The Political Consequences of Vaccination

Emilio Depetris-Chauvin Felipe González*

This pre-analysis plan was published online on May 14, 2021
Date of election under study: May 15-16, 2021

Abstract. Vaccines are reponsible for the largest increases in human welfare and the role of the state is known to be critical. Yet we know surprisingly little about the impact of vaccination campaigns in the political sphere. This pre-analysis plan outlines an empirical strategy to estimate the impact of a countrywide vaccination process on electoral outcomes in an important election. The context of the study is Chile, which provides us with a rare combination of a high-stakes election, voluntary voting, and a vaccination process that is halfway implemented by election day. Crucially, the roll-out of vaccines has a clear rule which we exploit for causal identification.

Keywords: vaccines, politics, election.

*Pontificia Universidad Católica de Chile, Instituto de Economía. We thank Sebastian Figari and Mounu Prem for comments and suggestions. Leonor Castro provided outstanding research assistance.

1 Introduction

Vaccines are reponsible for the largest increases in human welfare in the last century. The health consequences of immunization are well documented and the role of the state in public health is known to be critical. Yet we know surprisingly little about the impact of vaccination campaigns in the political sphere. We study one of the worst health crisis in modern history, the coronavirus pandemic, which has caused millions of deaths, depressed both supply and demand in the economy, and actived ambitious economic policies across all continents. Moreover, the health crisis triggered an unprecedented competition for the development of a vaccine and a subsequent race across nations to secure stocks for their populations. This pre-analysis plan outlines two empirical strategies to estimate the impact of a countrywide vaccination process on electoral outcomes.

The context of the study is Chile, which provides an ideal testing ground for at least three reasons. First, the country secured a stock of vaccines and has deployed the immunization since December 2020. Elder people and workers in certain occupations have gotten the vaccine first on a week-to-week rolling program that started with people older than 90 years old and health personnel. Importantly, all the vaccination data has been made public in real time. Second, the country faces one of the most important elections in its recent history. Five months before the pandemic outbreak, an intense wave of protests triggered a referendum asking citizens if they would like to replace the current Constitution, originally drafted by the Pinochet dictatorship in 1980. The referendum was held in October 2020 and 80% voted for a new Constitution. As a consequence, a new Constitution will be drafted by a Constitutional Convention composed by 155 members elected by popular vote in the election we study. And third, automatic registration and voluntary voting characterize all elections since 2012. The combination of a high-stakes election, voluntary voting, and a massive vaccination process halfway implemented constitute an ideal empirical setting.

The election we study will take place in May 15-16 of 2021 together with three other electoral processes. In particular, voters will be given four different ballots. The most important election is the Constitutional Convention Election in which voters will elect those who will write the new Constitution. Local Elections are arguably the second most important and particularly relevant given the role that mayors play in the implementation of the vaccination process. Two ballots are tied to the Local Election, one to choose the mayor and another one to choose the members of the local council. All 345 counties in the country simultaneously elect one mayor and 6, 8, or 10 councilors depending on the county population. The fourth ballot corresponds to the Regional Governors Election, in which voters will elect one governor for each of the 16 regions of the country. In this document we refer to these four electoral process as the Election. Below we outline a plan to make use of all of these elections to estimate the impact of the vaccination process.

Previous research in economics has focused on the role of information and historical vaccination campaigns in driving contemporary vaccination rates (e.g. Martínez-Bravo and Stegmann 2021; Lowes and Montero 2021). In contrast, there is substantially less research on the political effects of large vaccination campaigns. A related research agenda has studied the political impact of large public health policies such as the Medicaid Expansion (Haselswerdt, 2017; Clinton and Sances, 2018; Baicker and Finkelstein, 2019). We will contribute to this literature by providing estimates of the causal impact of vaccination on electoral outcomes in a high-stakes election.

Finally, we contribute to a small literature that uses pre-analysis plans with observational data. This type of analysis is relatively scarce in economics, particularly when compared to the use of this methodology in randomized controlled trials.[1] As emphasized by Christensen and Miguel (2018), the pioneering study and one of the few to this date is Neumark (2001). Operationally, we follow the recommendations of Christensen and Miguel (2018) and Burlig (2018) to construct this pre-analysis plan. As noted in previous research, the study of electoral outcomes is particularly suited for this type of analysis. Elections have the advantage of taking place in a specific and verifiable date, so as long the pre-analysis is published before election day the method works. To ensure that we pre-specified our statistical model before the election takes place in Chile, this document was registered in the website of the Open Science Framework.

