Supporting Information for "Differential Registration Bias in Voter File Data: A Sensitivity Analysis Approach"

Brendan Nyhan* Christopher Skovron† Rocío Titiunik‡

Contents

*Professor, Department of Government, Dartmouth College, nyhan@dartmouth.edu.

†Ph.D. Candidate, Department of Political Science, University of Michigan, cskovron@umich.edu.

‡James Orin Murfin Associate Professor, Department of Political Science, University of Michigan, titiunik@umich.edu.

S1 Quality of Catalist data

Ansolabehere and Hersh (2010) use Catalist data to analyze the quality of state voter files and find that "Identifying information such as birthdates are generally well collected." They do identify some problems with missing birth dates and unusual concentrations of voters with particular birth dates but these should not affect the validity of our design.[1] Catalist not only cleans and processes data from state voter files, which includes tracking individuals who move between states and/or are purged from voter files, but fills in exact birth dates from commercial sources when possible for states that only release month of birth, allowing us to use exact birthdates even in states that do not release them.

S2 Excluded Catalist data in Study 1

As mentioned in the paper, we excluded some individuals who were in the raw data provided to us by Catalist. Table S1 describes the observations excluded by category.

Table S1: Observations excluded from analysis in Study 1

Description	Observations
Original dataset	57,031
After dropping missing exact birthdates	54,332
After dropping outside birth targets	51,705
After dropping those recorded as voting when they should have been ineligible	51,472
After dropping those with no registration year listed	49,271

[1] The only unusual date within our window is November 11, which they find to be unusually prevalent in Texas, but we observe no evidence of a problem in our data (results available upon request).

S3 Balance checks for Study 1

We compare the demographic characteristics of just-eligible and just-ineligible voters for Study 1 in Table S2 using covariates in the Catalist data, which combines public and commercial records of gender, marital status, race/ethnicity, and religious affiliation. In the pooled data, the differences in means are small and generally not significant despite the very large sample size.[2] As we show in the paper, though, this seeming balance may mask consequential differences between the two groups in registration patterns.

[2]There are a few imbalances in the 1990 cohort, which may be the result of the shorter interval between the treatment election for this cohort (2008) and the year of data collection (2011).

Table S2: Balance statistics in Catalist data

All	Treatment	Control	p-value
Male	0.464	0.457	0.12
Married	0.095	0.095	0.84
Black	0.160	0.153	0.04
White	0.638	0.645	0.07
Hispanic	0.156	0.158	0.52
Catholic	0.265	0.272	0.09
Protestant	0.275	0.276	0.85
1986	Treatment	Control	p-value
Male	0.460	0.455	0.46
Married	0.126	0.123	0.48
Black	0.156	0.155	0.79
White	0.644	0.647	0.68
Hispanic	0.157	0.154	0.67
Catholic	0.262	0.267	0.50
Protestant	0.283	0.278	0.42
1988	Treatment	Control	p-value
Male	0.472	0.457	0.04
Married	0.085	0.089	0.33
Black	0.161	0.156	0.39
White	0.636	0.645	0.27
Hispanic	0.153	0.153	0.98
Catholic	0.265	0.268	0.67
Protestant	0.274	0.268	0.41
1990	Treatment	Control	p-value
Male	0.461	0.462	0.85
Married	0.070	0.061	0.02
Black	0.164	0.148	0.01
White	0.632	0.645	0.12
Hispanic	0.159	0.173	0.03
Catholic	0.267	0.285	0.02
Protestant	0.269	0.286	0.02

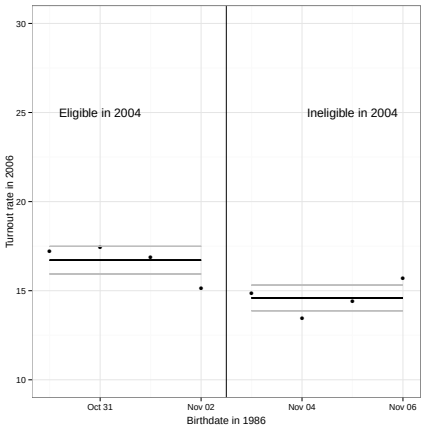
2011 Catalist data; n = 49,271 (1986: 18,326; 1988: 17,153; 1990: 13,792)

