OREGON HEALTH INSURANCE EXPERIMENT ANALYSIS PLAN: EVIDENCE FROM VOTING DATA

Introduction

In 2008, Oregon held a lottery to allocate a limited number of Medicaid slots to low-income uninsured adults who had signed up for a waiting list. Using this lottery (the Oregon Health Insurance Experiment) and data received from the Elections Division of the Oregon Secretary of State, we plan to estimate the effects of this Medicaid expansion on voting registration and voting behavior.

This document pre-specifies the planned analysis before comparing outcomes for treatment and control groups. Creating this record of our ex ante planned analysis helps to minimize issues of data mining and specification searching. This plan was completed prior to analysis of differences in outcomes; however, as detailed below, we examine the distribution of the outcomes in the control group to make specification decisions, run our first stage analyses, and perform treatment-control balance tests to check the validity of our empirical strategy.

This plan was constructed after completion of analyses using the lottery to estimate the effects of insurance using information from different data sets: from a mail survey and administrative data collected approximately one year after the lottery (Amy Finkelstein et al., 2012), in-person interview data collected approximately two years after the lottery (Katherine Baicker et al., 2013), social security administrative data (Katherine Baicker et al., 2014a), administrative emergency department data collected approximately eighteen months later (Sarah L. Taubman et al., 2014), and criminal charges data covering a period approximately two years after the lottery (Katherine Baicker et al., 2014b). The methods proposed here follow those of our prior analyses very closely; however, the outcome measures are new.

Background

This analysis will investigate effects of health insurance on voter registration, registered party affiliation, and voting. There are a number of channels by which the lottery (or receipt of Medicaid) might affect voter participation. First, winning (or losing) the lottery could affect (with either sign) one's desire to engage in the political process or to affiliate with a specific party. Second, voter participation may increase if beneficiaries of social programs are more likely to vote than non-beneficiaries. Medicaid receipt may also affect party affiliation; for example, it has been conjectured that the Medicaid expansion under the Affordable Care Act/Obamacare will create a new cohort of devoted Democrats.

Much of the recent empirical work examining how voters respond to government spending on social programs explores settings outside of the United States. Several studies have found that conditional cash transfer programs increase both voter turnout and share of votes to the incumbent political party ((Ana De La O, 2013), (Marco Manacorda et al., 2011), (Julien Labonne, 2013), (Cesar Zucco, 2013)). Targeted public spending has also been shown to increase support among program beneficiaries for the incumbent government party (Christian Pop-Eleches and Grigore Pop-Eleches, 2012).

Methods

Randomization and Intervention

Oregon opened a waiting list for a previously closed Medicaid program in early 2008 and then conducted eight lottery drawings from the waiting list between March and September 2008. Selected individuals won the opportunity – for themselves and any household member – to apply for health insurance benefits through Oregon Health Plan Standard (OHP Standard). OHP Standard provides benefits to low-income adults who are not categorically eligible for Oregon's traditional Medicaid program. To be eligible, individuals must be: ages 19-64; not otherwise eligible for Medicaid or other public insurance; Oregon residents; U.S. citizens or legal immigrants; without health insurance for six months; with income below the federal poverty level and assets below $2,000. Among the randomly selected individuals, those who completed the application process and met these eligibility criteria were enrolled in OHP Standard. OHP Standard provides relatively comprehensive medical benefits (including prescription drug coverage) with no consumer cost sharing and low monthly premiums (between $0 and $20, based on income), provided mostly through managed care organizations. The lottery process and OHP Standard have been described in more detail elsewhere (Amy Finkelstein et al., 2012).

Data Sources

The state provided us with the initial lottery list and with detailed data on Medicaid enrollment for every individual on the list. We use this to construct our primary measure of insurance coverage during the study period. These are analyzed and described in detail in (Amy Finkelstein et al., 2012).

