

Dual Effects of Vote-by-Mail Elections on Voter Turnout*

Yuki Atsusaka[†]
atsusaka@rice.edu

Robert M. Stein[‡]
stein@rice.edu

Andrew Menger
andrewmmenger@gmail.com

September 23, 2020

Abstract

There is growing evidence that Voter-by-Mail (VBM) elections increase voter turnout. However, previous research lacks a theoretical explanation for *how* VBM elections increase voter turnout, making the assessment of the election reform challenging. We identify two mechanisms through which VBM elections increase turnout among registered voters: notification about upcoming elections and a reduction in the costs of voting. We offer evidence for our explanations with more than 30 millions of individual voting records from Colorado, North Carolina, and New Mexico. A difference-in-differences analysis with exact matching shows that the VBM implementation significantly boosted turnout among frequent and older voters both in the 2014 midterm and 2016 presidential elections, suggesting the presence of a convenience effect. We also find that VBM elections substantially raised turnout among infrequent and younger voters only in 2014, implying the presence of a notification effect. Our theoretical framework not only assists future empirical studies of VBM elections, but also offers a reasonable explanation for previous studies that find that VBM effects do not vary substantially by voters' race, gender, and partisanship.

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

Word Count: 3791

*This study is supported by the New Venture Fund. For helpful comments, we would like to thank Vin Arceneaux, Barry Burden, Shiro Kuriwaki, Randy Stevenson, Soichiro Yamauchi, and 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 up to three weeks 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 mode 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 been further highlighted amid the COVID-19 pandemic, and it has been shown to enhance voter turnout without any evidence of partisan advantage (Barber and Holbein, 2020; Thompson et al., 2020; Bonica et al., 2020). Despite the large body of research on the turnout effects of VBM elections, the literature has not provided a unified explanation of *why* VBM could increase voter turnout, if any. Unfortunately, the lack of *theoretical* explanations makes it difficult for researchers and more importantly election administrators to assess whether VBM actually increases turnout among voters and how.

To resolve this issue and advance the research of VBM and political participation more broadly, we identify two mechanisms through which VBM may increase voter turnout. By focusing on the population of 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 We further theorize that the notification effect is more salient in low-informational races (e.g., local and midterm elections) than in high-informational contests (e.g., presidential elections) where most voters are already aware of upcoming elections. There are two key observable implications from our explanation for how VBM elections increase voter turnout. First, VBM elections increase turnout in high-informational elections (e.g., presidential elections) mostly among frequent and older (and thus more informed) voters who find it costly to return their ballots to the polling stations on Election Day. Second, VBM elections boost turnout among all registered voters including infrequent and younger (and thus less informed) voters in low-informational elections (e.g., midterm congressional elections) by notifying less attentive voters of an upcoming election *and* reducing the physical cost of voting for all voters.

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

²More 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 ballot on voter turnout.

We test these implications by using more than 30 millions individual voting records from Colorado, which adopted VBM in 2013, as well as North Carolina and New Mexico, which have never used VBM elections.³ For the 2014 midterm and 2016 presidential elections, we estimated the conditional average treatment effect on the treated (CATT) of VBM elections 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 frequent voters and older (as more informed) voters in both the 2014 midterm and 2016 presidential elections, suggesting the presence of the hypothesized convenience effect, whereas it substantially raised turnout among infrequent and younger (as less informed) voters only in 2014, implying the presence of a notification effect. Our theoretical framework not only assist future empirical studies of the consequences of VBM elections, but also offers a reasonable explanation for previous findings that VBM effects do not vary substantially by voters' race, gender, and partisanship.

Dual Effects of Notification and Convenience

We identify two ways VBM elections 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.⁴ We call these as 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:

$$\text{VBM Effect} = \text{Notification Effect} + \text{Convenience Effect} \quad (1)$$

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

³Colorado has made a statewide adoption of VBM as opposed to adoptions in a staggered fashion.

⁴We focus on voters who are “already registered to vote” at the time of VBM adoption for several reasons. First and foremost, VBM is intended to offer an alternative mode of voting primarily for registered voters. Second, even if VBM intends to encourage unregistered voters to be registered, it is unclear why and how such “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 clear our the theoretical problem, quantifying the notification, convenience, and registration effects will be extremely difficult on the empirical ground. For these reasons, we believe that it is reasonable to focus on registered voters when theorizing the effect of VBM adoption.

Strauss, 2009, 787).⁵

We argue that VBM increases voter turnout by providing registered voters with such a “reminder” that is noticeable, unavoidable and proximate to when they are likely to vote (i.e., complete their mail ballot). We further theorize that the notification is more salient among voters who are less informed about elections (e.g., infrequent and younger voters) and in low-informational contests such as local and midterm elections.⁶ In contrast, in high-informational contexts such as presidential elections, we expect the notification effect is negligible and formulate the VBM effect as follows:

$$\text{VBM Effect} \approx \text{Convenience Effect} \quad (2)$$

Our claim on the convenience effect draws from the literature of the cost 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 afford voters the opportunity to complete their ballot over several days (or even weeks) and at home or the location of their choosing and to return their ballots via mail. These features of VBM afford voters significant convenience and reduced costs, which has been shown to result in a higher number of cast ballots and completed ballots in VBM elections (Menger, Stein and Vonnahme, 2018). We claim that the convenience effect is present for all voters, but more salient among voters who are more susceptible to the physical cost of voting such as older voters.

The key observable implications from our theory are summarized in Figure 1. It illustrates that, in low-informational contexts, VBM has a dual effect of notification and convenience among uninformed and infrequent voters, whereas it has a convenience effect among informed and frequent voters. It also suggests that, in high-informational contexts, VBM primarily has a convenience effect among informed and frequent voters and a negligible convenience effect among uninformed and infrequent voters.⁷ Put differently, we must observe *varying magnitudes* of VBM effects by subpopulations (informed v. uninformed voters as well as frequent v. infrequent voters) and by the type of elections (low v. high-informational races). Below,

⁵The literature on voter mobilization (Green and Gerber, 2019; Green, Gerber and Nickerson, 2003; Nickerson, 2005, 2007; Michelson, 2006) links voter turnout to highly personalized messaging from candidates and parties. Dale and Strauss (2009) note, however, that empirical findings on personal and impersonal campaign messaging are not consistent on this point (Michelson, 2006; Nickerson, 2007; Green and Gerber, 2019, 37)

⁶This is because there is no efficacy in reminding voters of upcoming elections if almost all voters already know about them.