S4 RD plots for Study 1

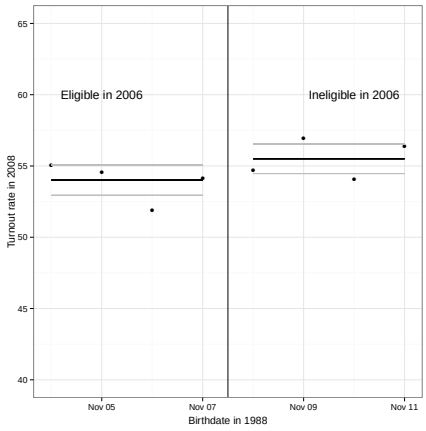
S4.1 Turnout-to-registration rates

As is conventional in RD analyses, we plot raw turnout rates among registrants binned by date of birth. Figure S1 illustrates how turnout varies by eligibility in the election after treatment (which we call E2), while Figure S2 presents corresponding results for the next two elections (E3 and E4).

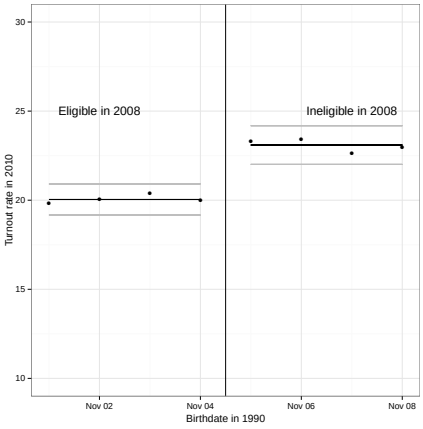
Figure S1: RD effects of voting eligibility on turnout in subsequent election



(a) Turnout effects in 2006 for 1986 cohort (E2)



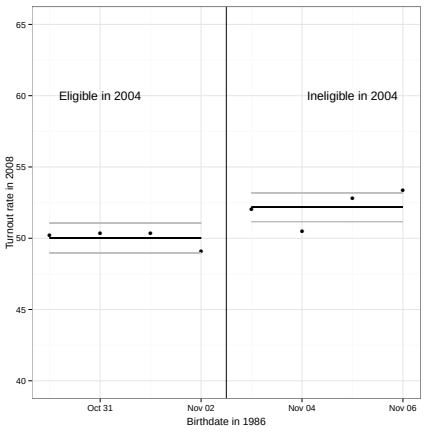
(b) Turnout effects in 2008 for 1988 cohort (E2)



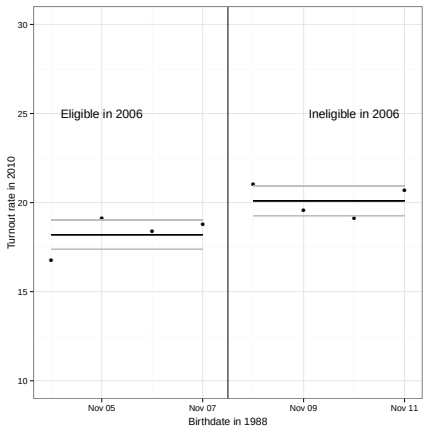
(c) Turnout effects in 2010 for 1990 cohort (E2)

2011 Catalist data; n = 49,271 (1986: 18,326; 1988: 17,153; 1990: 13,792). 95% confidence intervals in parentheses. Lines represent means and 95% confidence intervals for just-eligibles and just-ineligibles.