The statewide voter lists were obtained from the Office of the Secretary of State, Elections Office, in Oregon. Additional detail on these data are included in the Data Appendix. The voter list is a list of individuals who are or have been registered to vote, including both active voters and inactive voters (as defined below). Only registered individuals may vote, and they must update their voter registration in the case of a move, a name change, or if they wish to register or

change an association with a political party.[1] Individuals who have neither voted nor (re)registered within 5 years may be moved from the active voter file to the inactive voter file. Voters are also moved to the inactive file if mailings to them are returned or if they are incarcerated. Voters who die, move out-of-state, are inactive for more than 5 years, or ask to be removed are moved to a third set of cancelled voter files, not included in this analysis. We define a voter as "registered" (and hence eligible to vote) if he/she is on the "active" voter file.

We obtained two copies of the voter list: one in June 2010 (hereafter, "2010 data pull") and another in July 2013 (hereafter, "2013 data pull"). These voter lists are complete records of active and inactive voters in the state at the time. These datasets cover different elections and contain different sets of people (because of new registrations or movements to the cancelled voter list in the interim). Each voter list includes the date the voter record was last updated, current political party registration, and voter histories for each individual that indicate whether or not that individual voted in a given election Elections include statewide general elections (e.g. the November 2008 and November 2010 elections), primary elections (e.g. the May 2008 and May 2010 primaries), and local elections in which certain districts vote on particular measures or elect local politicians (e.g. school board elections). These elections are shown in Appendix Table A1. The 2013 file contains elections from May 2006 through November 2012, while the 2010 file contains elections from May 2008 through May 2010 and also some smaller local elections not included in the 2013 file.

We probabilistically matched the Oregon Health Insurance Experiment population to the voting data using LinkPlus software. This was done using name, date of birth and gender. Those on the lottery list who did not appear in the voter registration records were assumed not to be registered (as well as not to have voted). Figure 1 shows the evolution of the study population from submitting names to inclusion in the voting analysis.

Analytic Specifications

Intent-to-Treat Effect of the Lottery (ITT)

Our treatment group is comprised of those selected in the lottery and our controls are those who were not. We estimate the intent-to-treat (ITT) effect of winning the lottery (i.e. the difference between treatment and controls) by fitting the following OLS equation:

$$y_{ih} = b_0 + b_1 LOTTERY_h + X_{ih} b_2 + V_{ih} b_3 + e_{ih} \qquad (1)$$

where *i* denotes an individual and *h* denotes a household.

[1] OregonLaws.org: Oregon Revised Statutes 2013, Chapter 247: http://www.oregonlaws.org/ors/chapter/247 [September 10, 2015]

LOTTERY is an indicator variable for whether or not household *h* was selected by the lottery. The coefficient on *LOTTERY* (β_1) is the main coefficient of interest, and gives the average difference in (adjusted) means between the treatment group (the lottery winners) and the control group (those not selected by the lottery); it is interpreted as the impact of being able to apply for OHP Standard through the Oregon lottery.

We denote by X_{ih} the set of covariates that are correlated with treatment probability (and potentially with the outcome) and therefore must be controlled for so that estimates of β_1 give an unbiased estimate of the relationship between winning the lottery and the outcome. In all of our analyses, X_{ih} includes indicator variables for the number of individuals in the household listed on the lottery sign-up form (hereafter "household size"); although the state randomly sampled from individuals on the list, the entire household of any selected individual was considered selected and eligible to apply for insurance. As a result, selected (treatment) individuals are disproportionately drawn from households of larger household size.

We denote by V_{ih} a second set of covariates that can be included to potentially improve power by accounting for chance differences between treatment and control groups in variables that may be important determinants of outcomes. These covariates are not needed for β_1 to give an unbiased estimate of the relationship between winning the lottery and the outcome, as they are not related to treatment status, but may improve the precision of the estimates by explaining some of the variance in the outcome.[2] Our primary analysis includes no such V_{ih} covariates, but as a secondary analysis, we explore whether our results are sensitive to inclusion of V_{ih} covariates measured pre-randomization, as described below.

In all of our ITT estimates and in our subsequent instrumental variable estimates (see below), we estimate linear models even though our outcomes are binary. Because we are interested in the difference in conditional means for the treatments and controls, linear probability models would pose no concerns in the absence of covariates or in fully-saturated models (Joshua D. Angrist, 2001, Joshua D. Angrist and Jörn-Steffen Pischke, 2009). Our models are not fully saturated, however, so it is possible that results could be affected by this functional form choice, especially for outcomes with very low or very high mean probability. We therefore explore the sensitivity of our results to an alternate specification using logistic regression and calculating average marginal effects for all binary outcomes.