⁷Of course, this is rather a simplified view of the reality; it could be possible to identify a set of informed and frequent voters in low-informational contexts who might be benefited from a noticeable reminder by VBM.

we test these implications by comparing the magnitude of causal effects of VBM on turnout among different subpopulations. Meanwhile, we expect that the effect of VBM does not vary substantially by other subpopulations (e.g., by race, party, and gender) assuming that each group has a relatively similar level of informed and frequent voters.

	Uninformed & infrequent voters	Informed & frequent voters
Low-informational contexts	<i>notification + convenience</i>	<i>convenience</i>
High-informational contexts	<i>(convenience)</i>	<i>convenience</i>

Figure 1: **Implications of Dual Effects of VBM**

Note: This figure shows expected effects by subpopulations and by the types of elections. Uninformed and infrequent voters (left column) have both the notification and convenience effect in low-informational contests, but only the negligible notification in high-informational races. Informed and frequent voters (right column) are expected to have the convenience effect regardless of the types of elections.

Data and Identification Strategy

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

$$\tau_{CATT} = \mathbb{E}_{\mathbf{X}} \left[\mathbb{E}[Y_i^{d=1} - Y_i^{d=0} | D_i = 1, Z_i = z, \mathbf{X}_i] \right], \quad (3)$$

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 covariates. For effect modifiers, we consider whether a voter is a frequent voter, belongs to a particular racial group (White, Black, Hispanic, Asian, or other race), is female, is 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 newly collected voter history files from the 2012, 2014, 2016 presidential and midterm elections in Colorado and North Carolina. Colorado has adopted VBM elections in 2013, whereas North Carolina has not adopted VBM and is used as a counterfactual state.⁹ Substantively,

⁸Substantively, this means that we study VBM effects among voters to whom the election reform is targeted at and we are interested in how these effects vary by subpopulations (see Online Appendix C.1 for details). We code a voter as a frequent voter if she has voted in the 2010 midterm election. For age groups, we classify voters whether they are less than 35, between 35 and 65, or over 65.

⁹More specifically, we collected voter history files and registration records, coded individual race, imputed missing data, and examined descriptive statistics (see Online Appendix B).

this means that we test our hypotheses with individual-level turnout data in two states before (i.e., 2012) and after the intervention (VBM adoption) occurs (i.e., 2014 and 2016, respectively).¹⁰

For our identification strategy, we adopt a parametric linear difference-in-differences (DID) model. The key idea is to simulate voter turnout in Colorado (both in 2014 and 2016) had the VBM policy not been adopted and use such counterfactual turnout to estimate the causal effect of the policy adoption. We do so by using the information from the pre-intervention period in Colorado and information from North Carolina. 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.¹¹ Although we provide suggestive evidence that the assumption seems to hold in our setting (Online Appendix C.4), this assumption is inherently nonrefutable (i.e., cannot be empirically verified with *any* data) and our inferences could be jeopardized by the potential violation of the assumption.¹²

To mitigate this concern, we employ a semi-parametric approach in which we carry out exact matching prior to the application of the DID model so that we can restrict the set of control units to those with similar sociopolitical attributes to treated units. Specifically, we perform exact matching to preprocess the original data and then implement a DID model by a saturated weighted least squares regression, where weights (for control units) are obtained from exact matching (see also Online Appendix C.4).¹³ We further check the validity of our inferences by adopting the idea of *pattern specificity* (Rosenbaum, 2005) to compare our treated units to different control units from New Mexico, making sure that a common pattern emerges regardless of the choice of control units.

Empirical Findings

First, we examine the implication for the convenience effect. We do so by comparing the CATTs for different subpopulations in the 2016 election, where we expect the notification effect to be negligible. If our theory is plausible, we must observe that the VBM effect is larger for voters who would be most benefited from the reduction of physical cost (or increase in convenience) in voting, including older voters and frequent voters,

¹⁰Importantly, Colorado has simultaneously adopted VBM and the Same Day Registration (SDR) in 2013. However, because we limit the scope of inquiry to voters who had been registered between 2012 and 2016, the joint adoption of SDR does not affect our causal identification.

¹¹For 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.

¹²We also make the linearity assumption in the conventional DID model.

¹³This semi-parametric approach has been considered in O'Neill et al. (2016) and shown to reduce biases that could appear when the parallel trends assumption is violated in Ryan et al. (2019).

while it is negligible among less informed voters such as younger voters and infrequent voters. Figure 2 displays the estimated CATTs in 2016 with different effect modifiers based on control units from North Carolina (upper panel) and New Mexico (lower panel), respectively.