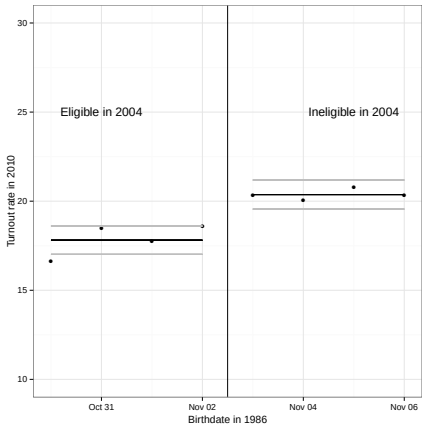
Figure S2: RD effects of voting eligibility on turnout in second and third subsequent elections



(a) Turnout effects in 2008 for 1986 cohort (E3)



(b) Turnout effects in 2010 for 1988 cohort (E3)



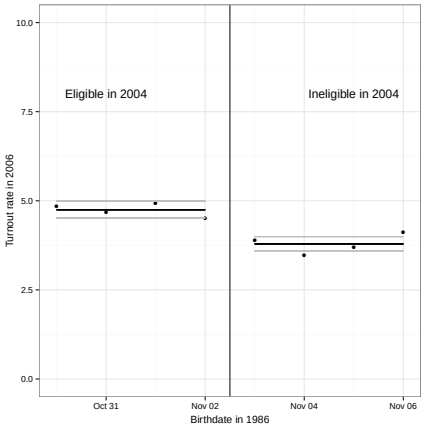
(c) Turnout effects in 2010 for 1986 cohort (E4)

2011 Catalist data; n = 49,271 (1986: 18,326; 1988: 17,153; 1990: 13,792). 95% confidence intervals in parentheses. Lines represent means and 95% confidence intervals for just-eligibles and just-ineligibles.

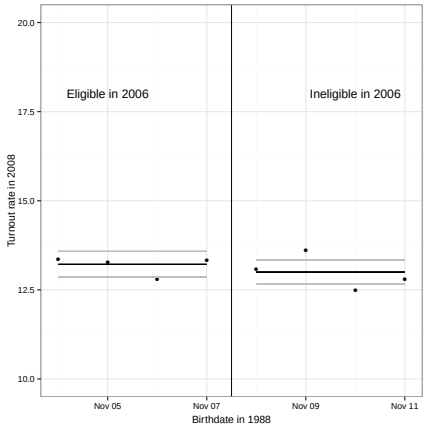
S4.2 Turnout-to-births rates

Figures S3 and S4 plots the raw data for turnout rates by date of birth when adjusted by birth totals rather than the number of registrants in the data.

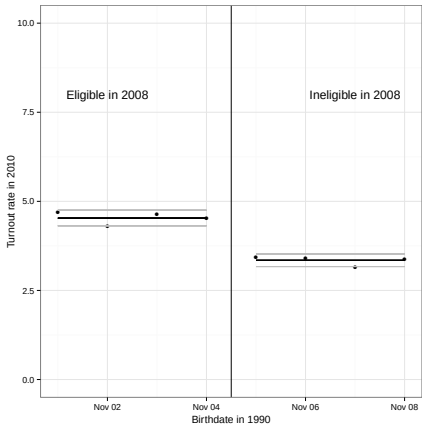
Figure S3: RD estimates of voting eligibility effects on population turnout rates (E2)



(a) Turnout effects in 2006 for 1986 cohort (E2)



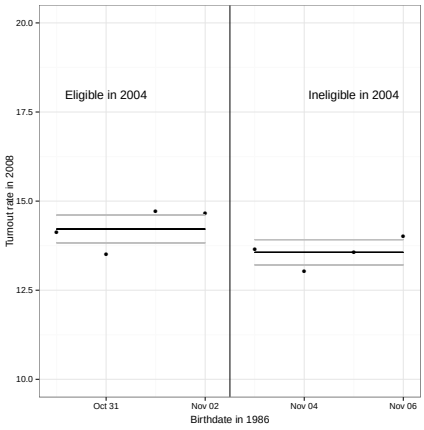
(b) Turnout effects in 2008 for 1988 cohort (E2)



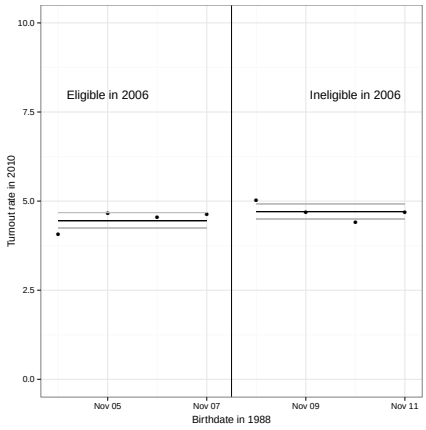
(c) Turnout effects in 2010 for 1990 cohort (E2)

2011 Catalist data; n = 49,271 (1986: 18,326; 1988: 17,153; 1990: 13,792). 95% confidence intervals in parentheses. Lines represent means and 95% confidence intervals for just-eligibles and just-ineligibles.

