Replication of a Research Claim from Kavanagh et al. (2020), from medRxiv BNrQ g66z

Replication Team Andrew Tyner, Center for Open Science

Replication Team Nick Huntington-Klein, Seattle University

Research Scientist Nick Fox, Center for Open Science

Last Updated 2020-11-05

Claim Summary.

Greater Republican political orientation is associated with reduced social distancing among U.S. counties in Kavanagh, Goel, and Venkataramani (2020). This reflects the following statement from the paper's abstract: "Using 15–17 million anonymized cell phone records, we find that lower per capita income and greater Republican orientation were associated with significantly reduced social distancing among U.S. counties."

Replication Criteria.

At the level of U.S. counties, support for Donald Trump in the 2016 presidential election will be negatively associated with social distancing behavior.

Replication Result.

The regression table for the key test of interest is shown below. This uses data from MIT Election Science Lab (2018), the American Community Survey, and Bureau of Transportation Statistics (2020). The key coefficient is for `Share of Trump Voters` This table can also be found here.

	March 19-April 1
Income per Capita (Thousands)	0.252 ***
	(0.053)
Share of Trump Voters	-0.078 ***
	(0.018)
Percent Male	-0.368 ***
	(0.079)
Percent Black	-0.228 ***

	(0.019)
Percent Hispanic	-0.023
	(0.019)
Percent with College Degree	0.340 ***
	(0.037)
Percent in Retail	0.155 **
	(0.060)
Percent in Transportation	0.030
	(0.072)
Percent in Health / Ed / Soc. Svcs	-0.004
	(0.037)
Percent Rural	-0.051 ***
	(0.007)
Percent Age 5-9	0.837
	(1.238)
Percent Age 10-14	1.063
	(1.238)
Percent Age 15-19	0.513
	(1.230)
Percent Age 20-24	0.900
	(1.232)
Percent Age 25-34	0.938
	(1.229)
Percent Age 35-44	1.289
	(1.232)

Percent Age 45-54	2.089
	(1.233)
Percent Age 55-59	0.848
	(1.229)
Percent Age 60-64	0.971
	(1.238)
Percent Age 65-74	1.310
	(1.229)
Percent Age 75-84	-0.066
	(1.241)
Percent 85+	-0.667
	(1.241)
N	3076
R2	0.662

*** p < 0.001; ** p < 0.01; * p < 0.05.

More-positive numbers indicate more stay-at-home activity. State fixed effects included.

The coefficient on the key variable of interest, Share of Trump Voters, is: $\hat{\beta} = -.078$, $t = -.078/.018 = -4.333$, $p = .000015$, based on a sample of 3,076 observations. This translates into a Cohen's f^2 value of 0.026.

This sample size is one observation smaller than the preregistered sample size due to some changes in the data construction leading to a missing observation in the percent-in-each-industry variables. This surpasses the Stage 1 sample size of 2990 detailed in the power analysis component of the OSF project.

This replicates the original result. The dependent variable is of an inverted scale to the original study, and so the positive significant coefficient in the original study is consistent with the negative significant coefficient in this replication.