In all of our analyses we cluster the standard errors on the household identifier since the treatment is at the household level. All analyses where outcomes are measured through July 15, 2010 are weighted to account for a new lottery conducted by the state starting in 2009 as described below.

[2] To determine whether to include these pre-randomization versions of the outcome, we estimated how much variance they explained in the control sample. The partial r-squareds ranged from 0.2 to 0.25.

Local Average Treatment Effect of Medicaid (LATE)

The intent-to-treat estimates from equation (1) provide an estimate of the causal effect of winning the lottery (i.e. winning the opportunity to apply for OHP Standard). This provides an estimate of the net impact of expanding *access* to public health insurance. We are also interested in the impact of insurance *coverage* itself. We model this as follows:

$$y_{ih} = \rho_0 + \rho_1 INSURANCE_{ih} + X_{ih}\rho_2 + V_{ih}\rho_3 + \eta_{ih} \quad (2)$$

where INSURANCE is a measure of insurance coverage and all other variables are as defined in equation (1). We estimate equation (2) by two stage least squares (2SLS), using the following first stage equation:

$$INSURANCE_{ih} = \delta_0 + \delta_1 LOTTERY_{ih} + X_{ih}\delta_2 + V_{ih}\delta_3 + \mu_{ih} \quad (3)$$

in which the excluded instrument is the variable *LOTTERY*.

We interpret the coefficient on insurance from instrumental variable estimation of equation (2) as the local average treatment effect of insurance, or LATE (Guido W. Imbens and Joshua D. Angrist, 1994). In other words, our estimate of π_1 identifies the causal impact of insurance among the subset of individuals who obtain insurance upon winning the lottery but who would not obtain insurance without winning the lottery (i.e. the compliers).

The LATE interpretation requires the additional identifying assumption that the only mechanism through which winning the lottery affects the outcomes studied is the lottery's impact on insurance coverage. We believe this is a reasonable approximation; in earlier work we discussed potential violations; where we could explore them we did not find cause for concern (Amy Finkelstein et al., 2012).

Analytic Weights

We use weights to adjust for a new lottery for OHP Standard which the state conducted beginning in the fall of 2009. Initially, the state mailed postcards to those on the original list that were not selected (our controls) asking if they would like to be included in this second lottery. Those who returned the postcard were added to the new waiting list and an initial draw was done just from that group. Following that initial draw, the state opened the new waiting list to the general public (including both our controls and our treatments as well people not on our original list); drawings from this list were conducted approximately monthly. Unlike the original 2008 waiting list, the new waiting list remained continuously open: individuals could sign up at any point. As with the original lottery, draws were done on individuals, but the opportunity to apply for OHP (treatment) was extended to the whole household. After each drawing, we probabilistically matched (using LinkPlus software) the new waiting list to our study population

to identify individuals who were eligible for selection by the state (called "opt-ins") and those who were actually selected in a given drawing (called "selected opt-ins").

Given the difficulty in interpreting the "treatment" received by those who were drawn in the new lottery, we drop the selected opt-ins from our analytic sample and use weights to correct for this. For each lottery drawing, the set of opt-ins is not a random sample of our study population: signing up for the new list was optional, and thus subject to the influence of factors such as underlying health. However, the set of selected opt-ins *is* a random sample of the opt-ins. We therefore use weights to adjust for the individuals dropped because of the second lottery using the following principle: within any (even non-random) subset of the original study population, a randomly selected group can be weighted to stand in for the non-selected remainder based on the probability of that random selection.

The weights we use are roughly analogous to weighting done for censoring or attrition in longitudinal data ((Stephen R. Cole and Miguel. A. Hernán, 2008), (Graham Kalton, 1986)). As in those settings, we weight each observation at each time point by the inverse probability of being in the sample, and we generate overall weights as the product of the weights across all time points. In our setting, the time points correspond to the continued lottery drawings from the new waiting list. We do not need to model the probability of being selected in the new lottery as a function of covariates; we know the process was random and we can observe the selection proportions.

