turnout and registration variables and save as *dataset_leip.dta*; (2) merges *dataset_leip.dta* file with *dataset_dataverse.dta* and save as *dataset_dataverse_with_leip.dta*; (3) run all the figure/table .do files.
- *setup leip.do* – Read in the (unincluded) raw proprietary data spreadsheets (see Section 3) and save as *dataset_leip.dta.*
- *table 1 - rd insurance.do*, ..., *table A22 A23 - rd turnout with interaction by party.do* – Same as described in Section 2.

The (unstated) steps required for replication were (assuming all third-party packages were installed):

1. Download all the replication files to a new directory, e.g. *Downloads/aca/*
2. Open Stata and set the working directory to the directory created in step 1.
3. Obtain the proprietary election files and download them to a new subdirectory *leip/*.
4. Run *replicate.do* to the end to build *dataset_dataverse_with_leip.dta* reproduce all figures and tables. Or, run *replicate.do* to the necessary line to build the data set only, then reproduce individual figures and tables by running the .do files *figure X.do* and *table X.do*.

Upon our attempt to complete these steps in late 2021 and early 2022, we encountered the following errors.

Error #2: Typo in replicate.do.

The original version of *replicate.do* began,

```
1  set more off
2
3  set scheme s1manual
4
5  ///// This chunk makes sure all .dta files are version 13 /////
6
7  local myfilelist : dir . files "*.dta"
8
9  foreach f in `myfilelist' {
10         di "`f'"
11         use "`f'", clear
12         describe
13         saveold "`f'", version(13) replacec
14 }
15
16 ///// This chunk cleans and merges the proprietary Leip data /////
17
18 do "setup leip"
```

Line 13 contained a typo: the subcommand *replace* was entered as *replacec*, which is unrecognized by Stata. The result is that users attempting to run *replicate.do* would receive an error message; *dataset_dataverse_with_leip.dta* is not created; and the figures and tables are not reproduced.

(The purpose of lines 7 through 14 was to cycle through all *.dta* files and check that they are saved in Stata Version 13 format. Presumably this was included because at the time, the co-authors had different versions of Stata. For instance, Sances may have had Version 13 and Clinton may have had Version 14. If any .dta files were created on Clinton's machine, by default they would be saved in Version 14 format and would not be accessible by Sances. See for example `https://ciser.cornell.edu/working-with-different-stata-versions-commands-and-data/`. In retrospect, these details are irrelevant for the replication file user and so the updated files do not use the *saveold* command.)

Error #3: Calls to missing data files.

The code files *figure X.do* and *table X.do* all made calls to *dataset_dataverse_with_leip.dta,* which was not included. Our intended steps for replication required users to first run *replicate.do* and create this file. However, (1) users could not have known this given the lack of documentation (*Error #1*), and (2) *replicate.do* included a coding error which prevented it from running (*Error #2).*

Error #4: Slight differences in treatment of extreme observations.

Footnote 15 of the article (page 175) reads:

"For each dependent variable, we code observations greater than the 99th percentile or less than the first percentile as missing. This eliminates implausibly large values of changes in participation, as high as 50% or even 90%, that sometimes occur in less populous counties. The results are substantively similar if we code these extreme observations as 100%."

In the original replication file, this coding occurs in the following lines of *replicate.do*:

```
26  ///// This chunk makes the participation variables /////
27
28  forvalues year = 2004(2)2016 {
29
30          * turnout
31          gen turnout`year' = totalvoting`year'/vap`year'*100
32          replace turnout`year' = . if turnout`year' == 0 | round(turnout`year') == 0
33          if `year' > 2004 {
34                  local year_2 = `year'-2
35                  gen dturnout`year'`year_2' = turnout`year' - turnout`year_2'
36                  su dturnout`year'`year_2', det
37                  replace dturnout`year'`year_2' = . if dturnout`year'`year_2' > r(p99) | dturnout`year'`year_2' < r(p1)
38          }
```

Specifically, the code loops over dependent variables and (1) finds the 1st and 99th percentile of the variable (line 36); (2) codes the variable as missing if it is greater than the 99th percentile or lower than the 1st percentile (line 37).