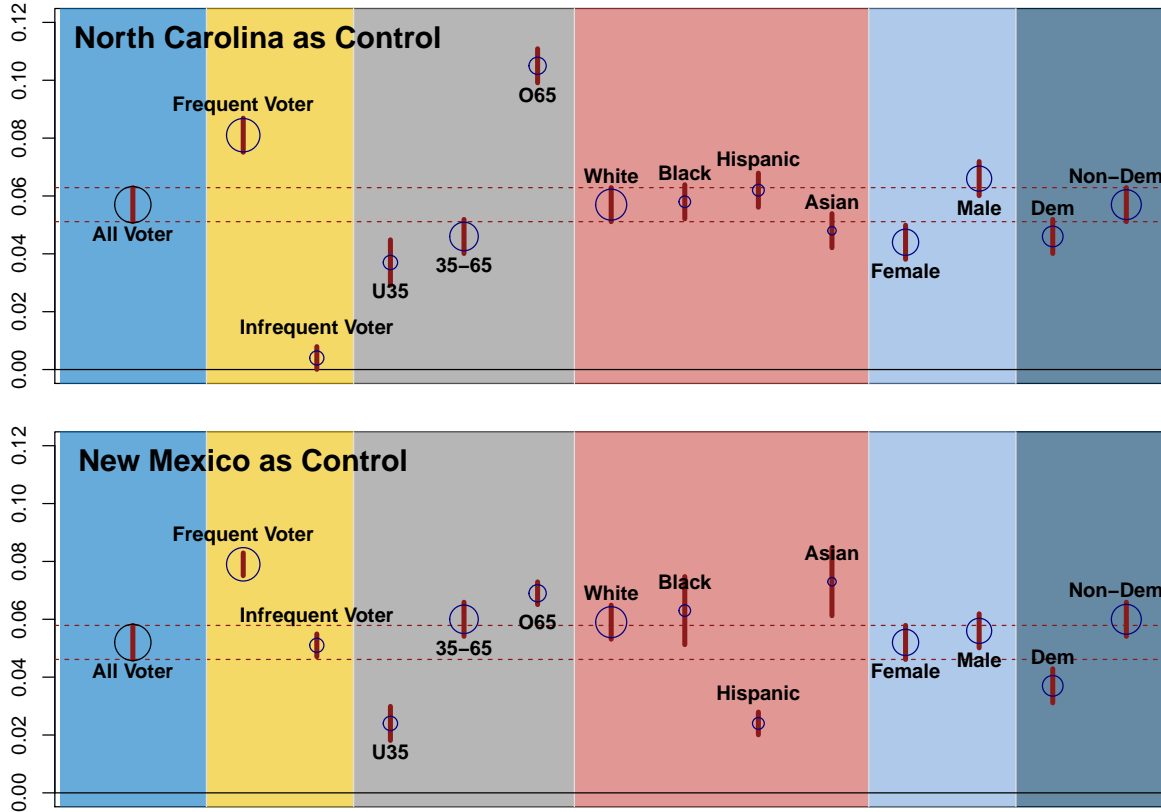


Figure 2: **Estimated Average Effects of VBM on Turnout with Different Effect Modifiers.**

Note: This figure visualizes the estimated effect (in 2016) on 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.

Consistent with our expectation, we find that the VBM effects are significantly larger among frequent voters and older voters than among infrequent voters and younger voters. Moreover, the result shows that the effect size does not vary much for other subpopulations. Notably, we did not find that VBM increases turnout only among Democratic or non-Democratic voters. The results imply that most of the VBM effects are from what we formulate as the convenience effect rather than the notification effect in the 2016 election. The finding for frequent voters is particularly consequential because it implies that VBM primarily increases turnout among voters who are already well mobilized in the presidential election.

Next, we test the implication for the notification effect by comparing the CATTs for voters both in 2014

and 2016. We expect to observe two results: (1) the VBM effect is small or even negligible among infrequent voters and younger voters in the 2016 presidential election and (2) the VBM effects among infrequent voters and younger voters are larger in 2014 than in 2016 because the notification effect is expected to be influential in (relatively) low-information races such as midterm elections. Figure 3 displays our results, showing that we indeed observe evidence for these expectations.

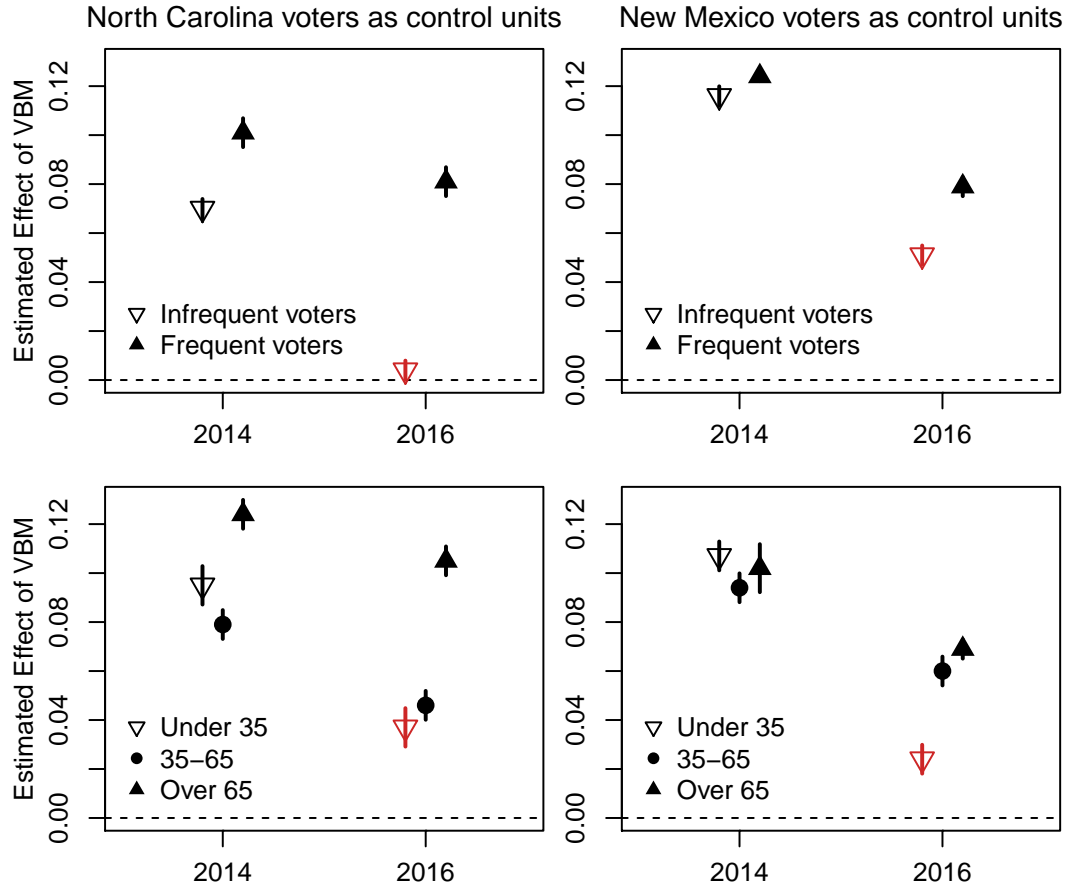


Figure 3: Estimated Average Effects in the 2014 Midterm and 2016 Presidential Elections.

Note: This figure visualizes the estimated effect on various subpopulations of interest when using North Carolina (left panels) and New Mexico (right panels) as control states. We expect the “red effects” in 2016 to be close to zero and smaller than their corresponding effects in 2014.

The figure presents the estimated effects for infrequent and frequent voters (upper panels) and younger, middle-age, and older voters (lower panels) based on North Carolina voters (left panels) and New Mexico voters (right panels) as control units, respectively. It shows that the estimated effects for infrequent voters and younger voters are substantially lower in 2016 (highlighted in red) than in 2014. It is also worth noting that, for every subpopulation, the estimated effect is higher in 2014 than in 2016. This is indeed natural given

our theoretical story, which claims that the total VBM effect is a *dual effect* of convenience and notification in low-informational contests, whereas it is a sole effect of convenience in high-informational races.