More formally, let O_t be the set of opt-ins in our study population eligible for new lottery drawing on date t. Let S_t be the set of opt-ins selected in drawing on date t. We define the

weight for individual *i* to be:

$$w_t(i) = \begin{cases} \frac{1}{1-p_t} \; if \; i \; in \; O_t \; and \; in \; S_t \\ 0 \; if \; i \; in \; S_t \\ 1 \; if \; i \; not \; in \; O_t \end{cases} \qquad (4)$$

where p_t is the probability of an opt-in being selected.

Selection probabilities varied by the number of household members on the new list, so in all cases, we estimated the selection probability separately by strata of "tickets" (household members on the new waiting list at time t).

The final analytic weight W is simply the product all the weights w_t introduced up to the end date. This end date is chosen based on the date of the outcome analyzed. Analysis of different outcomes use different weights.

We refer to the set of weights by their end date. (i.e. June 2010 weights use the product of weights up through June 2010). Analyses of November 2008 voting are unweighted (since this

occurred prior to the lottery. Analyses of registration (as of June 2010), and of "any other election in the data" (which includes elections through June 1, 2010) using the June 2010 weights; analyses of November 2010 voting use the November 2010 weights.

Table A4 shows the distribution of the June 2010 and November 2010 weights. One can see that the November 2010 weights involve a much greater share of individuals with zero weights (and a higher upweighting of the remaining individuals), reflecting several large new lottery draws that occurred between those dates. The control group is far more impacted by the weights than the treatment group as they were more likely to sign up for the new lottery.

Relationship between the Lottery and Insurance Coverage

Table A5 reports the control means and effects of lottery selection for various definitions of insurance coverage. Being selected in the lottery is associated with an increase of 24 percentage points (SE 0.381) in the probability of ever having Medicaid coverage between March 2008 and June 2010 and with an increase of 23 percentage points (SE 1.295) in the probability of ever having Medicaid coverage between March 2008 and November 2010. We use this increase in insurance coverage due to the lottery to estimate local average treatment effects.[3]

Lottery selection is associated with an average increase of 4.3 months on Medicaid by June 2010 and an average increase of 4.8 months on Medicaid by November 2010 (row 3) – both because only a subset of those selected in the lottery obtained coverage and because those who obtained coverage were not necessarily covered for the entire study period. For those who did obtain coverage through the lottery, there is an increase of 20.8 months on Medicaid. This is less than the 32 months in the study period for several reasons: lottery selection occurred in 8 draws between March and October 2008, initial enrollment in OHP took 1-2 months after lottery selection, and some of those enrolled in Medicaid through the lottery lost coverage by failing to recertify.

[3] There are two distinct Oregon Medicaid programs: the program for the traditional Medicaid population (OHP Plus) and the program for the expansion population (OHP Standard). We define someone as ever on "Medicaid" if they are on either Medicaid program, including both Plus and Standard. Since the lottery was for the OHP Standard program, that is where we would expect to find increases in coverage, and this is borne out in the data. In fact, the increase in OHP Standard is slightly greater than the increase in any Medicaid (25.9 percentage points compared to 23.4), suggesting that some of the increase in OHP Standard may have come from individuals who would have been on another Medicaid program at some point during the study period. The effect of the lottery on Medicaid coverage attenuates over time for two reasons. First, those who successfully enroll in OHP (through the lottery or other means) are required to recertify eligibility every six months, leading to attrition in coverage. Additionally over time, those not selected in the lottery may obtain Medicaid coverage through the OHP Plus program.

Balance Analysis

Table A2 shows treatment-control balance on pre-randomization characteristics. The characteristics measured at lottery sign up are balanced (F-statistic = 1.32, p-value = 0.23), as is a measure of whether the individual voted prior to the lottery.