We can go further and scale the effect by a one-IQR change in the Trump vote share:

```
##
##   Simultaneous Tests for General Linear Hypotheses
```

```
##
## Fit: felm(formula = formula_maker("prop_home_change_March", flat_data,
##     nr), data = flat_data)
##
## Linear Hypotheses:
##                                      Estimate Std. Error z value Pr(>|z|)
## 0.204829092905961 * trump_share == 0 -0.01589    0.00371  -4.283 1.84e-05
***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

We find a 1.59 percentage point decline in social distancing behavior, as opposed to 4.1 percentage points in the original study. So the effect is consistent but of reduced size.

Deviations from preregistration.

There were no deviations from the preregistration plan, or changes to the preregistered script (except to remove the line limiting analysis to a random 5% subset of the data, and the aforementioned missing observation).

Deviations from the original study.

The main deviation from the original study is the use of social distancing measurements from a different source. The original study used private data from Unacast, while the replication uses data from the Bureau of Transportation Statistics. So data quality issues will differ between the two. For example, the data dictionary highlights two observations (not in the replication sample) that are negative but should not be. The preregistration highlights that the replication data is marked as "experimental, and may not meet all quality standards."

Additional Analyses

The original study used a calculation of `percent_rural` that was likely erroneous. It used percentage of housing that is rural. Percent of population that is rural was intended, and this appears to be an error. The error is fixed in the below table.

	March 19-April 1
Income per Capita (Thousands)	0.257 ***
	(0.053)
Share of Trump Voters	-0.077 ***
	(0.018)
Percent Male	-0.394 ***

	(0.079)
Percent Black	-0.227 ***
	(0.019)
Percent Hispanic	-0.023
	(0.019)
Percent with College Degree	0.343 ***
	(0.037)
Percent in Retail	0.163 **
	(0.060)
Percent in Transportation	0.033
	(0.072)
Percent in Health / Ed / Soc. Svcs	-0.002
	(0.038)
Percent Rural	-0.047 ***
	(0.007)
Percent Age 5-9	0.871
	(1.240)
Percent Age 10-14	1.109
	(1.239)
Percent Age 15-19	0.540
	(1.232)
Percent Age 20-24	0.939
	(1.234)
Percent Age 25-34	0.996
	(1.231)

Percent Age 35-44	1.333
	(1.234)
Percent Age 45-54	2.130
	(1.234)
Percent Age 55-59	0.880
	(1.231)
Percent Age 60-64	0.994
	(1.240)
Percent Age 65-74	1.343
	(1.231)
Percent Age 75-84	-0.027
	(1.242)
Percent 85+	-0.624
	(1.243)
N	3076
R2	0.661

*** p < 0.001; ** p < 0.01; * p < 0.05.

More-positive numbers indicate more stay-at-home activity. State fixed effects included.

This result is still consistent with the original result.

I also perform two additional sets of additional results. First, I extend the original time frame to see if the result changes if using output data by August 16-29 compared to the same pre-COVID distancing rates, rather than in March. These results are in the following table

	Aug. 16-29, Housing Rural	Aug. 16-19, Pop. Rural
Income per Capita (Thousands)	0.382 ***	0.390 ***

	(0.058)	(0.059)
Share of Trump Voters	-0.177 ***	-0.176 ***
	(0.020)	(0.020)
Percent Male	-0.287 **	-0.331 ***
	(0.088)	(0.087)
Percent Black	-0.205 ***	-0.205 ***
	(0.021)	(0.021)
Percent Hispanic	-0.051 *	-0.052 *
	(0.020)	(0.021)
Percent with College Degree	-0.031	-0.029
	(0.041)	(0.041)
Percent in Retail	0.451 ***	0.458 ***
	(0.066)	(0.066)
Percent in Transportation	0.064	0.069
	(0.080)	(0.080)
Percent in Health / Ed / Soc. Svcs	0.215 ***	0.218 ***
	(0.041)	(0.042)
Percent Rural (housing)	-0.116 ***	
	(0.008)	
Percent Rural (pop)		-0.112 ***
		(0.007)
Percent Age 5-9	2.875 *	2.957 *

	(1.369)	(1.372)
Percent Age 10-14	3.089 *	3.198 *
	(1.368)	(1.372)
Percent Age 15-19	2.692 *	2.754 *
	(1.360)	(1.363)
Percent Age 20-24	3.076 *	3.158 *
	(1.362)	(1.365)
Percent Age 25-34	2.904 *	3.019 *
	(1.359)	(1.362)
Percent Age 35-44	3.332 *	3.414 *
	(1.362)	(1.365)
Percent Age 45-54	4.080 **	4.173 **
	(1.363)	(1.366)
Percent Age 55-59	2.600	2.681 *
	(1.359)	(1.362)
Percent Age 60-64	3.070 *	3.131 *
	(1.369)	(1.372)
Percent Age 65-74	2.989 *	3.068 *
	(1.359)	(1.362)
Percent Age 75-84	2.470	2.558
	(1.371)	(1.375)
Percent 85+	2.459	2.547
	(1.372)	(1.375)
N	3076	3076
R2	0.601	0.599

*** p < 0.001; ** p < 0.01; * p < 0.05.

More-positive numbers indicate more stay-at-home activity. State fixed effects included.

Results are much stronger by August than in the original March results.

Lastly, I intended in the preregistration to test the robustness of the original results by adding a control for spatial autocorrelation. Both stay-at-home behavior and Trump support are strongly geographically clustered, in a way that is unlikely to be handled by fixed effects for state. I calculate county neighbors using 5-nearest-neighbors clustering on county centroid latitude and longitude. Then, I estimate a standard spatial autocorrelation model, repeating the original analysis but with a spatial autocorrelation term.

However, when switching from the 5% sample to the full sample, the spatial autocorrelation analysis stopped working properly and would require changing code and procedures, so I do not include those results.

Description of materials provided.

The following materials are publicly available on the OSF project site: https://osf.io/zx7dn/files/

- R script for compiling data together: **kavanagh_g66z_data_prep.R**
- Data dictionary: **kavanagh_g66z_data_dictionary.tsv**
- Data files: **transportation.csv** and **county_variables.csv**
- File for combining the data files and running analysis: **kavanagh_analysis.R**
- File describing the construction of the data set: **README.txt**
- Main regression results: **main_regression_table.html**

Citations.

Bureau of Transportation Statistics. 2020. “Trips by Distance.” https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv.

Kavanagh, Nolan M, Rishi R Goel, and Atheendar S Venkataramani. 2020. “Association of County-Level Socioeconomic and Political Characteristics with Engagement in Social Distancing for Covid-19.” *medRxiv*.

MIT Election Science Lab. 2018. “County Presidential Election Returns 2000-2016.” Harvard.