Recognizing the limitations in making causal inferences in observational studies, we performed a number of robustness checks to confirm that our conclusion does not depend on specific decisions we made during our data collection, coding, modeling, and statistical inferences. Online Appendix D offers additional findings, robustness checks, and circumstantial evidence for the parallel trends assumption, confirming that our substantive conclusion remains unchanged. Taken together, we find supportive evidence for our theoretical claim.

Implications for Future Research

To advance the research of VBM and election reform, we have identified two mechanisms for how VBM elections affect voter turnout: notification and convenience. Our theoretical framework not only assists future evaluation of the VBM effect on voter turnout but also helps political scientists understand previous findings on varying (or not varying) magnitudes of VBM effects by frequency of voting and age as well as race, gender, and partisanship.

Testing the theory of dual effect of VBM on voter turnout more directly, however, presents several challenges. Observational studies of VBM elections are limited in their ability to deal with potential 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

- 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.
- Green, Donald P and Alan S Gerber. 2019. *Get out the vote: How to increase voter turnout*. Brookings Institution Press.
- Green, Donald P, Alan S Gerber and David W Nickerson. 2003. "Getting out the vote in local elections: Results from six door-to-door canvassing experiments." *The Journal of Politics* 65(4):1083–1096.
- 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, Robert M Stein and Greg Vonnahme. 2018. "Reducing the Undervote With Vote by Mail." *American Politics Research* 46(6):1039–1064.
- Michelson, Melissa R. 2006. "Mobilizing the Latino youth vote: Some experimental results." *Social Science Quarterly* 87(5):1188–1206.
- Nickerson, David W. 2005. "Partisan mobilization using volunteer phone banks and door hangers." *The Annals of the American Academy of Political and Social Science* 601(1):10–27.
- Nickerson, David W. 2007. "Quality is job one: Professional and volunteer voter mobilization calls." *American Journal of Political Science* 51(2):269–282.
- 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: Evidence for the positive turnout effects of Vote at Home (also known as vote by mail) in participation counties."

- PantheonAnalysis. 2018. “Colorado 2014. Comparisons 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 “Dual Effects of Vote-by-Mail Elections on Voter Turnout”

Table of Contents

A	VBM Elections and Previous Research	2
A.1	Brief History of VBM Elections	2
A.2	Previous Findings	2
B	Additional Information on Data	4
B.1	Data Collection	4
B.2	Coding Individual Race	4
B.3	Identification for Missing Data	5
C	Additional Discussion on Identification	6
C.1	Populations and Quantities of Interest	6
C.2	Using North Carolina and New Mexico as Counterfactuals	7
C.3	Preprocessing Data	7
C.4	Discussion on Assumptions for Identification	11
D	Additional Findings	12
D.1	Supporting Evidence for Main Results	12
D.2	Placebo Tests	14
D.3	Examining the Parallel Trends Assumption	14

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. Unlike absentee voters, registered voters in VBM-only precincts are automatically mailed a ballot whether they request one or not with a postage-paid return envelope. These voters do not have the option of voting in person. The only other development is the number of states that have adopted no-excuse absentee voting for the November election. But these adoptions are not permanent and only in response to the COVID-19.

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 effect based on stronger causal identifications than did earlier studies. 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 a similar experimental leverage, Thompson et al. (2020) report that VBM elections increase turnout by two to three percentage points, with no evidence of a partisan advantage. More recently, Bonica et al. (2020) show report turnout increases in excess of 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 reserch on the effects of VBM on 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 and North Carolina for the 2010, 2012, and 2016 elections. These files represent “snapshots” taken at the time of each election, and are made available online by the Colorado Secretary of State and the North Carolina State Board of Elections, respectively. These snapshot in time data mean that we are able to examine the behavior of all individuals registered to vote at the time of each election. These data stand in contrast to a “contemporary” voter file – that is, the voter file that a state elections office could generate on any given day – that will contain vote history data over time, but only for those voters who would be eligible to vote if an election were to be held on that day. It would not include voters who were registered to vote in each of the last three presidential elections, but who have died, moved out of state, or have become inactive after the 2016 election.

We began by keeping only those voters who were registered to vote on Election Day in all 2012 and 2016. 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 voting history for the 2014 midterm Congressional election. We first code for whether each person voted in the election, As required by federal law, The North Carolina voter files includes the voters self-identified race and ethnicity. This information is not required of Colorado voters and was imputed.

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 individual race. 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:

$\Pr(\text{Black} = 0.75 \mid \text{Smith, BlockGroup1001, age30, female, and democrat})$
 $\Pr(\text{White} = 0.15 \mid \text{Smith, BlockGroup1001, age30, female, and democrat})$
 $\Pr(\text{Asian} = 0.05 \mid \text{Smith, BlockGroup1001, age30, female, and democrat})$
 $\Pr(\text{Hispanic} = 0.05 \mid \text{Smith, BlockGroup1001, age30, female, and democrat})$

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 Tract each voter belongs to. After coding this information, we computed the above posterior probabilities by using `wru` in R developed 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 Identification for Missing Data

Like many other administrative records, our data contains missing values on multiple covariates. We assume that all variable, except for turnout in 2010, follows missing completely at random (MCAR). With the MCAR assumption, we implemented listwise deletion for (i.e., dropped) voters who contain any missing value on the covariates, except for turnout in 2010. For some unidentified reason, 58.08 % of this variable has missing values for Colorado voters.

For the variable denoting whether each unit voted in 2010 (hereafter `voted2010`), we impute missing values in three different ways. The first two strategies are based on the idea of partial identification. Specifically, we bound the estimated effects by using the lowest possible value and the highest possible value for `voted2010`. Let V_i be `voted2010`. Because V_i is a binary variable with a statistical support: $\text{Supp}[V_i] = \{0, 1\}$, this can be done by imputing 0 (the lowest possible value) and 1 (the highest possible value) in all missing values, respectively. In principle, we can then bound the lowest and highest possible values for the total VBM effects by applying our (causal) identification strategy to the two imputed data sets. Hence, we can expect that the “true” effects would be located somewhere between the lower and upper bounds.

