Why Do Vote-by-Mail Elections Boost Voter Turnout?*

Yuki Atsusaka[†]
atsusaka@rice.edu

Robert M. Stein[‡]
stein@rice.edu

Andrew Menger
andrewmmenger@gmail.com

October 31, 2020

Abstract

Despite growing evidence that Vote-by-Mail (VBM) elections boost voter turnout, previous research lacks a unified explanation for *why* VBM elections do so, assessing the election reform challenging. We theorize that VBM has a dual effect of mobilizing uninformed voters by notifying them about upcoming elections (notification effect) and making voting more convenient for all voters (convenience effect). Our theory implies that the VBM effect for frequent voters is larger in midterm elections than in presidential elections, while in presidential elections, the VBM effect is larger for infrequent voters than for frequent voters. Using 27 million individual voting records from Colorado, North Carolina, and New Mexico and a difference-in-differences design with exact matching, we find supportive evidence for our expectations. Our results offer insight to assess the impact of VBM elections on voter turnout as well as fruitful topics for future research.

Keywords: *Vote-by-mail, election reform, convenience voting, difference-in-differences, semiparametric approach*

Word Count: 3945

*This study is supported by the New Venture Fund. For helpful comments, we would like to thank Vin Arceneaux, Barry Burden, Andy Hall, Shiro Kuriwaki, Randy Stevenson, Soichiro Yamauchi, and the participants of the panels at MPSA and SPSA.

[†]Ph.D. Candidate in the Department of Political Science at Rice University, MS-24 105 Herzstein Hall, 6100 Main St, Houston, TX 77005 (https://atsusaka.org).

[‡]Lena Gohlman Fox Professor, Department of Political Science, Rice University, MS-24 105 Herzstein Hall, 6100 Main St, Houston, TX 77005.

Introduction

Vote-by-Mail (VBM) elections, where every registered voter is mailed a ballot before Election Day, have been touted as an antidote to various problems in American politics (Vote at Home, 2020). Previous research on VBM elections has shown this method of voting to increase voter turnout in many places, including Oregon (Southwell and Burchett, 2000; Southwell, 2009; Gronke and Miller, 2012), Washington (Gerber, Huber and Hill, 2013), California (Bergman and Yates, 2011; Kousser and Mullin, 2007), Colorado (PantheonAnalysis, 2017), Utah (PantheonAnalysis, 2018), and multiple states (Gronke, Galanes-Rosenbaum and Miller, 2007; Richey, 2008; Larocca and Klemanski, 2011).[1] Recently, the role of VBM has received increasing attention amid the COVID-19 pandemic, and research shows that it enhances voter turnout without any evidence of partisan advantage (Barber and Holbein, 2020; Thompson et al., 2020; Bonica et al., 2020). Despite growing attention to VBM, the literature has not addressed *in what mechanisms* VBM boost turnout, making it difficult for researchers and election administrators to assess when VBM actually increases turnout among different types of voters as well as how it does so.

To resolve this issue and advance the research of VBM, we first propose a unified explanation of why VBM increases voter turnout. By focusing on registered voters, we theorize that VBM has a *dual effect* of mobilizing uninformed voters by notifying them about upcoming elections (notification effect) *and* making voting more convenient for all voters (convenience effect).[2] Our theory implies that frequent voters have both convenience and notification effects in midterm elections, but only only receive a convenience effect in presidential elections. It also suggests that infrequent voters experience a convenience effect in both types of elections, while they may also be susceptible to a notification effect in midterm and presidential elections. Based on these implications, we hypothesize that the VBM effect for frequent voters is larger in midterm elections than in presidential elections, while in presidential elections, the VBM effect is larger for infrequent voters than for frequent voters.

We test our hypotheses by using more than 27 million individual voting records from Colorado, which adopted VBM statewide in 2013, as well as North Carolina and New Mexico, which did not use universal VBM for midterm and presidential elections between 2012 and 2016. For the 2014 midterm and 2016 presidential elections, we estimate the conditional average treatment effect on the treated of VBM elections

[1] Online Appendix A offers a brief history and previous research of VBM.

[2] More precisely, we consider the causal effect of *adopting VBM option* on voter turnout in a jurisdiction, or equivalently, the causal effect of the *availability* of universal mail balloting on voter turnout.

on turnout with various effect modifiers (i.e., for different subpopulations) with difference-in-differences models combined with exact matching. We find that VBM elections significantly boosted turnout among all types of voters in both midterm and presidential elections, but the magnitude of the effect varies by the types of elections and voter frequency in a manner consistent with our theory. Our results offer insight to assess the impact of VBM elections on voter turnout as well as fruitful topics for future research.

Dual Effect of Notification and Convenience

We identify two ways by which VBM elections can enhance political participation among registered voters: (1) giving voters a *notification* about an upcoming election and (2) increasing the *convenience* with which they cast a ballot.[3] We call these the notification and convenience effects, respectively. We then formulate the total effect of VBM adoption as an additive effect of the notification and convenience as follows:

$$\text{VBM Effect} = \text{Notification Effect} + \text{Convenience Effect} \tag{1}$$

The notification effect builds upon the *noticeable reminder theory* introduced by Dale and Strauss (2009). This theory argues that impersonalized messaging can be efficacious for turning out registered voters, who have already signaled their willingness to participate in the political process. According to this explanation, what these voters require to vote is "a reminder to make time in their busy schedules to go to the polls," where such messaging must be noticeable, unavoidable, and proximate to Election Day (Dale and Strauss, 2009, 787).

One method by which VBM may increase voter turnout is by providing registered voters with such a "reminder" in the form of a mailed ballot that is noticeable, unavoidable, and proximate to when they are likely to vote. One reading of Dale and Strauss's theory implies that this form of notification is more salient among those who are less informed about elections (e.g., infrequent voters) and in low-informational contests such as local and midterm elections.[4] However, an alternative reading of Dale and Strauss's theory is that a noticeable reminder may only be effective for infrequent voters in presidential elections since these

[3] We focus on voters who are "already registered to vote" at the time of VBM adoption for several reasons. First and foremost, VBM intends to offer an alternative mode of voting primarily for registered voters. Second, even if VBM intends to encourage unregistered voters to get registered, it is unclear why and how such a "registration effect" would occur. Empirically, Barber and Holbein (2020, 3) also show that the adoption of VBM has no association with registration rates. Finally, even if we do not have the theoretical problem, quantifying the notification, convenience, and registration effects will be difficult on empirical grounds.

[4] We assume that frequent voters are more informed than infrequent voters.

types of registered voters often do not have the "willingness to participate" in less salient midterm elections. If these individuals do not already have a strong inclination to vote in these elections, a noticeable reminder in the form of a mailed ballot may not be sufficient to entice them to vote. Due to the competing implications of this theory, we are agnostic about in which types of elections (presidential vs. midterm) infrequent voters may benefit from a notification effect.

On the other hand, by definition, frequent voters almost always participate in presidential elections and, therefore, will not benefit from a noticeable reminder. These individuals should experience a higher probability of voting from the reminder of a mailed ballot only in less-advertised midterm elections in which they already intend to participate but may need a "nudge" or may not be aware of the details on when, where, and how to vote.[5] Therefore, we expect that any notification effect for frequent voters should be visible only in midterm elections.

Our claim on the convenience effect draws from the literature regarding the costs of voting and the calculus of voting more generally. We argue that the convenience associated with VBM mitigates the high costs of locating and getting to a polling place on or before Election Day (Brady and McNulty, 2011; Fitzgerald, 2005; Haspel and Knotts, 2005). Indeed, VBM elections allow voters to complete their ballot over several days (or even weeks), to fill it out at home or the location of their choosing, and to return their ballot via mail. We claim that the convenience effect is present for all voters since voters of all age groups, socioeconomic levels, and from across both political parties experience high costs of voting in person (Menger, 2018). While some types of voters, such as those without transportation or the disabled, may benefit more from the cost reductions of VBM, we do not expect that these groups are systematically more present among subpopulations of frequent vs. infrequent voters or by demographic characteristics.

The key observable implications from our theory (with our effects of interest) are summarized in Figure 1. It illustrates that in midterm elections (as low-informational contexts), both notification and convenience effects will be present for more informed frequent voters, while the notification may be absent for infrequent voters. It also shows that in presidential elections (as high-informational contexts), VBM will only have a convenience effect among informed and frequent voters, while uninformed and infrequent voters may benefit from the dual effect of notification and convenience. Our theory implies that we should observe larger turnout effects from VBM in midterm elections than in presidential elections for frequent voters.

[5] Mailed ballots provide detailed instructions on how to fill out the ballot and where and when to return it. Additionally, they often provide information on in-person voting options for those who choose not to use this voting method.