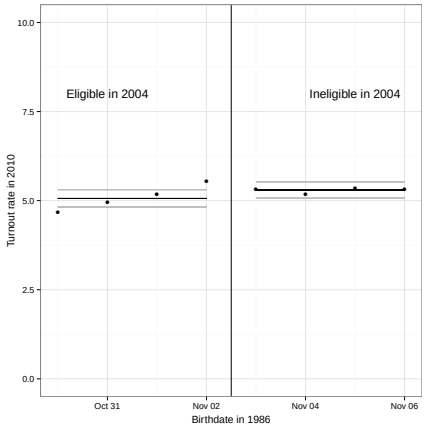
Figure S4: RD estimates of voting eligibility effects on population turnout rates (E3–E4)



(a) Turnout effects in 2008 for 1986 cohort (E3)



(b) Turnout effects in 2010 for 1988 cohort (E3)



(c) Turnout effects in 2010 for 1986 cohort (E4)

2011 Catalist data; n = 49,271 (1986: 18,326; 1988: 17,153; 1990: 13,792). 95% confidence intervals in parentheses. Lines represent means and 95% confidence intervals for just-eligibles and just-ineligibles.

S5 Results for Study 1 in states without preregistration

Holbein and Hillygus (2015) argue that preregistration increases mobilization for just-ineligible voters, who are exposed to the opportunity to preregister. At the time of the elections that we consider in Study 1, only Florida and Hawaii allowed voters to pre-register when they were 17 years old. Table S3 therefore replicates the main turnout-to-birth analysis of Study 1 (Table 7 in the main paper) excluding Florida and Hawaii from both the turnout and birth counts. The results are largely unchanged from the original analysis.

Table S3: Turnout rates by voting eligibility as a proportion of births, excluding preregistration states

	A. 1986 cohort (first election for just-eligibles: 2004 presidential)		
	E2 (2006 midterm)	E3 (2008 presidential)	E4 (2010 midterm)
Eligibility effect	0.89	0.59	-0.28
	[0.57, 1.21]	[0.06, 1.12]	[-0.62, 0.07]
Control group	3.79	12.93	5.16
	B. 1988 cohort (first election for just-eligibles: 2006 midterm)		
	E2 (2008 presidential)	E3 (2010 midterm)	E4 (2012 presidential)
Eligibility effect	0.32	-0.18	-
	[-0.18 , 0.83]	[-0.50, 0.13]	-
Control group	12.38	4.54	
	C. 1990 cohort (first election for just-eligibles: 2008 presidential)		
	E2 (2010 midterm)	E3 (2012 presidential)	E4 (2014 midterm)
Eligibility effect	1.20	-	-
	[0.91, 1.49]	-	-
Control group	3.26	-	-

2011 Catalist data; n = 44,167 (1986: 16,385; 1988: 15,371; 1990: 12,411). Brackets show 95% confidence intervals based on a differences-in-means Wald test.

S6 Birth, registration, and vote totals: Study 1

This table shows the total number of births, registered voters, and votes cast in the first election after the eligibility treatment election.