2 Empirical strategy based on the vaccination roll-out

We are interested in estimating the causal impact of vaccination on electoral outcomes, i.e. participation in the election and the corresponding political preferences for candidates, parties, and coalitions. We observe vaccination rates and electoral outcomes at the county level. Then we can write the relationship of interest as the following cross-sectional regression equation:

$$Y_c = \alpha + \beta V_c + \eta_c \tag{1}$$

where Y_c is an outcome of interest in county c, α is a constant term, V_c is the vaccination rate in county c, and η_c is a mean zero error term. The coefficient of interest is β. Unfortunately, a variety of different endogeneity problems prevent us from interpreting β as the causal relationship of vaccination on electoral outcomes. A leading concern is omitted variables W_c which can explain both the vaccination rates and electoral outcomes. One example is education, presumably associated with vaccination and electoral participation. However, there are potentially many omitted variables and even the bias in β in the naive regression (1) is difficult to bound or to put a sign on. In order to estimate the causal effect of vaccines, we need an identification strategy that exploits

[1] The use of pre-analysis plans in experimental studies has become more common and their use has increased rapidly in the past years. The number of registered studies in the AEA registry is an example of this trend.

exogenous variation in vaccination rates. We describe a strategy below.

2.1 Data sources

The empirical strategy described below uses the following four sources of data:

1. Individual-level data from the 2017 Census with the county of residence and age, gender, occupation, labor force participation, and unemployment status.

2. Administrative electoral data from the Electoral Service of Chile including county-level participation and vote preferences from 2012 until now.

3. Administrative data from the Ministry of Health with county-level information on the number of people who have been vaccinated up to a given date, the number of deaths and infections related to the pandemic, and the location of vaccination centers across the country.

4. Data from a nationally representative survey of approximately 270,000 individuals in 324 counties across de country in 2017 known as CASEN survey.

2.2 Plausibly exogenous variation from vaccination rules

We employ a two-stage least squares with multiple instruments for identification of Local Average Treatment Effects. The instruments $Z_c = \{z_{1c}, \ldots, z_{Jc}\}$ are vaccination priority groups defined by the central government. The scarcity of resources implies that a country-wide vaccination process such as the COVID campaign takes months (or even years) to reach a large fraction of the population. Then, the Government of Chile organized the vaccination process in such a way that people in priority groups were the first to get the vaccine. Moreover, the vaccination plan was released shortly before the first vaccines arrived to the country and consists of two parts. The first part states that older people and those with a chronic condition get a vaccine first. By the time of the election in May 15-16, all Chileans and foreign residents of 35 years old or older had the opportunity to get a vaccine. The second part of the plan states that workers in certain "critical" occupations also get the vaccine first. Examples of these occupations are those in the health sector, energy, gas, and water supply, public transportation, education, and public service, among others.

The existence of clear priority groups in the vaccination plan allows us to construct an instrument Z_c, defined as the share of the county population that was offered a vaccine before the election. Operationally, we use the individual-level data from the 2017 Census and take the union of the two priority groups to construct Z_c. The age of an individual is straightforward to collect. We identified individuals with a chronic condition using data from the Ministry of Health, which targets this population during the annual vaccination campaign related to the influenza disease. In terms of occupations, we are restricted by the categories in the census and we use the following: health personnel, public transportation, education, and public workers. Note that the existence of multiple priority groups allows us to construct many different instruments $z_{jc} \in Z_c$. For example, z_{1c} might denote the share of people in county c that are older than 35 years old, z_{2c} the share of people that have a chronic condition, and z_{3c} the share working as health personnel. Below we propose to exploit these sub-instruments to estimate multiple Local Average Treatment Effects.

2.3 Estimating equations

We build upon equation (1) and relate local electoral outcomes to the local COVID vaccination rates until the day before the election using the following parsimonious regression equation:

$$Y_{cp} = \beta V_c^k + \gamma X_c + \phi_p + \epsilon_{cp} \quad (2)$$

where Y_{cp} is an electoral outcome in county c, located in province p. Chile is divided in 346 counties and each county is located in one of 56 provinces. We use 343 counties in 54 provinces because one county lacks complete political data (Antarctica) and two counties are also a province which means their variation is absorbed by ϕ_p. The right-hand side variable of interest is the vaccination rate V_c^k which we defined as the number of people with $k = 1$ or $k = 2$ doses over the total number of people older than 18 years old (i.e. adult population) in the county as measured by the 2020 projections of the National Statistics Institute (INE). We also include a set of predetermined covariates X_c to improve the precision of estimates and control for county characteristics that correlate with the instrument. We use a mean zero error term ϵ_{cp} that we allow to be robust to heteroskedasticity. Finally, given that electoral outcomes arise from individual-level decisions, we estimate equation (2) using weighted least squares with the adult population in the county as weight.

We use two versions of the instrument, one for the exposure to one dose ($Z_{1,c}$) and another for the exposure to two doses ($Z_{2,c}$). To be clear, there are two versions of equation (2), each one with its own instrument. Assuming that the instrument is valid, then we can also estimate the causal effect of being offered the vaccine on electoral outcomes, i.e. the reduced form:

$$Y_{cp} = \delta Z_{k,c} + \gamma X_c + \phi_p + \epsilon_{cp} \quad (3)$$

where all variables are defined as before and δ represents the causal impact of the share of people who was offered the vaccine on electoral outcomes. In the randomized controlled trial literature the parameter δ could be be interpreted as the Intention to Treat (ITT) with the vaccine.