Thus, we consider a one-sided test with $H_{0,\text{Freq}} : \tau_{11} \leq \tau_{21}$ versus $H_{1,\text{Freq}} : \tau_{11} > \tau_{21}$, where H_0 and H_1 are our composite null and alternative hypotheses, respectively. It also implies that in presidential elections, we should see a larger turnout effect for infrequent voters than for frequent voters as infrequent voters may also experience a notification effect. For this expectation, we consider a one-sided test with $H_{0,\text{Pres}} : \tau_{21} \geq \tau_{22}$ versus $H_{1,\text{Pres}} : \tau_{21} < \tau_{22}$. Finally, we also consider an auxiliary one-sided test with $H_{0,\text{Mid}} : \tau_{11} \leq \tau_{12}$ versus $H_{1,\text{Mid}} : \tau_{11} > \tau_{12}$, although we are less certain about this expectation.[6] Below, we test these hypotheses by comparing the magnitude of the causal effects of VBM on turnout among frequent and infrequent voters. Meanwhile, we expect that the effect of VBM does not vary substantially by other subpopulations (e.g., by race, party, and gender) by assuming that each group has a relatively similar level of informed and frequent voters.

	Frequent voters	**Infrequent voters**
Midterm elections	$\tau_{11} = \textit{notification} + \textit{convenience}$	$\tau_{12} = \textit{(notification+) convenience}$
Presidential elections	$\tau_{21} = \textit{convenience}$	$\tau_{22} = \textit{(notification+) convenience}$

Figure 1: **Implications of dual effect of VBM.**
Note: This figure shows expected effects by subpopulations and by the types of elections as described above. Our main expectations are $\tau_{11} > \tau_{21}$ and $\tau_{21} < \tau_{22}$.

Data and Identification Strategy

We define our population of interest as a set of voters who had been registered in Colorado between 2010 and 2016 (before and after the adoption of VBM). We define our quantities of interest as a set of *conditional average treatment effects on the treated* (CATTs):

$$\tau_{CATT} = \mathbb{E}_{\mathbf{X}}\Big[\mathbb{E}[Y_i^{d=1} - Y_i^{d=0} | D_i = 1, Z_i = z, \mathbf{X_i}]\Big], \tag{2}$$

where $Y_i^{d=1}$ and $Y_i^{d=0}$ are potential outcomes (turnouts) for registered voter i, $D_i = 1$ the treatment status (VBM is available), $Z_i = z$ an *effect modifier* denoting voter i belongs to a subpopulation z, and $\mathbf{X_i}$ represents a vector of covariates.[7] For effect modifiers, we consider whether a voter is a frequent voter (our main variable), belongs to a particular racial group (White, Black, Hispanic, Asian, or other race),

[6] We assume that the nature and magnitude of notification and convenience effects do not vary across the types of elections and voters.

[7] This means that we study VBM effects among voters whose turnout the election reform intends to increase, and we examine how these effects vary by subpopulations (see Online Appendix C.1 for details)

is female, is a Democrat, and belongs to a particular age group.[8] We also assume that weak ignorability ($Y^{d=0} \perp\!\!\!\perp D_i | \mathbf{X_i}$) holds.

We estimate the CATTs by leveraging voter history files from the 2012 to 2016 presidential and midterm elections in Colorado, North Carolina, and New Mexico. Colorado adopted VBM elections in 2013, whereas North Carolina and New Mexico did not adopt VBM during this period and are used as counterfactual states (see Online Appendix B). In other words, we test our hypotheses with individual-level turnout data in the states before (i.e., 2012) and after the intervention (VBM adoption) occurs (i.e., 2014 and 2016, respectively).[9]

For our identification strategy, we adopt a parametric linear difference-in-differences (DID) model (controlling for voter's frequency of voting, age, race, gender, and partisanship) with two years and two states. The key idea is to simulate voter turnout in Colorado (both in 2014 and 2016) had the VBM policy not been adopted using voters from one control state (either North Carolina or New Mexico) and use such counterfactual turnout to estimate the causal effect of the policy adoption. The validity of this approach depends (largely) on the *parallel trends assumption*, which states that the unobserved confounders that create a systematic difference in turnouts in the two states should be constant over time.[10] We provide suggestive evidence that the assumption seems to hold in our setting (Online Appendix C.4), although this assumption is inherently non-refutable (i.e., cannot be empirically verified with *any* data).[11]

To mitigate the concern over the potential violation of the assumption, we employ a semi-parametric approach. Specifically, we carry out exact matching before the application of the DID model so that we can restrict the set of control units to those who have similar sociopolitical attributes to treated units.[12] We then implement a DID model by a saturated weighted least squares regression, where weights (for control units) are obtained from exact matching (see Online Appendix C.4).[13] We further check the validity of our inferences by adopting the notion of *pattern specificity* (Rosenbaum, 2005). The idea is to compare

[8] We code a voter as a frequent voter if she has voted in the 2010 midterm election by assuming that those who voted in 2010 are more likely than those who did not to have voted in previous elections (and thus being a frequent voter). For age groups, we classify voters whether they are less than 40, between 41 and 65, or over 65. We consider "Hispanic" as a racial group despite that it is an ethnicity.

[9] Colorado has simultaneously adopted VBM and the Same Day Registration (SDR) in 2013. However, because we only study voters who were registered between 2010 and 2016, the adoption of SDR does not affect our causal identification.

[10] For this reason, our identification is valid *even if Colorado and North Carolina voters have very different political cultures* that cannot be adjusted by available covariates as long as such difference remains the same across the two elections of comparison.

[11] We also make the linearity assumption in the conventional DID model.

[12] Due to the large size of our data, we carried out exact matching on a simple random sample of the original data for each effect modifier.

[13] This semi-parametric approach has been considered in ONeill et al. (2016) and shown to reduce biases that could appear when the parallel trends assumption is violated in Ryan et al. (2019). For a related semi-parametric method, see also Abadie (2005).

our treated units to different control units from North Carolina and New Mexico to verify if a common pattern emerges regardless of the choice of control units.[14] Although it is practically impossible to adjust for unobserved time-varying confounders, we can be more confident in findings that are consistent across different control units, which we find for our key variable.

Empirical Findings

Figure 2 presents the estimated effects of VBM on turnout for frequent and infrequent voters in 2014 and 2016 when using North Carolina voters (left panel) and New Mexico voters (right panel) as control units.[15] It shows that the estimated turnout effect from VBM for frequent voters is substantially and statistically significantly larger in 2014 than in 2016 (i.e., $\tau_{11} > \tau_{21}$) across both control states. Moreover, we find that the VBM effect is significantly larger among infrequent voters than among frequent voters in the 2016 presidential election (i.e., $\tau_{21} < \tau_{22}$) regardless of our choice of control states. Given these results, we reject $H_{0,\text{Freq}}$ and $H_{0,\text{Pres}}$. We also find evidence that frequent voters experience a larger turnout increase from VBM than infrequent voters in the midterm election (i.e., $\tau_{11} > \tau_{12}$) when using New Mexico voters as control units, but this does not hold across both control states. Being conservative on the hypothesis testing, we conclude that we fail to reject $H_{0,\text{Mid}}$.[16]

Next, we examine how the turnout effect of VBM varies by other subpopulations, including groupings by partisan affiliation and demographic characteristics. Figure 3 shows the estimated effects in the 2016 presidential election by voting frequency, age, race and ethnicity, gender, and party affiliation. The results show that the effect size does not vary much by other demographic characteristics in a manner consistent across the two states used as comparisons, except for age groups. Notably, we did not find that VBM increases turnout only or particularly among Democratic or non-Democratic voters consistently across the two control states. We find some suggestive evidence that the effect is smaller for middle-aged voters and larger for voters aged 65 and over, although the effect magnitude and significance are not consistent across

[14] An alternative strategy is to combine North Carolina and New Mexico data and only present the result based on it. We present our results based on this strategy in Online Appendix D.2. However, this approach may mask important limitations of our analysis by *assuming away* that potential issues in the identification. For this reason, we take a more conservative approach for our identification.

[15] To construct a 95% confidence interval, we use the (estimated) individual-level clustered standard error.

[16] We do not find consistent evidence that infrequent voters experience a larger turnout increase from VBM in the midterm than in presidential elections (i.e., $\tau_{12} > \tau_{22}$) (Dale and Strauss, 2009). This finding suggests that infrequent voters may benefit from the notification effect either equally in both types of elections, or more in high-salience presidential elections in which they already may intend to vote but need a reminder and/ or the relevant information that comes from a mailed ballot.