A concern with these data is the potential for endogenous presence in the voting data file based on post-lottery behavior. The voting data are meant to represent a complete history of anyone who had any activity in preceding years, but the nature of the match and exit to the cancelled data file leave open this possibility. For example, if the lottery affected voting behavior in 2008, and voting behavior in 2008 affected presence in the 2013 data pull, then using information in the 2013 data pull to infer the effect of the lottery on voting behavior in 2008 would be contaminated by differential selection of treatment and control groups into the sample. In Table A3 we test for this possibility by examining the balance between treatment and control groups in voting in pre-lottery (2006 or 2007) elections in the 2013 data pull (such outcomes are not available in the 2010 data pull). The lottery cannot have influenced this pre-lottery voting, so any differences between treatment and control groups would be indicative of differential matching to the voting data. There are no significant differences in pre-lottery voting or in the combination of characteristics measured at lottery sign-up and voting in elections that occurred before the lottery. To further probe potential concerns about endogenous presence in the post-lottery data pulls, we also examine whether "entry" into the data pull (i.e. appearing in 2013 data pull but not 2010 data pull) or "exit" (i.e. appearing in the 2010 data pull but not the 2013 data pull) is balanced across treatment and control groups (see appendix for more discussion); none of these tests provides evidence of imbalance.

Planned Analysis of Outcomes

Voter Registration

In Table 1, we analyze whether an individual is **an active registered voter** in June 2010, and whether he or she has a **registered party affiliation**. We consider all voters on the "active" voter files as currently registered and eligible to vote. We consider whether an individual was registered as a Democrat, Republican, affiliated with another political party, or a non-affiliated voter.[4]

[4] Other political parties are: Americans Elect, Constitution, Independent, Libertarian, Pacific Green, Progressive, and Working Families.

In June 2010, 42% of the control sample was on the active registered voter rolls. This is lower than figures for the Oregon population: the Census Bureau estimates that 73% of Oregon citizens (67.5% of the total population) was registered to vote in 2008.[5]

Voting

We analyze three different voting outcomes in Table 2: whether an individual voted in the **November 2008 election, the November 2010 election,** or **any other election between January 2008 and June 2010 (**excluding the 2008 November elections). Post-lottery elections occur from May 20, 2008 to June 1, 2010 (see Table A1).

In the control sample, 33.6% voted in the November 2008 election, 22.6% voted in the November 2010 election, and 20.5% voted in another election.

The rate of voting among the control sample was lower than both the statewide voting rate in Oregon and the national voting rate: 33.6% of the control sample voted in the 2008 presidential election, compared to the estimated 67.7% of citizens in Oregon and 63.6% of voting-age citizens nationally who voted in that election.[6] This disparity is likely slightly overstated, since our study sample does not include citizens over the age of 65, who vote at higher rates than younger individuals;[7] our list may also include non-citizens.

Heterogeneity

Table 3 and 4 explores potential heterogeneity in the voting and registration results along dimensions we can measure at the time of lottery sign-up: gender, age (19-49 vs. 50-64), whether an individual requested English language materials, and whether they did and did not live in a zip code located in a county where the majority then voted for Barack Obama in the 2008 presidential election.[8]