In the code files used in the production of the article, the 1st and 99th percentiles were defined with respect to the distribution of the dependent variables across all states. In the code files in the original replication files, the 1st and 99th percentiles were defined with respect to the distribution of the dependent variables across the 32 "border" states only. The code files for the article began by creating a data set using variables from all states, trimming the dependent variables based on the distribution of all the states, then subsetting to the 32 border states; the replication file began with *dataset_dataverse.dta* which was already restricted to the 32 border states, and trimmed the dependent variables based on the distribution within these 32 states.

We discovered this error by:

1. Downloading the original replication file
2. Correcting the typo described in *Error #2*
3. Copying the proprietary .xlsx files to a new directory *leip/*

	Registration 2014-2010				Registration 2016-2012			
	Published	Replication	Published	Replication	Published	Replication	Published	Replication
Medicaid	−0.35	-0.44	−0.27	-0.35	−0.03	0.08	0.09	0.2
expansion	[−2.76, 1.90]	[-2.71, 1.95]	[−2.14, 1.69]	[-2.27, 1.55]	[−1.70, 1.70]	[-1.62, 1.80]	[−1.66, 1.68]	[-1.61, 2.02]
High eligibility	−0.44	-0.55	−0.49	-0.6	0.03	0.1	−0.12	-0.07
	[−1.92, 1.08]	[-2.07, 0.79]	[−1.60, 0.70]	[-1.89, 0.60]	[−1.26, 1.27]	[-1.25, 1.27]	[−1.28, 1.10]	[-1.32, 1.21]
Expansion	3.6	3.94	2.7	2.99	4.09	4.23	3.93	4.09
X eligibility	[−0.17, 7.35]	[0.06, 7.64]	[0.51, 4.90]	[0.76, 5.15]	[1.93, 6.35]	[2.01, 6.57]	[1.67, 6.12]	[1.80, 6.67]
Number of Counties	1208	1213	1197	1202	1264	1266	1253	1255
Number of States	28	28	28	28	30	30	30	30
Window	100 miles	100 miles	100 miles	100 miles	100 miles	100 miles	100 miles	100 miles
Covariates	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.09	0.1	0.25	0.25	0.1	0.11	0.13	0.14

Table 1: Replication of Table 2 in the article using original replication data.

4. Running lines 1 through 112 of *replicate.do*

5. Running the code to replicate the figures and tables in the main text and noticing slight discrepancies in point estimates and numbers of observation for Table 2 and Table 3. See comparisons in the tables below.

We have verified that the updated replication files eliminate these discrepancies.

	Turnout 2014-2010				Turnout 2016-2012			
	Published	Replication	Published	Replication	Published	Replication	Published	Replication
Medicaid	0.74	0.77	1.3	1.34	0.27	0.33	0.31	0.36
expansion	[−3.11, 4.57]	[-2.96, 4.30]	[−1.71, 4.44]	[-2.14, 4.54]	[−1.07, 1.74]	[-1.05, 1.74]	[−0.74, 1.25]	[-0.56, 1.40]
High eligibility	−3.26	-3.24	−2.23	-2.25	0.99	1	0.61	0.7
	[−5.79, −0.53]	[-5.77, -0.86]	[−4.05, −0.40]	[-4.06, -0.33]	[0.19, 1.88]	[0.02, 2.04]	[−0.07, 1.31]	[0.10, 1.40]
Expansion	3.96	3.78	2.36	2.27	0.14	0.13	−0.17	-0.21
X eligibility	[0.32, 8.05]	[0.39, 7.42]	[−0.63, 5.13]	[-0.58, 5.06]	[−1.53, 1.63]	[-1.57, 1.84]	[−1.30, 1.01]	[-1.41, 0.79]
Number of Counties	1320	1325	1309	1314	1320	1325	1309	1314
Number of States	32	32	32	32	32	32	32	32
Window	100 miles	100 miles	100 miles	100 miles	100 miles	100 miles	100 miles	100 miles
Covariates	No	No	Yes	Yes	No	No	Yes	Yes
R-squared	0.12	0.12	0.39	0.39	0.09	0.09	0.32	0.33

Table 2: Replication of Table 3 in the article using original replication data.