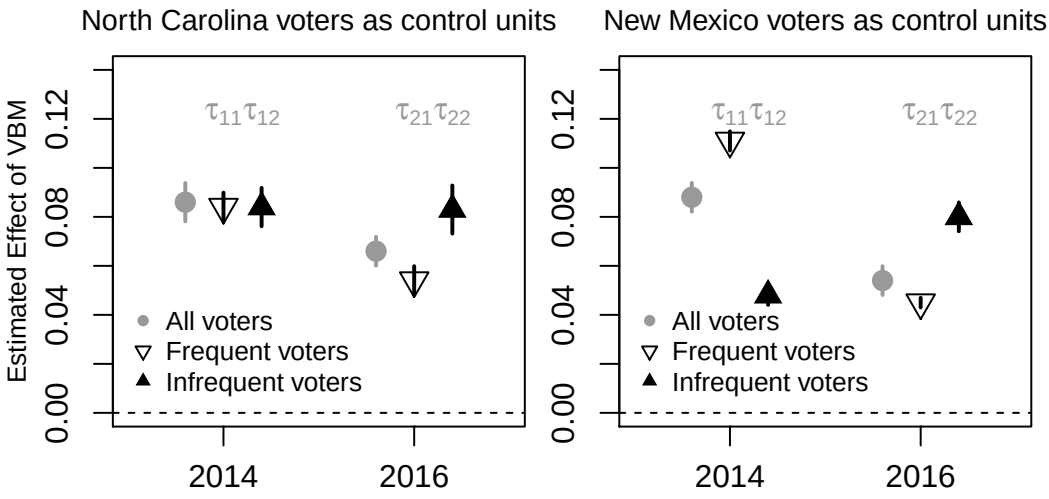


Figure 2: **Estimated Average Effects in the 2014 Midterm and 2016 Presidential Elections.**
Note: This figure visualizes the estimated VBM turnout effect for various subpopulations by voting frequency when using North Carolina (left panels) and New Mexico (right panels) as control states. We find $\tau_{11} > \tau_{21}$ and $\tau_{21} < \tau_{22}$.

the control states.

We performed multiple robustness checks to confirm that our conclusions do not depend on specific decisions we made during our data collection, coding, modeling, and statistical inferences. Online Appendix D offers additional findings, robustness checks, placebo tests, and circumstantial evidence for the parallel trends assumption, confirming that our substantive conclusions hold.

Implications for Future Research

We have identified two mechanisms for how VBM elections boost voter turnout: notification and convenience. Consistent with our theory, we find that, for frequent voters, the effect of VBM on turnout is substantially larger in midterm than in presidential elections, whereas, in presidential elections, the effect is larger for infrequent voters than for frequent voters. We do not find consistent evidence on whether infrequent voters experience a larger notification effect in midterm elections vs. presidential elections. We also do not find consistent evidence that the effect of VBM varies by partisan or demographic characteristics, except for a slightly smaller effect for middle-aged voters than for other groups.

Not only our theoretical framework assists future evaluations of the impact of VBM on voter turnout

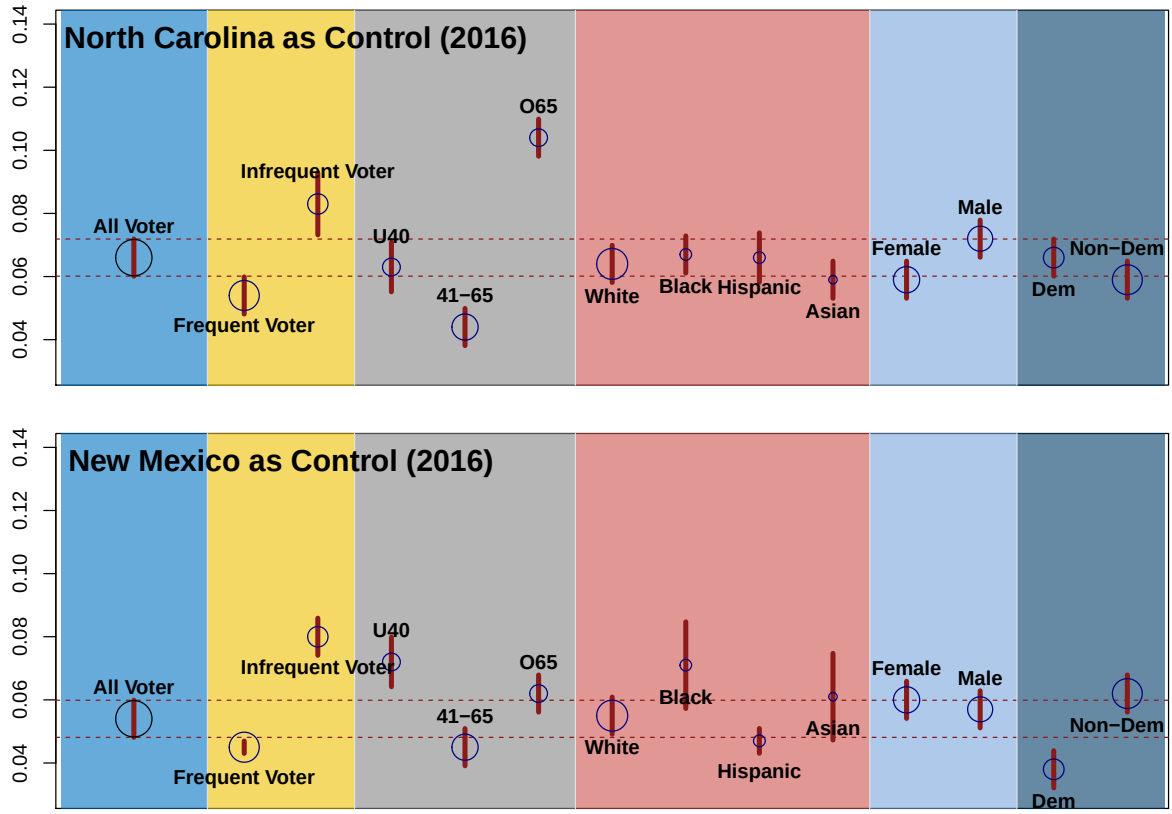


Figure 3: **Estimated Average Effects of VBM on Turnout with Different Effect Modifiers.**
Note: This figure visualizes the estimated effect (in 2016) with a 95% confidence interval among various subpopulations of interest when using North Carolina (upper panel) and New Mexico (lower panel) as control states. For each group, the size of its open circle is proportional to the square-root of the proportion of the group in Colorado.

but also helps political scientists understand previous studies that find varying (or not varying) magnitudes of VBM effects by frequency of voting and age as well as by race, gender, and partisanship (e.g., Barber and Holbein, 2020; Thompson et al., 2020; Bonica et al., 2020). Testing the implications of the dual effect more directly, however, presents several challenges. Observational studies of VBM elections are limited in their ability to deal with unobserved (time-varying) confounders and violation of the stable unit treatment value assumption (e.g., Keele and Titiunik, 2018) among other inferential problems. One fruitful area in future research is to use survey experiments conducted with registered voters to provide a partial means of addressing challenges to estimate the notification and convenience effects more directly. We also leave future research to expand the population of interest into eligible voters (as opposed to already registered voters) and incorporate a potential registration effect both theoretically and empirically.

References

Abadie, Alberto. 2005. "Semiparametric difference-in-differences estimators." *The Review of Economic Studies* 72(1):1–19.

Barber, Michael and John B Holbein. 2020. "The participatory and partisan impacts of mandatory vote-by-mail." *Science Advances* 6(35):eabc7685.

Bergman, Elizabeth and Philip A Yates. 2011. "Changing election methods: How does mandated vote-by-mail affect individual registrants?" *Election Law Journal* 10(2):115–127.

Bonica, Adam, Jacob M. Grumbach, Charlotte Hill and Hakeem Jefferson. 2020. "All-Mail Voting in Colorado Increases Turnout and Reduces Turnout Inequality." *Working Paper* .

Brady, Henry E and John E McNulty. 2011. "Turning out to vote: The costs of finding and getting to the polling place." *American Political Science Review* 105(1):115–134.

Dale, Allison and Aaron Strauss. 2009. "Don't forget to vote: Text message reminders as a mobilization tool." *American Journal of Political Science* 53(4):787–804.

Fitzgerald, Mary. 2005. "Greater convenience but not greater turnout: The impact of alternative voting methods on electoral participation in the United States." *American Politics Research* 33(6):842–867.

Gerber, Alan S, Gregory A Huber and Seth J Hill. 2013. "Identifying the effect of all-mail elections on turnout: Staggered reform in the evergreen state." *Political Science Research and Methods* 1(1):91–116.

Gronke, Paul, Eva Galanes-Rosenbaum and Peter A Miller. 2007. "Early voting and turnout." *PS: Political Science & Politics* 40(4):639–645.