Throughout the analysis we will use the following five specifications to learn about the robustness of results and the role of omitted variables using a coefficient stability approach:

1. Without province fixed effects ϕ_p and without controls X_c.

2. Including province fixed effects ϕ_p and without controls X_c.

3. Including ϕ_p and the following basic controls $x_{1,c} \in X_c$: the log of the distance (in km.) from the county to the national capital, the log of the distance (in km.) from the county to the regional capital, one indicator for counties with less than 50,000 inhabitants, and one indicator for counties hosting between 50,000 and 100,000 people. These controls aim to capture basic predetermined differences in the geographic location and size of counties.

4. Including ϕ_p, $x_{1,c}$, and the following extended controls $x_{2,c} \in X_c$ which below we find to be correlated with the instrument: turnout in the 2017 presidential election, labor participation rate, share of women in population, labor participation and unemployment rate of women, prevalence of permanent health conditions, average household subsidy (in logs), total COVID deaths per 10,000 inhabitants (in logs), and number of vaccination centers per 100,000 inhabitants.

5. Including ϕ_p, $x_{1,c}$, $x_{2,c}$ and the following controls from the 2020 plebiscite $x_{3,c} \in X_c$: turnout and vote share for the approval option. These controls aim to capture predetermined political differences across counties in a recent election also held during the pandemic.

For specifications 1-3 we observe 343 counties. However, we only observe 323 counties when we use specifications 4-5 because two covariates come from the 2017 National Survey.

2.4 Validity of the research design

The validity of the instruments rests on the condition that it has sufficiently strong predictive power of the endogenous variable and on the assumption that it affects the outcomes of interest only through the endogenous variable (i.e. exclusion restriction) after we condition on a small set of predetermined covariates (i.e. conditional exogeneity). Below we will show that the instrument, i.e. the share of people in priority groups, has a strong predictive power of the percentage of people vaccinated before the Election. Regarding the exclusion restriction, we provide suggestive evidence supporting this identification assumption using the correlation between the instrument a wide range of variables covering the political and economic dimensions of counties.

2.4.1 Correlation between the instrument and predetermined covariates

Table 1 presents summary statistics for 17 variables describing local political participation and preferences, and the predicted power of the instrument ($Z_{2,c}$) on these variables. We have organized this table to study political participation (panel A), and preferences (panel B) in the 2020 plebiscite, and for left-wing, right-wing, and independent candidates in all elections since 2012 when automatic registration and voluntary voting was introduced. To classify candidates as left-wing and right-wing, we follow previous work using data from these elections (Bautista et al., 2021). Table 2 examines 14 additional variables from the 2017 Census. Table 3 studies other 10 variables but now from the 2017 National Survey. Finally, Table 4 presents four variables related to the COVID pandemic. In sum, we estimated the correlation between the instrument and 46 variables covering elections, the labor market, health conditions, state subsidies, and the pandemic, and we observe 8 statistically significant differences at the 10% level. The number of differences is slightly above the 5 derived from a 10% statistical test ($0.10 \times 46 = 4.6$), which in this case was reasonable to expect given that the instrument should correlate with characteristics of the elder population as we explain below. Importantly, only one of the 17 political variables is correlated with the instrument at the 5% level, which is what we expected of a 5% statistical test ($0.05 \times 17 = 0.85$).

Overall, we interpret Tables 1-4 as supporting the validity of the research design in the sense that the instrument has little predictive power of political participation or political preferences at the local level as measured by the five elections held between 2012 and 2020. Moreover, the signs of coefficients do *not* suggest systematic political differences across counties with varying exposure to the vaccination process. For example, the standardized correlation between the instrument and the vote share of left-wing candidates in local elections changes from 0.29 in 2012 to -0.07 in 2016, and a similar picture emerges in the case of right-wing or independent candidates.

The few differences in Tables 1-4 confirm that the vaccination process prioritized the elder population. As women tend to live longer, it was expected to observe a higher population of women in counties with more priority groups. Similarly, as older people are less likely to work, we also expected lower participation rates in the labor force in places more exposed to the vaccines, and more people with permanent health conditions and who receive more state subsidies. In other words, the instrument is expected to correlate with variables that characterize the elder population, including COVID deaths and the number of vaccination centers. More critical is the lack of a correlation with predetermined political preferences or with economic conditions and educational levels, all which are likely to affect political outcomes. In that sense, it is reassuring that the instrument is uncorrelated with household per capita income, poverty rates, rural population, different education measures, malnutrition, lack of health insurance, and lack of basic services. It is also reassuring that the instrument is *not* associated with the number of COVID infections and the prevalence of lockdowns, which arguably proxy for the negative economic impacts of the pandemic. Regardless, the fourth specification controls for the variables correlated with the instrument.