The third strategy is based on the missing at random (MAR) assumption and uses a binary choice model to impute the missing values. Informally, we estimate a logistic regression with V_i as the dependent variable and other covariates as predictors (while deleting units with missing values). We then generate predicted probabilities for V_i and code 0 if the predicted probability is less than 0.5 and 1 if it is greater than or equal to 0.5. The expectation is that this enables us to point identify the missing value for each `voted2010` (if missing) and that the estimated effects based on this imputed data would be between the above bounded effect estimates. Below, Table D.1 shows that this is actually the case and our main results are based on the third approach.

More formally, let $M_i \in \{0, 1\}$ be a binary variable denoting whether a value is missing for V_i . We then assume that the following two conditions are satisfied: $V_i \perp\!\!\!\perp M_i | \mathbf{X}_i$ (independence of `voted2010` and the missingness conditional on covariates) and $\Pr(M_i = 1 | \mathbf{X}_i) > 0$ (nonzero probability of voting in 2010 conditional on covariates). With these assumptions, we can show that $\Pr(V_i = 1 | M_i = 0, \mathbf{X}_i) = \Pr(V_i = 1 | \mathbf{X}_i) = \Pr(V_i = 1 | M_i = 1, \mathbf{X}_i)$. This implies that we can use the (population) predicted probabilities for V_i given covariates in non-missing data to impute the (population) predicted probabilities in missing data. Since we do not know the population probabilities for non-missing data, we estimate them using a logistic regression with all other covariates assuming that it is a good approximation of the underlying conditional expectation function. According to a plug-in principle, we then use $\hat{\Pr}(V_i = 1 | M_i = 0, \mathbf{X}_i)$ (predicted probabilities based on the logistic regression) to impute $\hat{\Pr}(V_i = 1 | M_i = 1, \mathbf{X}_i)$. We then code if 0 if the predicted probabilities are less than 0.5 and 1 if greater than or equal to 0.5.

C Additional Discussion on Identification

C.1 Populations and Quantities of Interest

Our primary population of interest is *a set of voters who had been registered in Colorado between the 2012 and 2016 elections*. In addition, 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 *or* 2016 presidential election.

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 (with respect to the primary population or “all voters”) as the average treatment effect on the treated (ATT):

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

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 actually 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):

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

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.

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

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 a multidimensional 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 has used VBM elections, but offers all the other modes 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, 4% in North Carolina, and 9% in New Mexico (SOURCE?). Both states have a racially and ethnically diverse population. 36% of North Carolinas population is non-white white and 31% of Colorados population is non-white. 19% of New Mexico’s population is non-white, with 48% Hispanic (SOURCE?). 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 (SOURCE?). A fifth of all three states population (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.

	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**

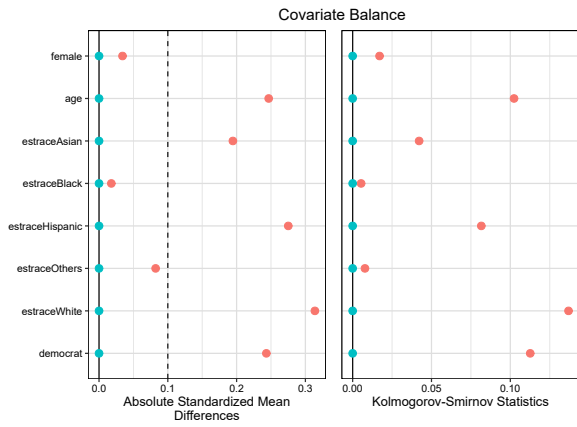
Note: This table shows the comparison between Colorado and North Carolina on several state-level variables of interest. The same day registration is available in Colorado since 2013 and in North Carolina since 1994.

C.3 Preprocessing Data

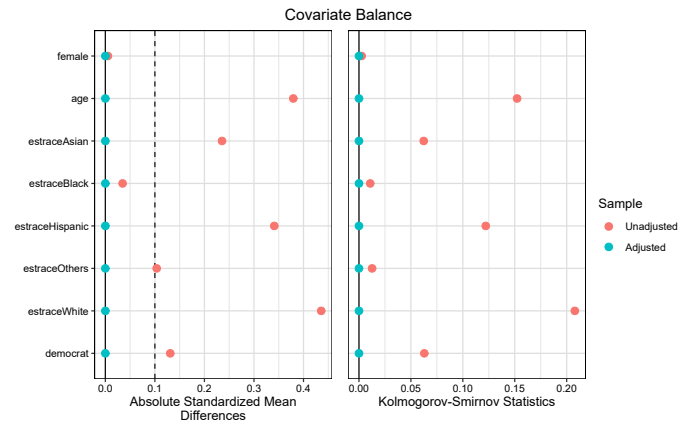
For credible observational studies, it is critical to reduce the imbalance in the covariate distributions between treated and control units in data so that our estimation and inference will be less model dependent (Ho

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 extraordinary large size of observations (≈ 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.

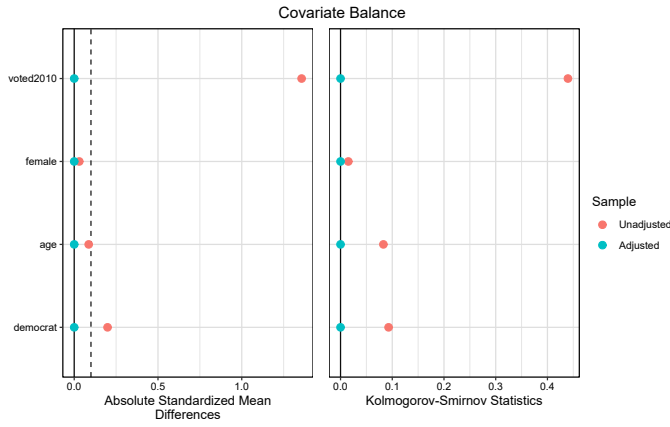
As a quick visual inspection, Figure C.1 compares the standardized bias and Kolmogorov-Smirnov statistics for each covariate before and after matching for each population of interest. The results demonstrate that after preprocessing our data achieves a higher balance on covariates than raw data.