Gronke, Paul and Peter Miller. 2012. "Voting by mail and turnout in Oregon: Revisiting Southwell and Burchett." *American Politics Research* 40(6):976–997.

Haspel, Moshe and H Gibbs Knotts. 2005. "Location, location, location: Precinct placement and the costs of voting." *Journal of Politics* 67(2):560–573.

Keele, Luke and Rocío Titiunik. 2018. "Geographic Natural Experiments with Interference: The Effect of All-Mail Voting on Turnout in Colorado." *CESifo Economic Studies* 64(2):127–149.

Kousser, Thad and Megan Mullin. 2007. "Does voting by mail increase participation? Using matching to analyze a natural experiment." *Political Analysis* 15(4):428–445.

Larocca, Roger and John S Klemanski. 2011. "US state election reform and turnout in presidential elections." *State Politics & Policy Quarterly* 11(1):76–101.

Menger, Andrew. 2018. The Costs of Voting and their Impacts. PhD thesis Rice University.

ONeill, Stephen, Noémi Kreif, Richard Grieve, Matthew Sutton and Jasjeet S Sekhon. 2016. "Estimating causal effects: considering three alternatives to difference-in-differences estimation." *Health Services and Outcomes Research Methodology* 16(1-2):1–21.

PantheonAnalysis. 2017. "Utah 2016: Evience for the positive turnout effects of Vote at Home (also known as vote by mail) in participation counties.".

PantheonAnalysis. 2018. "Colorado 2014. Comparisions of Predicted and Actual Turnout.".

Richey, Sean. 2008. "Voting by mail: Turnout and institutional reform in Oregon." *Social Science Quarterly* 89(4):902–915.

Rosenbaum, Paul R. 2005. "Observational study." *Encyclopedia of statistics in behavioral science* .

Ryan, Andrew M, Evangelos Kontopantelis, Ariel Linden and James F Burgess Jr. 2019. "Now trending: Coping with non-parallel trends in difference-in-differences analysis." *Statistical methods in medical research* 28(12):3697–3711.

Southwell, Priscilla L. 2009. "Analysis of the turnout effects of vote by mail elections, 1980–2007." *The Social Science Journal* 46(1):211–217.

Southwell, Priscilla L and Justin I Burchett. 2000. "The effect of all-mail elections on voter turnout." *American Politics Quarterly* 28(1):72–79.

Thompson, Daniel M, Jennifer A Wu, Jesse Yoder and Andrew B Hall. 2020. "Universal vote-by-mail has no impact on partisan turnout or vote share." *Proceedings of the National Academy of Sciences* .

Vote at Home. 2020. *Proven Track Record*.
URL: *https://www.voteathome.org/project/proven-track-record/*

Online Appendix

For "Why Do Vote-by-Mail Elections Boost Voter Turnout?"

Table of Contents

A VBM Elections and Previous Research

A.1 Brief History of VBM Elections

First adopted by Oregon in 1996, Washington and Colorado have also implemented VBM elections statewide for all elections in 2005 and 2013, respectively. In 2014 Utah initiated VBM elections on an optional county-wide basis, and most recently California has introduced optional countywide VBM elections beginning in 2020. Hawaii adopted VBM voting for the 2020 Presidential election. Since 2000 California counties have had the discretion to conduct VBM elections in precincts where there are 250 or fewer registered voters, presumably to save money on the costs of poll workers, voting stations, and polling locations. Registered voters in VBM-only precincts are automatically mailed a ballot whether they request one or not with a postage-paid return envelope. Unlike in most other forms of VBM elections, these voters do not have the option of voting in person. More recently, several states have adopted no-excuse absentee voting or universal VBM for the November 2020 presidential election, but these adoptions are not permanent and only in response to the COVID-19 pandemic.

A.2 Previous Findings

To date, a body of research has examined the effect of adopting VBM elections on voter turnout. Table A.1 lists published studies on this topic with the place of interest, the type of elections, data source, and main findings on estimated effects (positive, negative, or null). Overall, previous research has found positive effects of VBM elections on turnout, although the estimated magnitudes vary across studies.

While most research shows a significant and positive effect of VBM on voter turnout, these findings are not fully consistent and some researchers report either no effect on turnout or even a negative effect (Gronke and Miller, 2012; Southwell, 2009; Bergman and Yates, 2011; Kousser and Mullin, 2007; Keele and Titiunik, 2018) The findings of negative effects are restricted to California counties discretionary use of VBM elections, but some studies of VBM in Oregon have reported the absence of a positive effect from the policy adoption.

Importantly, more recent studies (including unpublished works) have studied the VBM turnout effect based on stronger causal identification strategies than earlier studies employed. For example, using a staggered implementation of VBM as a quasi-natural experiment, Gerber, Huber and Hill (2013) find that VBM increases turnout by two to four percentage points. Based on similar experimental leverage, Thompson et al. (2020) report that VBM elections increase turnout by two to three percentage points, with no evidence of partisan advantage. More recently, Bonica et al. (2020) report turnout increases over nine percentage points in Colorado and by larger margins for subpopulations (i.e., younger voters and voters of color). Barber and Holbein (2020) also report significant positive turnout effects from VBM elections.

Authors	Place	Election Type	Data	Effect
Southwell and Burchett (2000)	Oregon	Federal	TS, state level	+
Gronke et al. (2007)	All states	Presidential	Panel, state level	+
Richey (2008)	All states	Federal	Panel, state level	+
Gerber, Huber and Hill (2013)	WA (counties)	Federal	Panel, individual	+
Larocca and Klemanski (2011)	All states	Federal	CS, survey	+
Gronke and Miller (2012)	Oregon	Fed & Special	TS, state level	null & +
Southwell (2009)	Oregon	All	TS, state level	null & +
Bergman and Yates (2011)	CA (precincts)	Local	CS, individual	−
Kousser and Mullin (2007)	CA (precincts)	Federal	CS, individual	−
Sled (2008)	8 states	All	Panel, local level	+
PantheonAnalysis (2017)	Colorado	Federal	CS, individual	+
PantheonAnalysis (2018)	Utah	Federal	CS, individual	+
Thompson et al. (2020)	3 states	federal	Panel, county	+
Bonica et al. (2020)	Colorado	federal	Panel, individual	+
Barber and Holbein (2020)	6 states	federal	CS, individual	+

Table A.1: **Previous Research on the Effects of VBM on Voter Turnout.**
Note: WA = Washington, CA = California, TS= Time Series, CS = Cross Sectional, Gronke et al (2007) = Gronke, Galanes-Rosenbaum and Miller (2007).

B Additional Information on Data

B.1 Data Collection

We collected historical voter files from Colorado, North Carolina, and New Mexico for the 2010, 2012, 2014, and 2016 elections. These files are made available online by the Colorado Secretary of State, the North Carolina State Board of Elections, and New Mexico Secretary of State, respectively. Among the raw data, we only keep those voters who were registered to vote on Election Day in the period from 2010 to 2016. Reducing the data in this way allows us to observe the behavior of the same voters across the pairs of elections (2012 vs. 2014 and 2012 vs. 2016) being studied, which allow us to clearly define our population and counterfactual of interest. We discuss this point further in Appendix C.1. The dataset includes a variety of information about each voter, including their age at the time of the election, their gender, and their party registration as well a prior registration date and voting history. We first code for whether each person voted in the election. For all three states, we imputed voters' race and ethnicity as described in the next subsection, even though, the North Carolina voter files includes the voters self-identified race and ethnicity as required by federal law. We do so to make sure that we have the same coding scheme across states.

B.2 Coding Individual Race

To evaluate the conditional average treatment effects on the treated by racial groups, we also need to code for individual race and ethnicity (hereafter "race"). However, neither voter history files nor registration data include information about individuals' race in the state of Colorado. Given this limitation, we predict individual's race by drawing on a Bayesian approach proposed by Imai and Khanna (2016).

Specifically, for each registered voter, we predicted her race by choosing the race that has the highest posterior probability that she belongs to the group conditional upon her background characteristics including surname, residential address, age, gender, and registered party. For example, when our posterior probabilities for a voter look like:

$$\Pr(\text{Black} = 0.75 \mid \text{Smith, BlockGroup1001, age 30, female, and Democrat})$$
$$\Pr(\text{White} = 0.15 \mid \text{Smith, BlockGroup1001, age 30, female, and Democrat})$$
$$\Pr(\text{Asian} = 0.05 \mid \text{Smith, BlockGroup1001, age 30, female, and Democrat})$$
$$\Pr(\text{Hispanic} = 0.05 \mid \text{Smith, BlockGroup1001, age 30, female, and Democrat})$$

we code her as a Black voter.