2.4.2 Preliminary first-stage results

Table 5 presents the first-stage. These results are only preliminary because we only observe vaccination data until May 7, but the analysis will use data until May 14. As mentioned in subsection 2.3, we always present results from five different specifications and one of two endogenous variables, i.e. share of adults with one or two doses.[2] Column 1 in panel A shows the partial correlation between the instrument and the share of adults with two doses, and panel B the share with one dose. Four patterns emerge from this table. First, the share of people in priority groups is a strong predictor of the share of adults one or two doses. The F-statistics are always larger than 49, alleviating concerns about potential weak instruments (Stock and Yogo, 2005). Second, the first-stage coefficient is remarkably stable across the five different specifications and hovers between 0.66 and 0.81. Moreover, the differences between columns 3 and 4 suggest that the correlations between the instrument and predetermined covariates are unlikely to be an empirical concern as, if anything, the correlation becomes stronger when including these covariates as controls. Third, the first-stage coefficient is lower than one, which means that there is imperfect compliance with the vaccination process, i.e. approximately 70-80% of the people who were eligible to get vaccinated decided to take the vaccine. And fourth, the covariates related to the only election held during the pandemic (2020 plebiscite) have predictive power of vaccination rates as measured by the R-squared.

The first-stage estimation results in Table 5 pushed us to make the following three empirical decisions. First, we will use the fifth specification when estimating the impact of the vaccination process on electoral outcomes. The reason behind this decision is the explanatory power of the covariates related to the 2020 plebiscite, which will increase the precision of our estimates, and the statistically significant correlations between the instrument and predetermined covariates. Second, given the similarity across panels, we will focus on specifications in which the endogenous variable is the share of adults with two doses. Third, we will add as control one lag of the corresponding dependent variable to improve the precision of our estimates (details below).

2.5 Outcomes

We will examine the impact of the vaccination process on two sets of outcomes Y_c. The first is *Turnout*, defined as total votes in election ℓ (including null and blank votes) over total number of people who are eligible to vote (i.e. *electores*), with ℓ being Local Elections (mayor), Local Elections (councilors), Constitutional Convention, and Governors. The second set of outcomes are *Vote Shares*, defined as votes for option j in the election over total number of votes, with j being:

[2] The two instruments are measured as the share of people in critical occupations plus the share of adults older than 40 ($Z_{1,c}$) or 50 ($Z_{2,c}$) years old. Ages com from the vaccination calendar and will be updated accordingly.

1. **Local Election**

   1.1 *Incumbent*, defined as the incumbent mayor running for reelection or the candidate from his/her coalition when the mayor is not running.

   1.2 *Left-wing*, defined as those running in the following coalitions: Unidad por el Apruebo, Chile Digno Verde y Soberano, Unidos por la Dignidad, Dignidad Ahora,

   1.3 *Right-wing*, defined as those running in the following coalitions: Chile Vamos, Republicanos, Independientes Cristianos, Ciudadanos Independientes, Nuevo Tiempo.

   1.4 *Independent*, defined as those running in the following coalitions: Ecologistas e Independientes, Independientes fuera de pacto.

   1.5 *Councilors*, same outcomes as the previous four but defined in the separate local election for councilors.

2. **Constitutional Convention Election**

   2.1 *Left-wing*, defined as candidates running in the following lists: Lista del Apruebo (YB), Apruebo Dignidad (YQ), Partido Humanista (XG), Partido Ecologista (XA).

   2.2 *Right-wing*, defined as candidates running in the list Vamos por Chile (XP).

   2.3 *Independent*, defined as candidates in any of the 74 lists (A-ZZ) that are different from the five lists composed by candidates from left- or right-wing political parties.

   2.4 *Invalid*, defined as null or blank votes over the total number of casted votes. This measure attempts to capture the level of confusion or disinformation in the population. Recent media articles suggest that some people appear to believe that they have to vote for multiple candidates. The confusion is understandable given that this is the first time a Constitutional Convention will be elected and there are reserved seats for women and indigenous people.

3. **Regional Governors Election**

   3.1 *Left-wing*, defined as candidates in the following coalitions: Unidad Constituyente, Frente Amplio, Igualdad para Chile, Humanicemos Chile, Partido de Trabajadores Revolucionarios, Por Dignidad Regional,

   3.2 *Right-wing*, defined as candidates in the following coalitions: Chile Vamos, Partido Republicano, Unión Patriótica, Partido Nacional Ciudadano, Independientes Cristianos,

   3.3 *Independents*, defined as candidates in the following coalitions: Ecologistas e Independientes, Regionalistas Verdes, Independientes fuera de pacto.

2.6 Additional empirical exercises

We will implement the following additional empirical exercises:

- We will estimate heterogeneous effects in Local Elections (mayors) using an indicator for counties where the incumbent mayor was banned from reelection. A new law enacted shortly before the pandemic outbreak establishes that incumbent mayors can only go for reelection for a maximum of two periods, for a total of three periods in power (12 years).
- We will estimate heterogenous treatment effects using all posible combinations of z_{jc} to trace out variation in the Local Average Treatment Effect. The implementation of this exercise rests on the possibility that different instruments have predictive power of the endogenous variables. Thus, we will only use subsets of instruments with a sufficiently strong first-stage.
- We will calculate standard errors that are corrected for spatial correlation (Conley, 1999).
- We will calculate additional p-values based on randomization inference.
- We will use the method proposed by Abadie et al. (2002) to characterize the population of compliers and trace out the importance of these characteristics to explain the LATE.