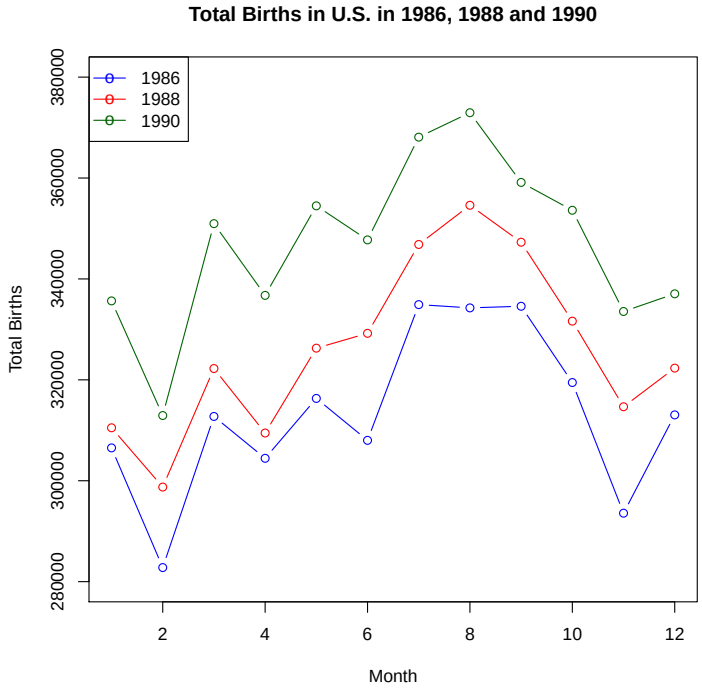
Table S4: Births, registration, and vote totals

	1986 treated	1986 control	1988 treated	1988 control	1990 treated	1990 control
Births	31,476	36,096	34,083	37,646	35,537	39,801
Registration	8,945	9,381	8,340	8813	8,029	5,763
Votes	1,495	1,369	4,504	4,892	1,609	1,331

Registration and vote totals from 2011 Catalist data; n = 49,271 (1986: 18,326; 1988: 17,153; 1990: 13,792).

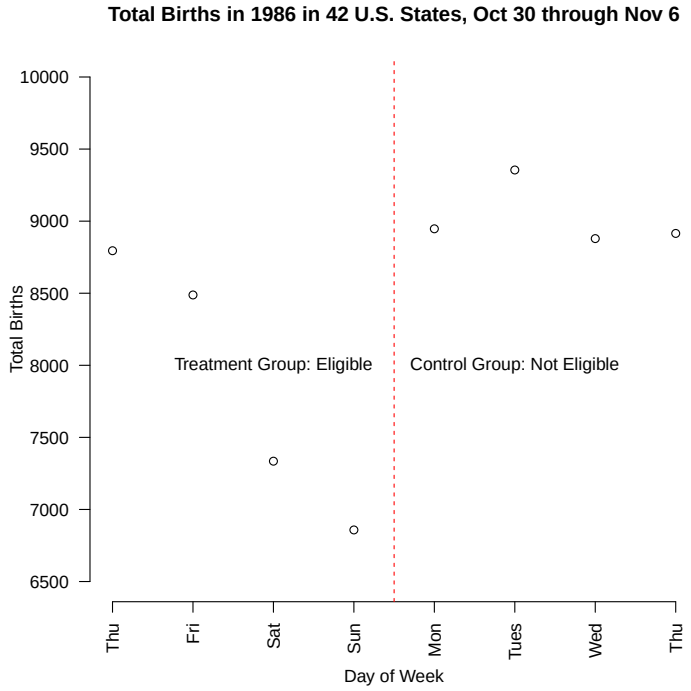
S7 Total births by day and month

Figure S5: Total births by month in the United States

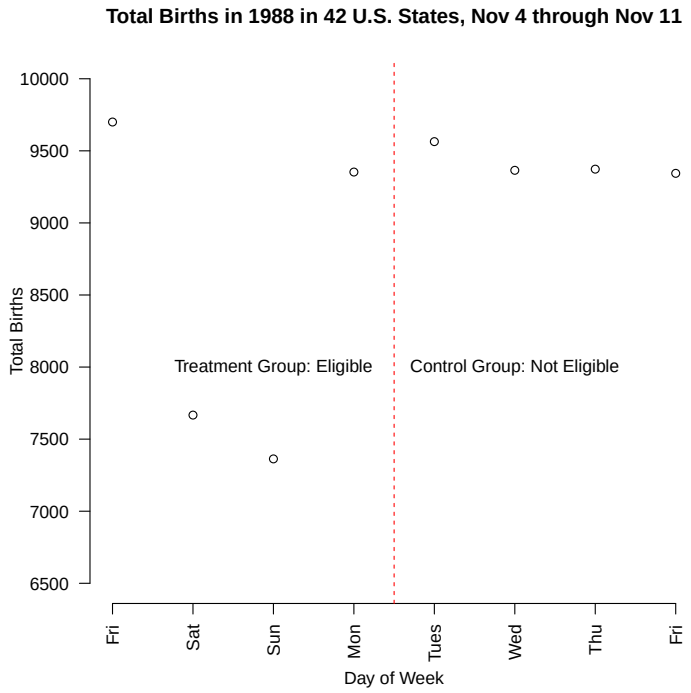


Source: Vital Statistics of the United States for 1986, 1988 and 1990, Volume I, Natality.