To apply this approach, we first geocoded the set of registered voters by using `censusgeocode` module in Python and obtained which Block Group and Census Track each voter belongs to. After coding this information, we computed the above posterior probabilities by using `wru` in R develoepd by (Khanna and Imai, 2019). Finally, for each voter, we coded her race by choosing the group that gives the highest posterior probability.

B.3 Descriptive Statistics

After our careful data construction, we obtain four panel data sets (CO-NC-12-14, CO-NC-12-16, CO-NM-12-14, CO-NM-12-16). Table B.1 shows the turnout rate in each state in each year.

	Colorado	North Carolina	New Mexico
2010	0.685	0.479	0.583
2012	0.828	0.694	0.697
2014	0.704	0.461	0.480
2016	0.819	0.624	0.641

Table B.1: **Voter Turnout by State and Year**

C Additional Discussion on Identification

C.1 Populations and Quantities of Interest

Our primary population of interest is *a set of voters who were registered in Colorado between the 2010 and 2016 elections*. Besides, we consider a set of politically salient subpopulations in the primary population as our secondary populations of interest. This means that we do not consider voters who registered in Colorado *only* at the time of the 2012, 2014, *or* 2016 presidential elections.

Defining the populations of interest this way is critical when one wants to accurately identify and estimate the information and convenience effects of VBM elections. First, we can isolate the information and convenience effects from the registration effect (the adoption of VBM encourages more people to register and turn out). Second, we can be more certain that the composition of the primary population and subpopulations will (almost) stay the same.

Given our populations of interest, we now define our quantities of interest. Let Y_i be a binary random variable for the outcome, denoting whether a registered voter i turned out to vote in an election. Let D_i denote a binary random variable for the treatment status, representing if the same voter had an option to use a mail ballot. Based on the potential outcomes framework, let us define $Y_i^{d=1}$ as a potential outcome (turnout) for the voter had she been assigned to a jurisdiction with a mail ballot, and $Y_i^{d=0}$ as a potential outcome (turnout) for the same voter had she been assigned to a jurisdiction without the VBM adoption. Here, we assume that the consistency assumption holds as $Y_i = D_i Y_i^{d=1} + (1 - D_i) Y_i^{d=0}$. Assuming an additive effect measure, we define our first quantity of interest (for the primary population or "all voters") as the average treatment effect on the treated (ATT):

$$\tau_{ATT} = \mathbb{E}[Y_i^{d=1} - Y_i^{d=0} | D_i = 1]. \tag{1}$$

Substantively, this means that we study what turnout rate would have looked like in the 2016 presidential election in Colorado if VBM elections had not been adopted in 2013 (in contrast to the reality in which voters were exposed to VBM elections).

Now, we consider an *effect modifier* Z_i as a categorical variable taking a discrete value z, which denotes that voter i belongs to group z. Specifically, we are interested in several group structures including the frequency of voting, race and ethnicity, gender, partisanship, and age. These groups are called effect modifiers because causal effects of interest (i.e., ATT in our case) are expected to be modified by (or conditioned upon) such groups. Based on this, we define our next quantities of interest as the conditional average treatment effect on the treated (CATT):

$$\tau_{CATT} = \mathbb{E}[Y_i^{d=1} - Y_i^{d=0} | D_i = 1, Z_i = z]. \tag{2}$$

In practice, since we use observed outcomes of voters who had the option of use mail ballots and those who had not (and not a variant of randomized experiments), we are concerned with the CATT conditional on a set of covariates.

$$\tau_{CATT}^{X} = \mathbb{E}_X \big[\mathbb{E}[Y_i^{d=1} - Y_i^{d=0} | D_i = 1, Z_i = z, \mathbf{X} = \mathbf{x}] \big], \tag{3}$$

where $\mathbf{X}$ is a vector of p covariates and $\mathbf{x}$ denotes a vector of specific covariate values, defined over the p-dimensional covariate space $\mathcal{X}$. We assume that the conditional ignorability holds such that within multi-dimensional strata of background characteristics the potential outcomes of voters in the treatment condition and control condition do not depend on the treatment assignment.

C.2 Using North Carolina and New Mexico as Counterfactuals

The key to our identification strategy is to assume that North Carolina and New Mexico voters are good counterfactuals for Colorado voters (Keele and Minozzi, 2013). We believe that North Carolina is an appropriate control state for several conditions. Most importantly, neither North Carolina and New Mexico have used VBM elections, but both offer all the other methods of voting as Colorado does. During the period under study, North Carolina and New Mexico voters were able to vote absentee by mail with no excuse, in-person before Election Day, and in-person on Election Day. These options for casting a ballot were also available to Colorado voters before and after the adoption of VBM elections in 2013.

Moreover, all three states have a competitive partisan environment. In 2016, the two-party vote share differential was 5% in Colorado[1], 4% in North Carolina[2], and 9% in New Mexico.[3] Both states have a racially and ethnically diverse population. 31% of North Carolinas population is non-white white[4], 36% of Colorados population is non-white[5], and 19% of New Mexico's population is non-white, with 48% Hispanic.[6] The majority-minority population in Colorado is Hispanic (21%) and an identical proportion of North Carolinas population is African-American (21.5%). The age distribution of both states is comparable with 16% and 14% of Colorados and North Carolinas population over the age 65 respectively, 18% of New Mexicos population is over 65 years of age. A fifth of all three states populations (22%) are under the age of 18. These data indicate that North Carolina and New Mexico are good comparison states with which to evaluate the adoption of VBM elections in Colorado. These comparisons are displayed in Table C.1.[7]

	Colorado	**North Carolina**	**New Mexico**
Partisan competition (2016)	5% Dem	3.6% Rep	9.2% Dem
U.S. House seats	7 (4-D, 3-R)	13 (9-R,3-D, 1-I)	3 (1-R,2-D)
Non-White	31%	36%	19%
Under 18	22%	23%	23%
Over 65	16%	14%	18%
In-person before Election Day	✓	✓	✓
Early voting	✓	✓	✓
Same day registration	✓	✓	✓
No excuse mail voting	✓	✓	✓

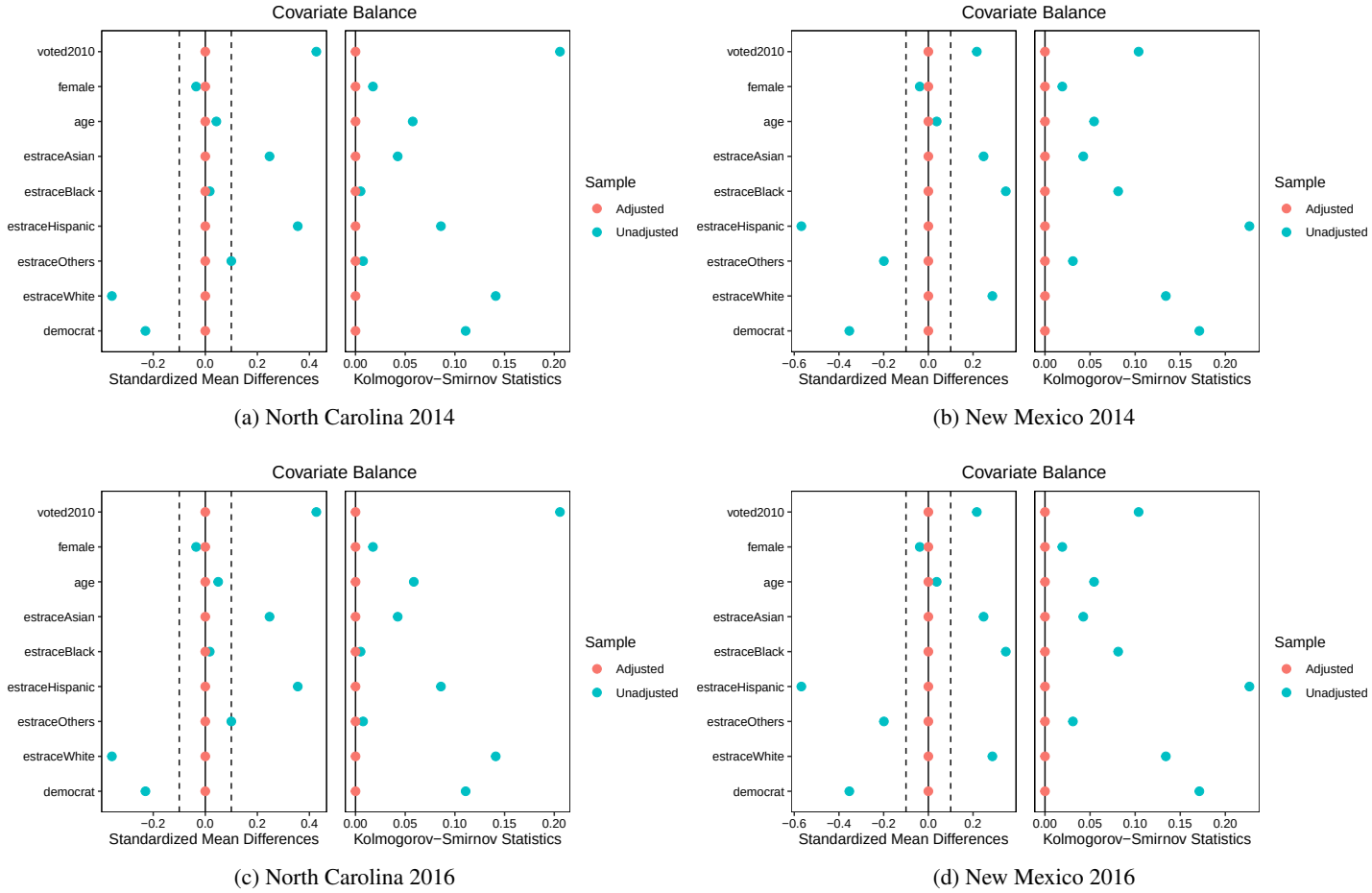
Table C.1: **Comparison between Treated and Control States.**
Note: This table shows the comparison between Colorado and North Carolina on several state-level variables of interest. The same day registration is available in Colorado since 2013 and in North Carolina since 1994.