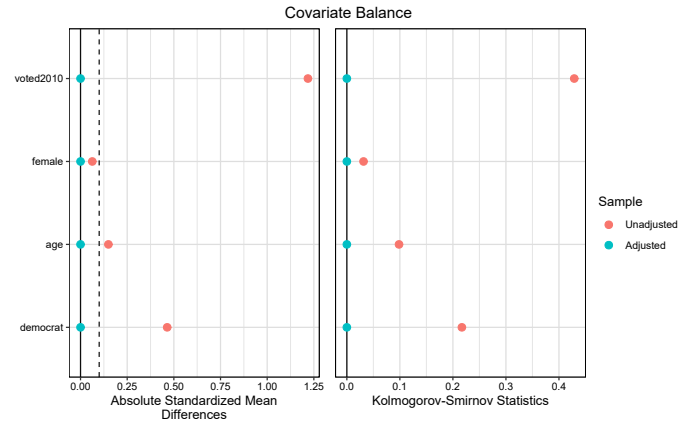
(a) Frequent voters



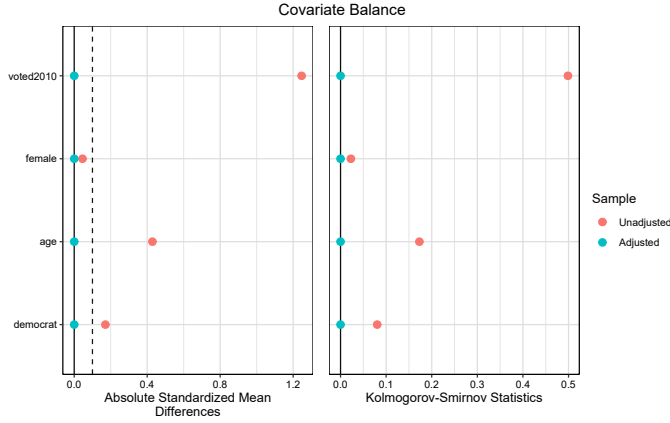
(b) Infrequent voters



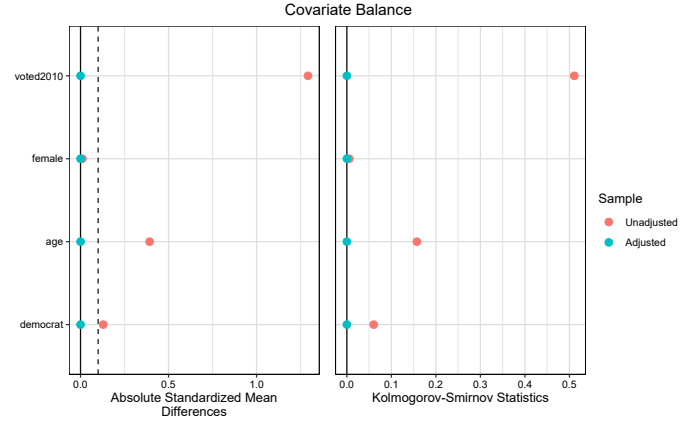
(c) White



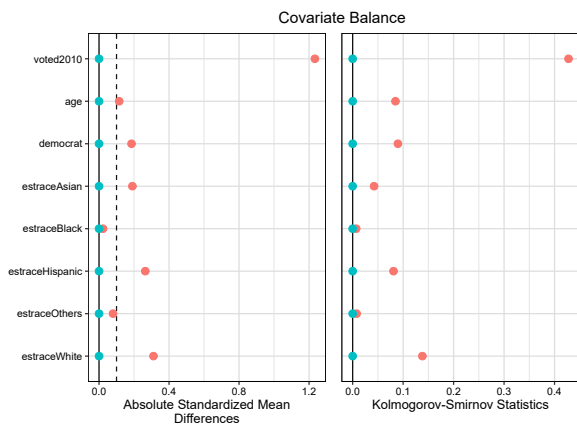
(d) Black



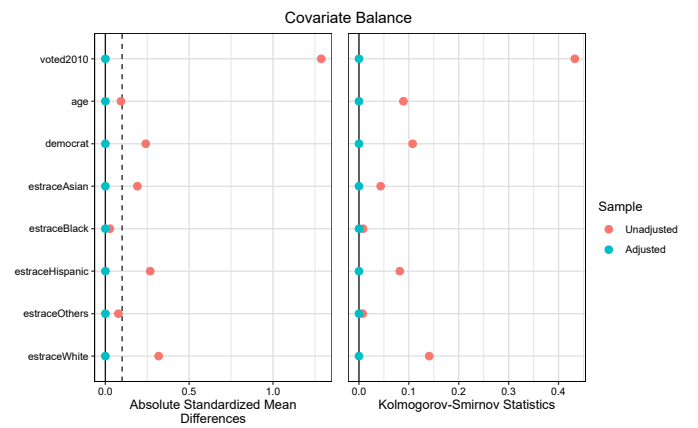
(e) Hispanic



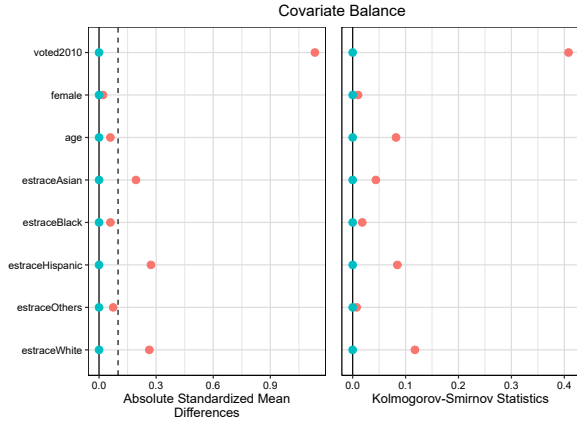
(f) Asian



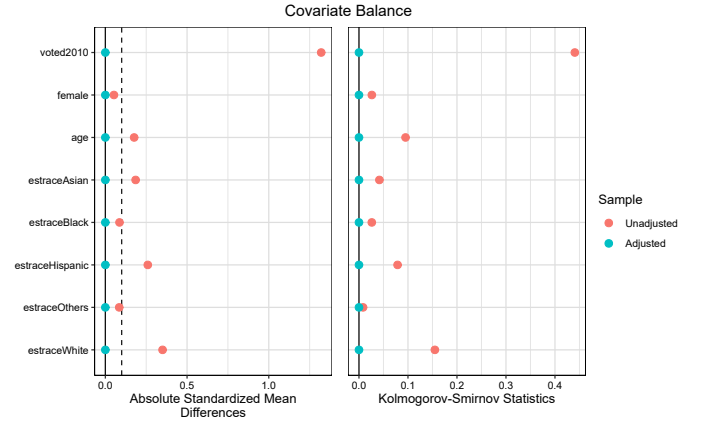
(g) Female



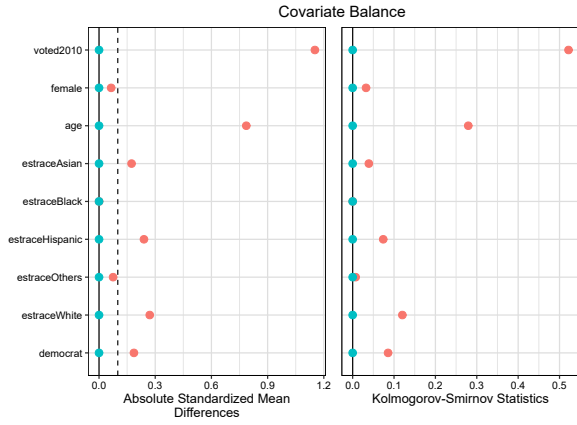
(h) Male



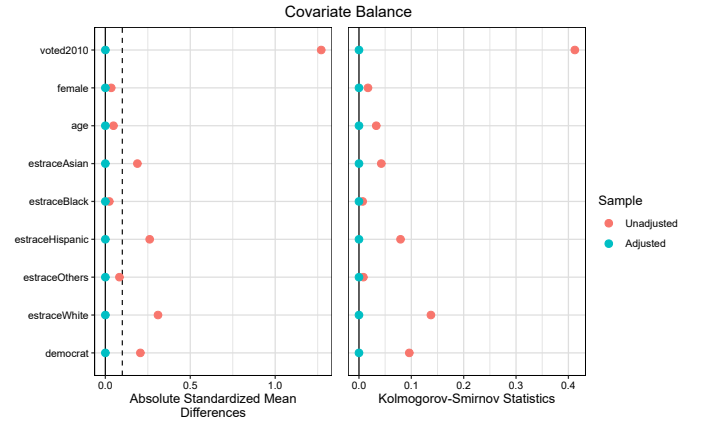
(a) Democrat



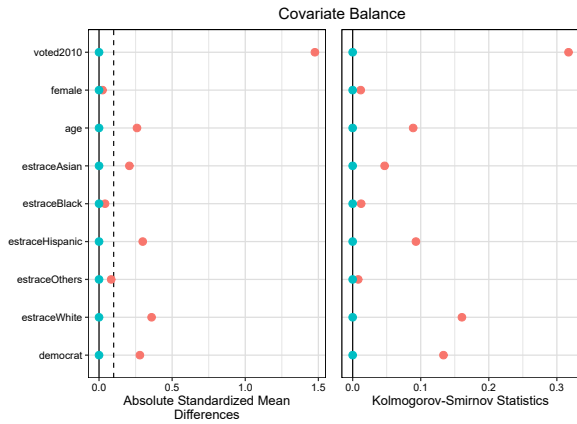
(b) non-Democrat



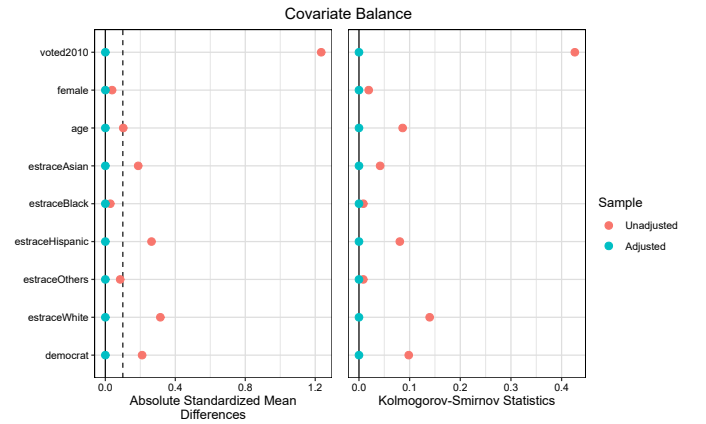
(c) Under 35



(d) 35 to 65



(e) Over 65



(f) All voters

Figure C.1: Covariate Balance After Preprocessing

Note: This figure visualizes the covariate balance before and after preprocessing data. We preprocess our data with exact matching (one-to-one matching without replacement) on four covariates. 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 by replacing the true intervention with a false intervention). In Online Appendix D.3, nevertheless, we offer circumstantial evidence that the turnout trends could be the same in the two states absence 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 own 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 an interference between the two where one’s outcome is now a function of her own 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

D.1 Supporting Evidence for Main Results

To confirm that our substantive conclusion is robust to several decisions that we made to derive our main results, we estimate our quantities of interest by not performing preprocessing (exact matching) and using different strategies to impute voted2010 variable (discussed in Online Appendix B.3). The results are displayed in Table D.1, which confirms that our substantive conclusion remains the same.

Our main results are based on Column (4). Columns (1)-(3) are results based on a saturated (difference-in-differences) OLS without preprocessing, while Columns (4)-(6) are results based on a saturated weighted least squares (WLS) with preprocessing. Columns (1) and (4) are results from data where missing values for voted2010 were imputed by logistic regressions, whereas Columns (2) and (5) and Columns (3) and (6) are results based on partial identifications in which the lowest and highest possible values (i.e., 0 or 1) are used for imputation, respectively.

It is worth noting that the estimated effects for all categories in Column (4) are located between the estimated effects in Column (5) and Column (6), which offer lower and upper bounds of the estimated effects via partial identification (for missing data). More importantly, our substantive conclusions (i.e., effects are larger among frequent voters than infrequent voters, effects are larger among old voters than other voters, effects do not vary much by other groups) are not susceptible to the different “specifications” of our models.