3 Identification strategy based on vaccination centers

3.1 Data sources

This empirical analysis will use following sources of data:

1. List of all people in the country that have the right to vote in the 2021 Election. These data is known as Electoral Registry, it is constructed by the Electoral Service and for each person we observe their age, gender, home address, and the booth in which they can vote.
2. Electoral outcomes at the booth-level. Booths are groups of maximum 350 people and there are approximately 45,000 in the country located inside thousands of polling stations.
3. The location of all vaccination centers. There were approximately 1,400 in January 2021 and they have increased to 1,900 before the Election. In Tables 4 and 5 we use the 1,400 centers but we will update this number to the closest one before the Election.

We will geocode the entire Electoral Registry, i.e. the home addresses of the 15 million people in Chile who can vote, the location of all booths, and the location of all vaccination centers.

3.2 Plausibly exogenous variation

We propose to exploit the location of vaccination centers as a source of within-county exposure to the vaccination campaign. The intuition is simple, people who happen to live farther away from vaccination centers face a larger cost of getting vaccinated and therefore lower vaccination rates. Because the location of centers was presumably unknown ex-ante, the distance from people's homes to these places should be a valid source of variation. Moreover, people are assigned randomly to booths within their county of residence based on their national ID number and the explicit goal of reaching 350 voters per booth. Therefore, the average distance from people in a booth to the closest vaccination venue should vary quasi-randomly across booths. Below we propose to test the exogeneity of this distance by looking at its relationship with previous voting patterns.

3.3 Estimating equation

We will estimate the following cross-sectional equation:

$$y_{ij} = \tau d_i + \gamma x_i + \phi_j + \varepsilon_{ij} \tag{4}$$

where d_i is a vector of distances from people's homes in booth i to specific locations, and y_{ij} is an electoral outcome in booth i located in county j. We will use the same outcomes as those described in section 2.5 but now measured in 45,000 booths instead of 343 counties. In contrast to the previous strategy, our interest is now on the average distance from people's homes in a booth to the closest vaccination venue within their county of residence. As geographic controls, we will also include the distance from people's homes to the booth and the county hall for a total of three distance variables. Equation (4) also includes a vector for the characteristics of people in a booth, x_i. This vector includes the percentage of women, the average age, and the total number of people registered in the booth. In order to make comparisons within counties we include a full vector of county-level fixed effects ϕ_j and we allow the error term ε_{ij} to be correlated within counties.

3.4 Validity of the research design

We will provide evidence to support this research design using the same analysis as in Table 1. In this case the exogenous variable is the distance to the closest vaccination venue, we replace the province fixed effects by county fixed effects, and the controls by the distance from people's homes to the booth and the county hall, the percentage of women, the average age of people, and the total number of people in the booth. The variables to be included as covariates to examine are again turnout and vote shares in the 2020, 2017, 2016, 2013, and 2012 elections. Note that in this case we expect to miss some booths because as population increases new booths are opened.

3.5 Additional empirical exercises

We will implement the following additional empirical exercises:

- We will check for the robustness of results when replacing the distance to the nearest vaccination center by (i) the average distance to all vaccination centers in the county, and (ii) the distance to the nearest vaccination center in the province of residence.

4 Presentation of results

Tables to be presented in the paper:

- **Table 1**: Descriptive statistics and validity of research design (Table 1 in this document).
- **Table 2**: Descriptive statistics and validity of research design for the booth-level design. This table mimics Table 1 in this document but using 45,000 booths following section 3.4.
- **Table 3**: Priority groups and vaccination process (first-stage, Table 5 in this document). The results in this document were estimated using vaccination data until May 7 of 2021. We will update results until May 14 of 2021, the day before the Election under study.
- **Table 4**: The impact of the vaccination process on political participation. Four columns using turnout as outcome in the four elections: (1) Constitutional Convention, (2) Local Election, (3) Councilors, and (4) Governors. Panel A shows reduced form results, panel B the 2SLS results, and panel C the OLS results for comparison. The additional control for the lag of the dependent variable is turnout in the 2016 local election in all four columns.[3] Panel D presents the booth-level results using the same four outcomes and one specification.
- **Table 5**: Partisanship in Local Election for mayors. Four columns using vote shares for the following groups as dependent variable: (1) Incumbents, (2) Left-wing candidates, (3) Right-wing candidates, (4) Independent candidates. Panel A shows reduced form results, panel B the 2SLS results, and panel C the OLS results for comparison. The additional control for the lag of the dependent variable are vote shares in the 2016 local election in all four columns. Panel D presents the booth-level results using the same four outcomes and one specification.

[3]The Constitutional Convention election and the Regional Governors election are held for the first time.

- **Table 6**: Partisanship in Constitutional Convention. Four columns using vote shares for the following groups as dependent variable: (1) Left-wing candidates, (2) Right-wing candidates, (3) Independent candidates. Panel A shows reduced form results, panel B the 2SLS results, and panel C the OLS results for comparison. Panel D presents the booth-level results using the same three outcomes and one specification.