[1] https://en.wikipedia.org/wiki/2016_United_States_presidential_election_in_Colorado
[2] https://en.wikipedia.org/wiki/2016_United_States_presidential_election_in_North_Carolina
[3] https://en.wikipedia.org/wiki/2016_United_States_presidential_election_in_New_Mexico
[4] https://www.census.gov/quickfacts/CO
[5] https://www.census.gov/quickfacts/NC
[6] https://www.census.gov/quickfacts/NM
[7] See the above urls.

C.3 Preprocessing Data

For credible observational studies, it is critical to reduce the imbalance in the covariate distributions between treated and control units in data so that our estimation and inference will be less model dependent (Ho et al., 2007). We thus preprocess our data by exact matching and then apply a difference-in-differences design to the matched data. To apply this semi-parametric approach (e.g., Abadie, 2005), we ideally want to perform exact matching on all available covariates including turnout in 2010, age, gender, race and ethnicity, and party affiliation, except for the covariate we used for an effect modifier using all units. Due to the large size of observations ($\approx$ 1.7 million), however, we perform a simple random sampling on the original data (sampling 3% of the entire data) and apply matching to the sampled data.

As a quick visual inspection, Figure C.1 compares the standardized bias and Kolmogorov-Smirnov statistics for each covariate before and after matching for each population of interest. The results demonstrate that after preprocessing our data achieves a higher balance on covariates than raw data.



(a) North Carolina 2014

(b) New Mexico 2014

(c) North Carolina 2016

(d) New Mexico 2016

Figure C.1: **Covariate Balance After Preprocessing (Selected Results).**
Note: This figure visualizes the covariate balance before and after proprocessing data (for all voters as selected results). We preprocess our data with exact matching (one-to-one matching without replacement). It shows that preprocessing increased the covariate balance, ensuring that treated units and control units look similar on observed features.

C.4 Discussion on Assumptions for Identification

To draw valid inferences under the difference-in-differences design, we must assume that the **turnout trends in the absence of the intervention in both states are the same** (i.e., the *parallel trends assumption*) (Wing, Simon and Bello-Gomez, 2018; Bertrand, Duflo and Mullainathan, 2004; Angrist and Pischke, 2008, 227-233). In the two-states and two-years design including ours, there is no way to empirically verify the parallel trends assumption (such as replacing the true intervention with a false intervention). In Online Appendix D.4, nevertheless, we offer circumstantial evidence that the turnout trends could be the same in the two states in the absence of the intervention using aggregate election data.

Moreover, we assume that there is only one type of treatment effect and each unit's outcome is only a function of her treatment status and not others (i.e., the *stable unit treatment value assumption* (SUTVA)). As discussed below, our research design might violate the SUTVA by allowing the interference *among treated units*. Indeed, by allowing voters to fill in their ballots *at their convenient locations* and *with anybody*, VBM might induce a form of interference between voters especially in the intervened location (e.g., state, county, precinct). For example, consider two voters who are sharing the same residence (e.g., married couples, partners, roommates, etc). If one of them received a mail ballot (or mail ballots) and talked to the other person about the upcoming election (or even about the mail ballot), it creates interference between the two where one's outcome is now a function of her treatment status and the other person's treatment condition. Estimation of the VBM effects with the presence of interference requires advanced techniques (Sobel, 2006). Here, we assume that no interference between units exists, but it must be scrutinized in future analysis. In contrast, we suspect that the SUTVA may not be violated due to the interference *between* treated and control units as discussed by Keele and Titiunik (2018).

D Additional Findings

This section presents additional findings that support our empirical analyses in the main text, including the full results for the 2014 midterm election, results based on the combined control states, placebo tests, and circumstantial evidence for the parallel trends assumption.

D.1 Full Results for 2014

In this subsection, we report the full results for the 2014 midterm election. Figure D.1 visualizes the estimated effect of VBM on voter turnout for different subgroups when using North Carolina (upper panel) and New Mexico (lower panel) as our control states, respectively. We find consistent evidence that the magnitude of the effect does not vary by gender. We do not find any consistent results for the varying size of the VBM effect for other profiles, including age, race and ethnicity, and partisanship.

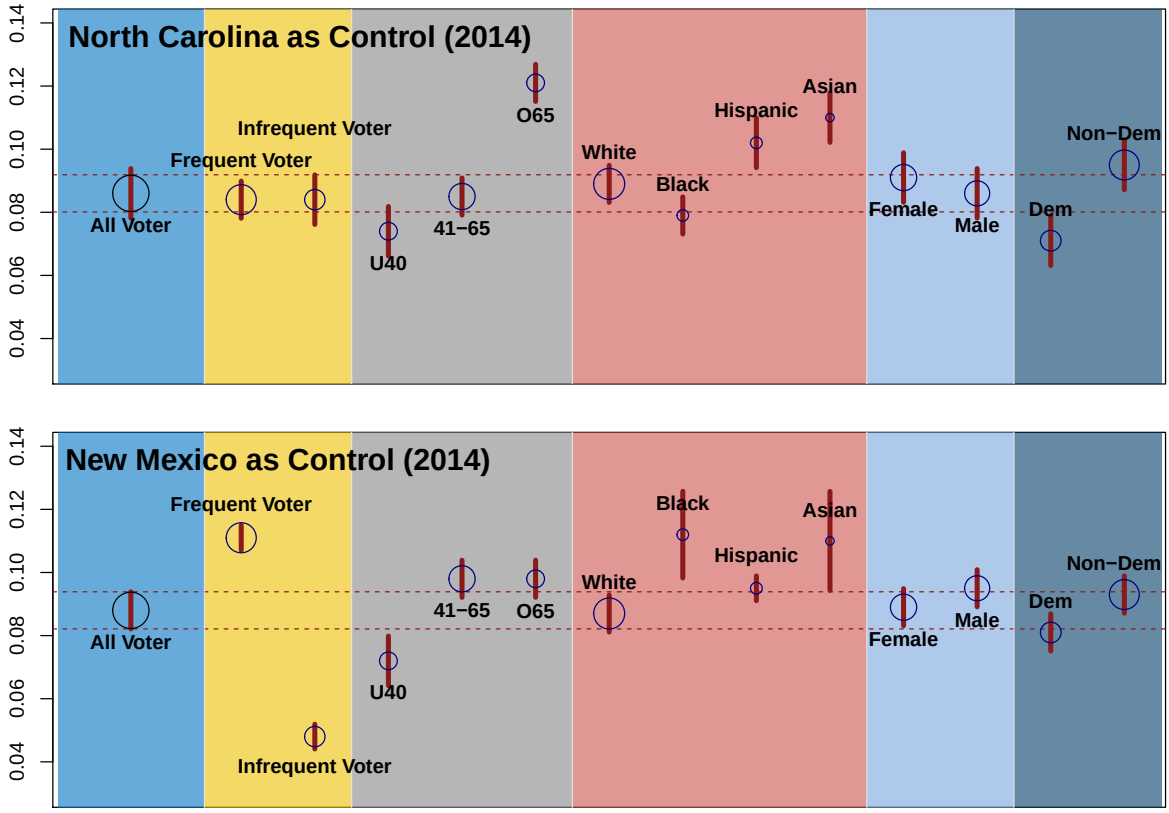


Figure D.1: **Estimated Average Effects of VBM on Turnout with Different Effect Modifiers.**
Note: This figure visualizes the estimated effect (in 2014) with a 95% confidence interval among various subpopulations of interest when using North Carolina (upper panel) and New Mexico (lower panel) as control states. For each group, the size of its open circle is proportional to the square-root of the proportion of the group in Colorado.