Figure S6: Total births by day in 42 U.S. states included in Study 1



(a) Daily birth counts in 1986



(b) Daily birth counts in 1988

Source: Vital Statistics of the United States for 1986 and 1988, Volume I, Natality.

S8 Sensitivity analysis example: Voting rights restorations

The problem of differential registration in studies based on voter files is less severe than many missing data problems because we know that all eligible voters who are not registered did not cast a vote, leaving only the differential registration factor k to be varied. This approach can be applied to other research designs in which the size of the treatment and control populations are not known but outcomes are known with certainty and an intervention could differentially affect the likelihood of a treatment case being observed compared to a control.

Meredith and Morse (N.d.), for instance, consider differences by race in the rate at which voting rights restoration applications by ex-felons in Alabama are denied due to outstanding legal financial obligations (LFOs). In this case, African American ex-felons who are eligible to apply for restoration of their voting rights are the treatment group, and eligible non-African American ex-felons are the control group. All outcomes are observed among individuals who petition to have their voting rights restored—the group that is the equivalent of registered voters in turnout studies. Moreover, the outcome of interest—voting rights—is known to be 0 among those ex-felons who do not apply to have those rights restored. However, the size of the treatment and control group populations are unknown due to limitations on data from the Alabama criminal courts system. The sensitivity analysis approach we propose can be applied in this case to assess how sensitive these results are to potential differences in application rates by race.

Another example comes from the literature on international relations. Many analysts study the likelihood of escalation between states among observed disputes (e.g., Senese 1997). However, this research design neglects how a treatment of interest might also influence the likelihood of dispute initiation among the unknown set of *potential* disputes that could be initiated. One potential approach is to estimate a two-stage selection model (Senese and Vasquez 2003) or a joint model of the likelihood of onset and escalation (Reed 2000), but scholars who prefer to avoid the strong distributional assumptions that these approaches typically require could use our sensitivity analysis approach instead. For instance, Senese (1997) considers the effect of joint democracy (the treatment of interest) on dispute escalation among the set of qualifying observed disputes between states. However, joint democracy might affect the likelihood of a dispute being observed among the universe of potential interstate disputes, producing a form of differential selection bias. Schol-

ars could therefore estimate the sensitivity of an observed difference in escalation rates by dyad regime type to differential selection among the set of potential disputes.[3]

References

Ansolabehere, Stephen and Eitan Hersh. 2010. "The Quality of Voter Registration Records: A State-by-State Analysis." Report, Harvard University. Downloaded August 22, 2015 from http://www.eitanhersh.com/uploads/7/9/7/5/7975685/reg_quality_report_8-5-10.pdf.

Holbein, John B. and D. Sunshine Hillygus. 2015. "Making Young Voters: The Impact of Preregistration on Youth Turnout." *American Journal of Political Science* Early View.

Meredith, Marc and Michael Morse. N.d. "Discretionary Disenfranchisement: The Case of Legal Financial Obligations." Unpublished manuscript.

Reed, William. 2000. "A unified statistical model of conflict onset and escalation." *American Journal of Political Science* 44(1):84–93.

Senese, Paul D. 1997. "Between dispute and war: The effect of joint democracy on interstate conflict escalation." *Journal of Politics* 59(1):1–27.

Senese, Paul D. and John A. Vasquez. 2003. "A unified explanation of territorial conflict: Testing the impact of sampling bias, 1919–1992." *International Studies Quarterly* 47(2):275–298.

[3] An alternate approach that is more common in the literature is to consider selection effects among dyad-years where the treatment and control populations can be fully enumerated.