Saturated (Difference-in-Differences) OLS/WLS (North Carolina)						
	(1)	(2)	(3)	(4)	(5)	(6)
All voters	0.048 (0.001)	0.059 (0.001)	0.059 (0.001)	0.057 (0.001)	0.041 (0.003)	0.069 (0.003)
Frequent	0.082 (0.001)	0.061 (0.001)	0.093 (0.001)	0.081 (0.003)	0.054 (0.002)	0.081 (0.003)
Infrequent	0.041 (0.001)	0.050 (0.001)	-0.015 (0.001)	0.004 (0.002)	0.035 (0.002)	-0.035 (0.002)
White	0.049 (0.001)	0.059 (0.001)	0.059 (0.001)	0.057 (0.003)	0.042 (0.003)	0.066 (0.003)
Black	0.057 (0.001)	0.067 (0.002)	0.067 (0.001)	0.058 (0.003)	0.048 (0.003)	0.068 (0.003)
Hispanic	0.038 (0.002)	0.048 (0.002)	0.048 (0.002)	0.062 (0.003)	0.050 (0.003)	0.074 (0.003)
Asian	0.030 (0.003)	0.039 (0.003)	0.039 (0.003)	0.048 (0.003)	0.034 (0.003)	0.065 (0.003)
Female	0.043 (0.001)	0.054 (0.001)	0.054 (0.001)	0.044 (0.003)	0.038 (0.003)	0.054 (0.003)
Male	0.054 (0.001)	0.064 (0.001)	0.064 (0.001)	0.066 (0.003)	0.045 (0.003)	0.076 (0.003)
Democrat	0.056 (0.001)	0.069 (0.001)	0.069 (0.001)	0.046 (0.003)	0.041 (0.004)	0.057 (0.003)
non-Democrat	0.039 (0.001)	0.048 (0.001)	0.048 (0.001)	0.057 (0.003)	0.035 (0.004)	0.066 (0.003)
Under 35	0.034 (0.001)	0.066 (0.001)	0.061 (0.001)	0.037 (0.004)	0.044 (0.004)	0.082 (0.004)
35 to 65	0.026 (0.001)	0.027 (0.001)	0.026 (0.001)	0.046 (0.003)	0.027 (0.003)	0.047 (0.003)
Over 65	0.079 (0.001)	0.080 (0.001)	0.079 (0.001)	0.105 (0.003)	0.076 (0.003)	0.105 (0.003)
Covariates	✓	✓	✓	✓	✓	✓
Exact matching				✓	✓	✓
Imputing 2010 turnout	Logit	Lowest	Highest	Logit	Lowest	Highest

Table D.1: **Estimated Effects of VBM Adoption on Turnout**

Note: This table reports estimated effects of the VBM adoption on voter turnout with different specifications. Our main results are based on Column (4) and shown in bold.

D.2 Placebo Tests

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 being Democrat, female, and white 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.1 visualizes the results. It indicates that our placebo tests do not find any statistically and/or substantially significance effect on our false outcomes, suggesting that our DID estimation is a valid procedure.

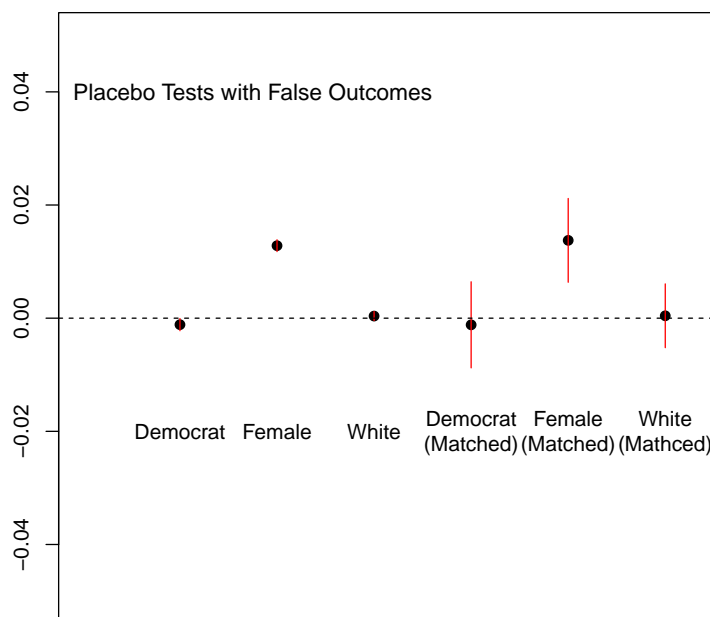


Figure D.1: **Placebo Tests with False Outcomes and Treatments**

Note: This figure presents the results of placebo tests with false outcomes and treatments, respectively.

D.3 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.2 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 a suggestive evidence that the parallel trends assumption is more or less satisfied, which grants a confidence in our causal identification. Nevertheless, this result can only be a 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.2 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 cautions 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).

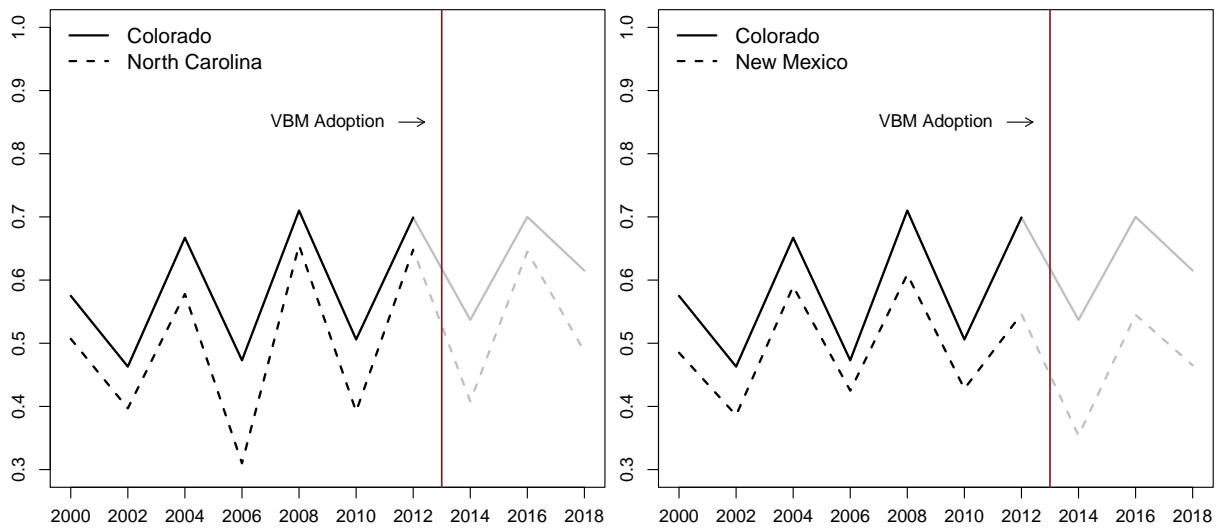


Figure D.2: Time Trends in Voter Turnout in Colorado, North Carolina, and New Mexico.

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: Evidence for the positive turnout effects of Vote at Home (also known as vote by mail) in participation counties."
- PantheonAnalysis. 2018. "Colorado 2014. Comparisons 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.