Tables to be presented in the Online Appendix:

- **Table A.1**: Descriptive statistics from the 2017 census (Table 2 in this document).
- **Table A.2**: Descriptive statistics from the 2017 national survey (Table 3 in this document).
- **Table A.3**: Descriptive statistics from the COVID pandemic (Table 4 in this document).
- **Table A.4**: Partisanship in Local Election for councilors. Four columns using vote shares for the following groups as dependent variable: (1) Incumbents, (2) Left-wing candidates, (3) Right-wing candidates, (4) Independent candidates. The additional control for the lag of the dependent variable is turnout in the 2016 local election in all four columns. Panel A shows reduced form results, panel B the 2SLS results, and panel C the OLS results for comparison. Panel D presents the booth-level results using the same four outcomes and one specification.
- **Table A.5**: Partisanship in Regional Governors Election. Three columns using vote shares for the following groups as dependent variable: (1) Left-wing candidates, (2) Right-wing candidates, (3) Independent candidates. Panel A shows reduced form results, panel B the 2SLS results, and panel C the OLS results for comparison. Panel D presents the booth-level results using the same four outcomes and one specification.
- **Table A.6**: Characterization of the complier population.

References

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.

Baicker, K. and Finkelstein, A. (2019). The Impact of Medicaid Expansion on Voter Participation: Evidence from the Oregon Health Insurance Experiment. *Quarterly Journal of Political Science*, 14(4):383–400.

Bautista, M. A., González, F., Martínez, L., Muñoz, P., and Prem, M. (2021). The geography of repression and opposition to autocracy. *American Journal of Political Science*.

Burlig, F. (2018). Improving transparency in observational social science research: A pre-analysis plan approach. *Economics Letters*, 168:56–60.

Christensen, G. and Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3):920–80.

Clinton, J. D. and Sances, M. W. (2018). The politics of policy: The initial mass political effects of medicaid expansion in the states. *American Political Science Review*, 112(1):167?185.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 91(1):1–45.

Haselswerdt, J. (2017). Expanding medicaid, expanding the electorate: The affordable care act's short-term impact on political participation. *Journal of health politics, policy and law*, 42(4):667–695.

Lowes, S. and Montero, E. (2021). The legacy of colonial medicine in Central Africa. *American Economic Review*, 111(4):1284–1314.

Martínez-Bravo, M. and Stegmann, A. (2021). In vaccines we trust? the effects of the cia's vaccine ruse on immunization in Pakistan. *Journal of the European Economic Association*.

Neumark, D. (2001). The employment effects of minimum wages: Evidence from a prespecified research design the employment effects of minimumwages. *Industrial Relations: A Journal of Economy and Society*, 40(1):121–144.

Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter 5, pages 80–108.

Table 1: Descriptive statistics and validity of the instrument

		Univariate regression of covariate on instrument (mean instrument 64.3, st. dev. 9.27)			
	Mean st. dev.	unconditional	conditional on province F.E.	conditional on province F.E. and controls	Standardized effect from (4)
Panel A: Political participation	(1)	(2)	(3)	(4)	(5)
Turnout 2020 Plebiscite	43.9	-0.267*	0.178	0.208	0.18
	10.4	(0.140)	(0.214)	(0.172)	
Turnout 2017 Presidential Election	46.1	-0.2	-0.401**	-0.410***	0.35
	10.9	(0.161)	(0.173)	(0.154)	
Turnout 2016 Local Election	47.3	0.641***	0.333***	0.034	0.03
	12.2	(0.090)	(0.100)	(0.098)	
Turnout 2013 Presidential Election	49.1	0.076	-0.202	-0.229	-0.20
	10.5	(0.172)	(0.181)	(0.145)	
Turnout 2012 Local Election	53.6	0.562***	0.298***	0.059	0.05
	10.8	(0.100)	(0.079)	(0.079)	
Panel B: Political preferences					
Supports new constitution 2020	75.7	-0.19*	0.074	0.177	0.17
	9.9	(0.101)	(0.141)	(0.170)	
Supports convention 2020	71.8	-0.199**	0.045	0.163	0.18
	8.4	(0.091)	(0.128)	(0.151)	
Vote share right-wing 2017	46.7	0.088	-0.074	-0.212	-0.23
	8.6	(0.110)	(0.134)	(0.153)	
Vote share right-wing 2016	36.7	-0.299	-0.082	-0.005	0.00
	19.7	(0.268)	(0.293)	(0.352)	
Vote share right-wing 2013	23.7	-0.124	-0.08	-0.156	-0.21
	7.0	(0.085)	(0.124)	(0.150)	
Vote share right-wing 2012	35.6	-0.122	-0.25	-0.137	-0.07
	18.1	(0.265)	(0.264)	(0.334)	
Vote share left-wing 2017	53.3	-0.088	0.074	0.212	0.23
	8.6	(0.111)	(0.134)	(0.153)	
Vote share left-wing 2016	41.8	0.183	-0.054	-0.147	-0.07
	18.5	(0.220)	(0.279)	(0.337)	
Vote share left-wing 2013	64.7	0.135*	0.122	0.15	0.20
	7.0	(0.080)	(0.110)	(0.132)	
Vote share left-wing 2012	44.7	0.22	0.535*	0.558	0.29
	17.7	(0.189)	(0.316)	(0.356)	
Vote Share Independent 2016	17.9	0.158	0.074	0.052	0.02
	22.8	(0.329)	(0.435)	(0.516)	
Vote Share Independent 2012	16.0	-0.014	-0.254	-0.411	-0.18
	20.9	(0.321)	(0.443)	(0.523)	
Counties	343				