D.2 Combining the Two Control States

An alternative approach to estimate our counterfactuals of interest is to combine data from North Carolina and New Mexico in our analysis. Figure D.2 visualizes the estimated effects for frequent and infrequent voters in 2014 and 2016 obtained by this approach, showing that we find evidence against $H_{0,\text{Freq}}$ (i.e., $\tau_{11} > \tau_{21}$), $H_{0,\text{Pres}}$ (i.e., $\tau_{21} < \tau_{22}$), and $H_{0,\text{Mid}}$ (i.e., $\tau_{11} > \tau_{12}$), respectively. Figure D.3 shows the full results for all groups in 2014 and 2016, respectively.

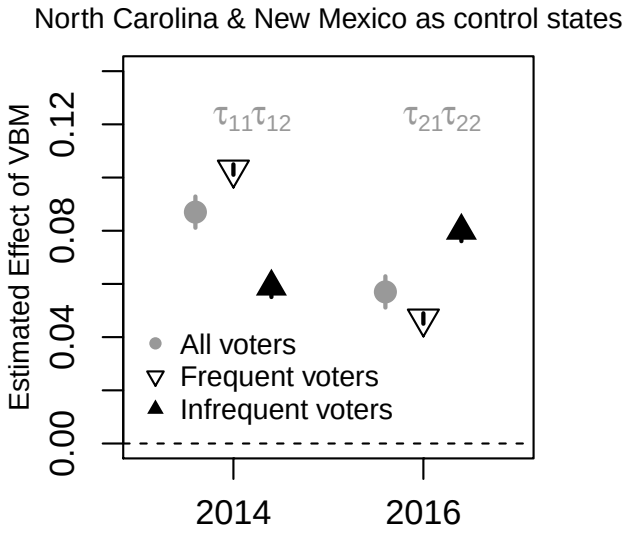


Figure D.2: **Estimated Average Effects in the 2014 Midterm and 2016 Presidential Elections.**
Note: This figure visualizes the estimated VBM turnout effect for frequent and infrequent voters in 2014 and 2016 when using North Carolina and New Mexico (combined) as control states. We find $\tau_{11} > \tau_{21}$ and $\tau_{21} < \tau_{22}$.

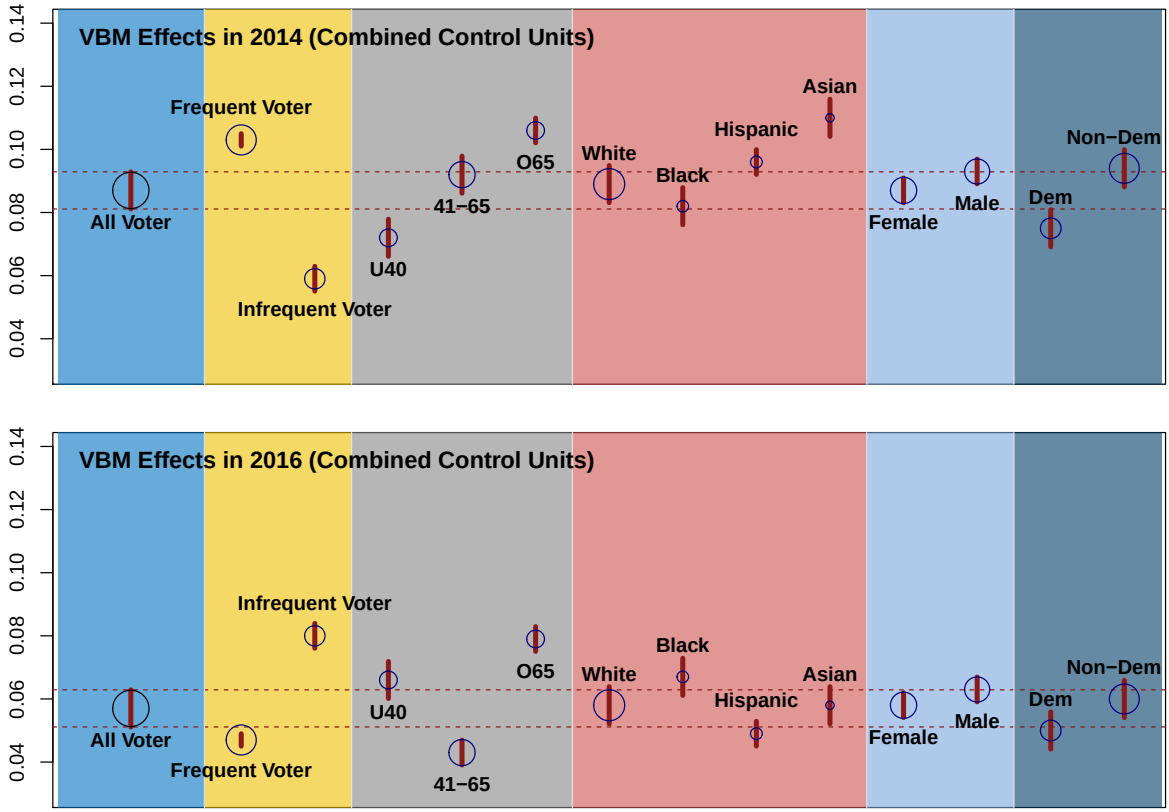


Figure D.3: **Additional Results for the 2014 Midterm and 2016 Presidential Elections.**
Note: This figure visualizes the estimated VBM turnout effect for various subpopulations by voting frequency in 2014 (upper panel) and 2016 (lower panel) when using North Carolina and New Mexico (combined) as control states. For each group, the size of its open circle is proportional to the square-root of the proportion of the group in Colorado.

D.3 Placebo Tests (False Outcomes)

To further confirm the internal validity of our analysis, we perform simple robustness checks. Specifically, we perform a series of placebo tests by replacing the original outcome (i.e., turnout) with several *false outcomes*. In particular, we use a dummy variable for voted in 2010 as well as being White, Black, Hispanic, Asian, Female, and Democrat as our placebo outcome for each test, and we apply the tests for data sets with and without preprocessing. If our identification strategy "works" as we intended, we should expect to find null results (or should not find effects) for these placebo tests. The left panel of Figure D.4 visualizes the results. It indicates that our placebo tests do not find any statistically and/or substantially significant effect on our false outcomes, suggesting that our DID estimation is a valid procedure.

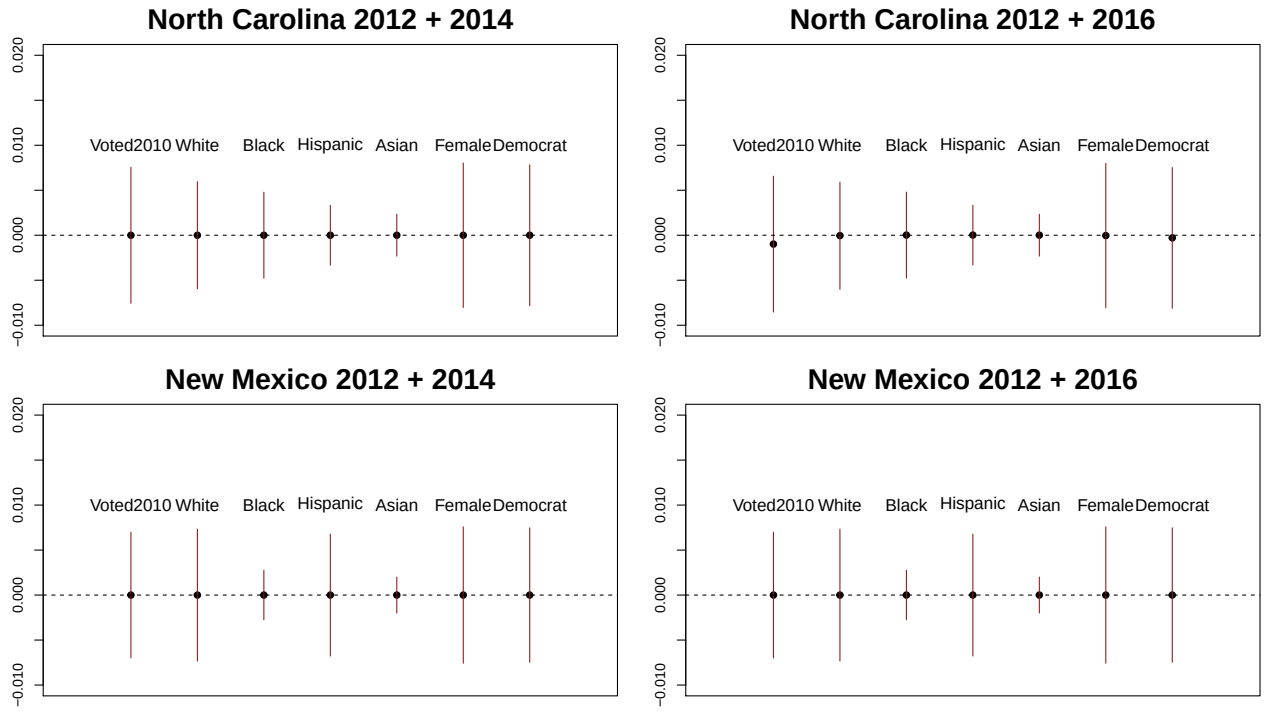


Figure D.4: **Placebo Tests with False Outcomes.**
Note: This figure presents the results of placebo tests with false outcomes.