Notes: Column 1 reports the mean and standard deviation for 17 variables from previous elections (listed at the left). Columns 2 to 4 report point estimates and robust standard errors from OLS regressions of each covariate on the instrument (i.e., share of people in priority groups). Column 2 shows unconditional results, column 3 conditions on 54 province fixed effects, and column 4 conditions on province fixed effects and a restricted set of controls including distance to the national capital (in logs), distance to the regional capital (in logs) and two indicators of population size (i.e., less than 50 thousand inhabitants and between 50 thousands and 100 thousands inhabitants). All regressions are weighted by county adult population in 2020. Statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Table 2: Descriptive statistics from the 2017 Census

	Mean st. dev.	Univariate regression of covariate on instrument (mean instrument 64.3, st. dev. 9.27) unconditional	conditional on province F.E.	conditional on province F.E. and controls	Standardized effect from (4)
	(1)	(2)	(3)	(4)	(5)
Population women	49.0	0.037*	0.041	0.060*	0.10
	5.6	(0.021)	(0.042)	(0.033)	
Population 0 to 4 yrs old	6.4	-0.037	-0.027	-0.036	-0.30
	1.1	(0.024)	(0.031)	(0.024)	
Population 5 to 12yrs old	10.8	-0.006	0.008	-0.027	-0.14
	1.7	(0.055)	(0.076)	(0.057)	
Population 12 to 18 yrs old	9.3	0.021	0.032	0.011	0.06
	1.7	(0.046)	(0.068)	(0.052)	
Labor Participation Rate	59.8	-0.582***	-0.434***	-0.400***	-0.38
	9.7	(0.056)	(0.059)	(0.058)	
Labor Participation Rate, women	47.0	-0.698***	-0.540***	-0.448***	-0.40
	10.3	(0.093)	(0.109)	(0.097)	
Unemployment Rate	7.0	0.030*	0.022	0.031	0.13
	2.3	(0.016)	(0.018)	(0.020)	
Unemployment Rate, women	11.5	0.112***	0.091**	0.070*	0.15
	4.3	(0.035)	(0.035)	(0.039)	
Poor Household Rate (extensive)	6.4	-0.067**	-0.044	-0.037	-0.12
	2.9	(0.032)	(0.050)	(0.049)	
Poor Household Rate (intensive)	1.4	-0.013	-0.009	-0.007	-0.09
	0.7	(0.008)	(0.011)	(0.010)	
Rural Population	0.4	0.009***	0.005***	0.001	0.03
	0.3	(0.001)	(0.002)	(0.001)	
Population with Primary Education	0.3	0.004***	0.003**	0.001	0.10
	0.1	(0.001)	(0.001)	(0.001)	
Population with Secondary Education	0.4	0.001	0.001	0.001	0.17
	0.1	(0.001)	(0.001)	(0.001)	
Population with Terciary Education	0.2	-0.005***	-0.005	-0.003	-0.31
	0.1	(0.002)	(0.003)	(0.003)	
Counties	343				

Notes: Column 1 reports the mean value and standard deviation for 14 demographic and labor market variables from 2017 Census (listed at the left). Columns 2 to 4 report point estimates and robust standard errors from OLS regressions of each covariate on our instrument (i.e., share of people in priority groups). Column 2 shows unconditional results, column 3 conditions on 54 province fixed effects, and column 4 conditions on province fixed effects and a restricted set of controls including distance to the national capital (in logs), distance to the regional capital (in logs) and two indicators of population size (i.e., less than 50 thousand inhabitants and between 50 thousands and 100 thousands inhabitants). All regressions are weighted by county adult population in 2020. Statistical significance: *** p <0.01, ** p <0.05, * p <0.1.

Table 3: Descriptive statistics from the 2017 National Survey

		Univariate regression of covariate on instrument (mean instrument 64.3, st. dev. 9.27)			
	Mean st. dev.	unconditional	conditional on province F.E.	conditional on province F.E. and controls	Standardized effect from (4)
	(1)	(2)	(3)	(4)	(5)
Log household income	12.5	-0.016***	-0.009	-0.006	-0.18
	0.3	(0.004)	(0.006)	(0.007)	
Poverty Rate	12.4	0.228***	-0.018	-0.038	-0.05
	7.3	(0.040)	(0.050)	(0.058)	
Poverty Rate, multidimensional	26.1	0.095	0.156	0.031	0.03
	10.5	(0.095)	(0.120)	(0.124)	
Self-reported health score	18.1	0.135***	0.062**	0.053	0.15
	3.2	(0.031)	(0.031)	(0.038)	
Permanent health condition	12.7	0.189***	0.098**	0.101**	0.20
	4.6	(0.034)	(0.039)	(0.040)	
Malnutrition	7.4	0.052	0.046	0.018	0.04
	3.9	(0.042)	(0.060)	(0.057)	
Lack of health insurance	5.3	-0.166***	-0.082	-0.091	-0.20
	4.3	(0.041)	(0.067)	(0.075)	
Lack of social security	36.4	0.079	0.281**	0.204	0.17
	11.5	(0.124)	(0.137)	(0.145)	
Lack of basic services	14.3	0.313***	0.138*	0.008	0.01
	12.6	(0.062)	(0.075)	(0.053)	
Log household subsidy	9.5	0.034***	0.021***	0.017***	0.37
	0.4	(0.004)	(0.005)	(0.005)	
Counties	323				

Notes: Column 1 reports the mean value and standard deviation for 12 demographic and labor market variables from 2017 Census (listed at the left). Columns 2 to 4 report point estimates and robust standard errors from OLS regressions of each covariate on our instrument (i.e., share of people in priority groups). Column 2 shows unconditional results, column 3 conditions on 54 province fixed effects, and column 4 conditions on province fixed effects and a restricted set of controls including distance to the national capital (in logs), distance to the regional capital (in logs) and two indicators of population size (i.e., less than 50 thousand inhabitants and between 50 thousands and 100 thousands inhabitants). All regressions are weighted by county adult population in 2020. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Descriptive statistics for the pandemic before the vaccines

		Univariate regression of covariate on instrument (mean instrument 64.3, st. dev. 9.27)			
	Mean st. dev.	unconditional	conditional on province F.E.	conditional on province F.E. and controls	Standardized effect from (4)
	(1)	(2)	(3)	(4)	(5)
Share of lockdown days	7.0	-0.310**	-0.137	0.002	0.00
	9.7	(0.151)	(0.113)	(0.104)	
COVID infections per 10,000	277.7	-4.595**	1.042	1.701	0.10
	159.7	(1.931)	(1.788)	(1.842)	
COVID deaths per 10,000	5.8	-0.161**	0.256**	0.278**	0.50
	5.2	(0.076)	(0.112)	(0.111)	
Vaccination centers per 100,000	24.3	0.540***	0.445***	0.351***	0.07
	48.4	(0.080)	(0.139)	(0.103)	
Counties	343				

Notes: Column 1 reports the mean value and standard deviation for 4 variables related to the pandemic (listed at the left). All covid figures are measured until first day of the vaccination campaign (December 23, 2020). Columns 2 to 4 report point estimates and robust standard errors from OLS regressions of each covariate on our instrument (i.e., share of people in priority groups). Column 2 shows unconditional results, column 3 conditions on 54 province fixed effects, and column 4 conditions on province fixed effects and a restricted set of controls including distance to the national capital (in logs), distance to the regional capital (in logs) and two indicators of population size (i.e., less than 50 thousand inhabitants and between 50 thousands and 100 thousands inhabitants). All regressions are weighted by county adult population. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Priority groups and vaccination process (first-stage)

	Dependent variable: Share of adults with two doses				
Panel A	(1)	(2)	(3)	(4)	(5)
Share of people in priority groups	0.723***	0.706***	0.689***	0.750***	0.663***
	[0.052]	[0.086]	[0.098]	[0.095]	[0.089]
R-squared	0.405	0.524	0.539	0.757	0.778
Mean of dependent variable	47.61	47.61	47.61	46.34	46.34
Mean of instrument	64.17	64.17	64.17	63.52	63.52
Panel B	Dependent variable: Share of adults with one dose				
Share of people in priority groups	0.746***	0.718***	0.710***	0.809***	0.672***
	[0.059]	[0.095]	[0.111]	[0.113]	[0.106]
R-squared	0.370	0.477	0.498	0.724	0.741
Mean of dependent variable	56.75	56.75	56.75	55.46	55.46
Mean of instrument	81.6	81.6	81.6	80.90	80.90
Province fixed effects	N	Y	Y	Y	Y
Basic controls	N	N	Y	Y	Y
Unbalanced covariates	N	N	N	Y	Y
2020 Plebiscite controls	N	N	N	N	Y
Observations	343	343	343	323	323

Notes: The share of target population is computed as the sum of population working in health services, transportation, education, and public administration, population with chronic diseases, and population older than 50 (40) years old for two (one) dose(s); all as shares of adult population. The basic set of controls includes distance to national capital (in logs), distance to regional capital (in logs) and two indicators of population size (i.e., less than 50 thousand inhabitants and between 50 thousands and 100 thousands inhabitants). The set of unbalanced covariates includes turnout in 2017 presidential election, labor participation rate, share of women in population, labor participation rate of women, unemployment rate of women, prevalence of permanent health conditions, average household subsidy (in logs), total covid deaths per 10,000 inhabitants (in logs), and number of vaccination centers per 100,000 inhabitants. 2020 Plebiscite controls include turnout and vote share for approval. Regressions are weighted by voting age population. Robust standard errors in parenthesis. Statistical significance: *** p <0.01, ** p <0.05, * p <0.1.