D.4 Examining the Parallel Trends Assumption

As noted above, the validity of our estimates hinges upon the parallel trends assumption on voter turnout in the two states without the presence of the intervention. However, in our data, no empirical method can verify the assumption, leaving us no choice but to merely believe that the assumption holds. To (at least) provide circumstantial evidence, nonetheless, we visualize voter turnout in both states between 2000 and 2018 using the voting-eligible population (VEP) turnout data from McDonald's United States Elections Project (McDonald, 2008). It should be noted that the VEP is different from our primary population of interest (i.e., a set of voters who had been registered between the 2012 and 2016 elections). With this in mind, we nevertheless check if "overall turnout trends" would look-alike in Colorado and North Carolina over time before the VBM adoption in 2013.

The left panel of Figure D.5 shows that the time trends seem to be fairly similar before the adoption of VBM in Colorado in 2003. From the visual inspection, we could not make a credible claim about the validity of the assumption, but at least have suggestive evidence that the parallel trends assumption is more or less satisfied, which grants a degree of confidence in our causal identification. Nevertheless, this result can only be circumstantial evidence since our main populations of interest are limited within specific groups both in terms of over time registration status and politically salient attributes. We also check the turnout trend in the pre-intervention period in New Mexico using the same data source. The right panel of Figure D.5 suggests that the time trend in voter turnout in this specific population of voters seems to be parallel before 2013, although there seems to be a gap between the two states in 2012. Again, we need to see this result with caution because this result does not directly apply to our population of interest (i.e., the set of voters who had been registered between 2012 and 2016).

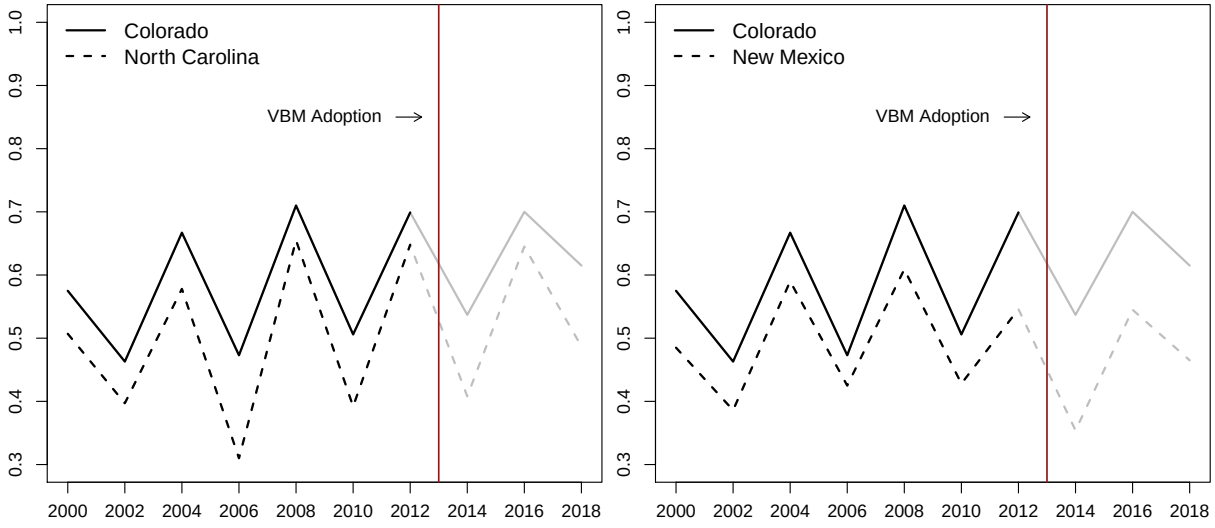


Figure D.5: **Time Trends in Voter Turnout in Colorado, North Carolina, and New Mexico.**
Note: This plot portrays the overtime trend in voter turnout in Colorado and North Carolina based on the highest office Voting-eligible population (VEP) turnout collected by McDonald (2008).

References for Online Appendix

Abadie, Alberto. 2005. "Semiparametric difference-in-differences estimators." *The Review of Economic Studies* 72(1):1–19.

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Barber, Michael and John B Holbein. 2020. "The participatory and partisan impacts of mandatory vote-by-mail." *Science Advances* 6(35):eabc7685.

Bergman, Elizabeth and Philip A Yates. 2011. "Changing election methods: How does mandated vote-by-mail affect individual registrants?" *Election Law Journal* 10(2):115–127.

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics* 119(1):249–275.

Bonica, Adam, Jacob M. Grumbach, Charlotte Hill and Hakeem Jefferson. 2020. "All-Mail Voting in Colorado Increases Turnout and Reduces Turnout Inequality." *Working Paper* .

Gerber, Alan S, Gregory A Huber and Seth J Hill. 2013. "Identifying the effect of all-mail elections on turnout: Staggered reform in the evergreen state." *Political Science Research and Methods* 1(1):91–116.

Gronke, Paul and Peter Miller. 2012. "Voting by mail and turnout in Oregon: Revisiting Southwell and Burchett." *American Politics Research* 40(6):976–997.

Ho, Daniel E, Kosuke Imai, Gary King and Elizabeth A Stuart. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political analysis* 15(3):199–236.

Imai, Kosuke and Kabir Khanna. 2016. "Improving ecological inference by predicting individual ethnicity from voter registration records." *Political Analysis* 24(2):263–272.

Keele, Luke and Rocío Titiunik. 2018. "Geographic Natural Experiments with Interference: The Effect of All-Mail Voting on Turnout in Colorado." *CESifo Economic Studies* 64(2):127–149.

Keele, Luke and William Minozzi. 2013. "How much is Minnesota like Wisconsin? Assumptions and counterfactuals in causal inference with observational data." *Political Analysis* pp. 193–216.

Khanna, Kabir and Kosuke Imai. 2019. *wru: Who are You? Bayesian Prediction of Racial Category Using Surname and Geolocation*. R package version 0.1-9.
URL: *https://CRAN.R-project.org/package=wru*

Kousser, Thad and Megan Mullin. 2007. "Does voting by mail increase participation? Using matching to analyze a natural experiment." *Political Analysis* 15(4):428–445.

Larocca, Roger and John S Klemanski. 2011. "US state election reform and turnout in presidential elections." *State Politics & Policy Quarterly* 11(1):76–101.

McDonald, Michael. 2008. "United States election project." *United States Elections Project* .

PantheonAnalysis. 2017. "Utah 2016: Evience for the positive turnout effects of Vote at Home (also known as vote by mail) in participation counties.".

PantheonAnalysis. 2018. "Colorado 2014. Comparisions of Predicted and Actual Turnout.".

Richey, Sean. 2008. "Voting by mail: Turnout and institutional reform in Oregon." *Social Science Quarterly* 89(4):902–915.

Sled, Sarah Marie. 2008. "It's in the mail: The effect of vote by mail balloting on voter turnout and policy outcomes in US elections.".

Sobel, Michael E. 2006. "What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference." *Journal of the American Statistical Association* 101(476):1398–1407.

Southwell, Priscilla L. 2009. "Analysis of the turnout effects of vote by mail elections, 1980–2007." *The Social Science Journal* 46(1):211–217.

Southwell, Priscilla L and Justin I Burchett. 2000. "The effect of all-mail elections on voter turnout." *American Politics Quarterly* 28(1):72–79.

Thompson, Daniel M, Jennifer A Wu, Jesse Yoder and Andrew B Hall. 2020. "Universal vote-by-mail has no impact on partisan turnout or vote share." *Proceedings of the National Academy of Sciences* .

Wing, Coady, Kosali Simon and Ricardo A Bello-Gomez. 2018. "Designing difference in difference studies: best practices for public health policy research." *Annual review of public health* 39.