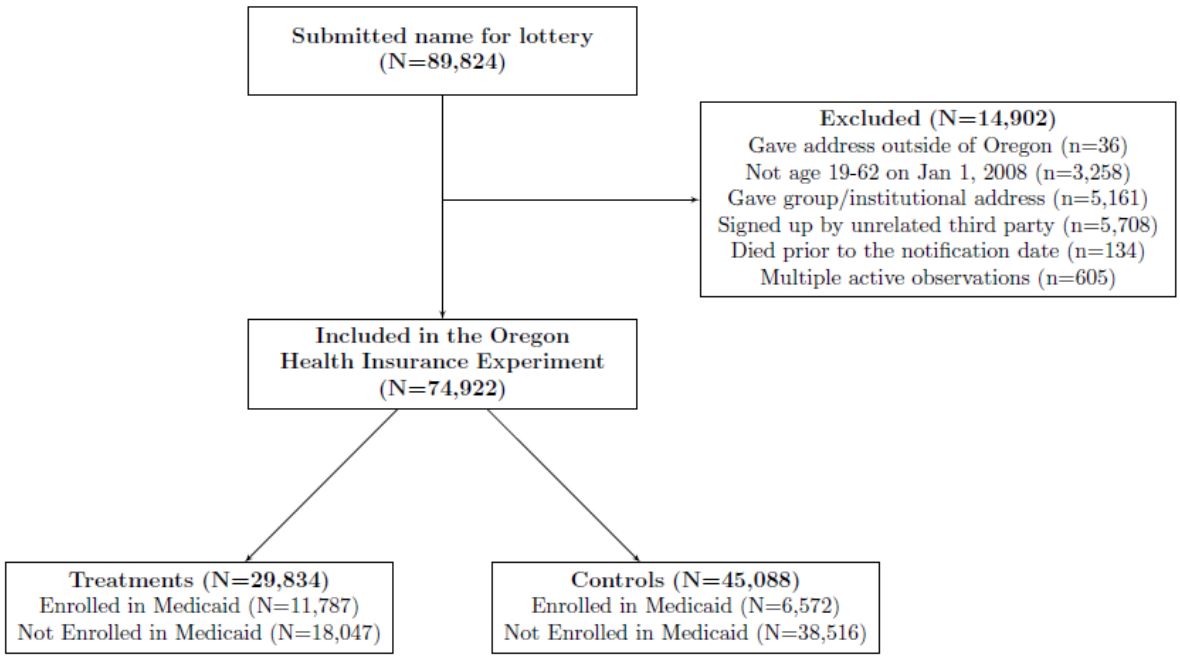
Sensitivity

We examine the sensitivity of our baseline results in Table 1 and 2 for voting and registration to several alternative specifications; these results are shown in Table A6 (voting) and A7 (registration). In Panel A we analyze the sensitivity to measuring outcomes in the 2013 data file that were previously measured in the 2010 data file. In Panel B, we investigate the sensitivity of

[5] http://www.census.gov/hhes/www/socdemo/voting/publications/historical/index.html (Table A-5b) [2 May 2014]
[6] Tables 3 and 1, respectively. http://www.census.gov/prod/2010pubs/p20-562.pdf [April 23, 2014]
[7] Ibid – Table 2.
[8] Zip codes were linked to counties using a crosswalk available via the Department of Housing and Urban Development (Department of Housing and Urban Development (HUD), 2015). Where a zip code was split across two counties, it was assigned to the county in which the greatest share of addresses are located (115/411 Oregon zip codes in the HUD data were split across counties). We were able to assign the majority of the study sample to a county (0.78% were unassigned).

our baseline results in Tables 1 and 2 to controlling for pre-randomization covariates. In Panel C, we investigate the sensitivity of our results to estimating logit models rather than linear probability models. In Panel D (Table A6 only), we explore the sensitivity of our voting analysis to controlling for whether the individual voted in elections in 2006 or 2007 (prior to lottery), which is available only in the 2013 data.

Figure 1


Submitted name for lottery
(N=89,824)
Excluded (N=14,902)
Gave address outside of Oregon (n=36)
Not age 19-62 on Jan 1, 2008 (n=3,258)
Gave group/institutional address (n=5,161)
Signed up by unrelated third party (n=5,708)
Died prior to the notification date (n=134)
Multiple active observations (n=605)
Included in the Oregon
Health Insurance Experiment
(N=74,922)
Treatments (N=29,834)
Enrolled in Medicaid (N=11,787)
Not Enrolled in Medicaid (N=18,047)
Controls (N=45,088)
Enrolled in Medicaid (N=6,572)
Not Enrolled in Medicaid (N=38,516)

References

Angrist, Joshua D. 2001. "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice." *Journal of Business & Economic Statistics*, 19(1), 2-16.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Baicker, Katherine; Amy Finkelstein; Jae Song and Sarah Taubman. 2014a. "The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment." *American Economic Review*, 104(5), 322-28.

Baicker, Katherine; Amy Finkelstein and Sarah Taubman. 2014b. "The Oregon Health Insurance Experiment: Evidence from Criminal Charges Data," Analysis Plan: Archived April 30, 2014:

Baicker, Katherine; Sarah L. Taubman; Heidi L. Allen; Mira Bernsterin; Jonathan H. Gruber; Joseph P. Newhouse; Eric C. Schneider; Bill J. Wright; Alan M. Zaslavsky and Amy N. Finkelstein. 2013. "The Oregon Experiment -- Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine*, 368(18), 1713-22.

Cole, Stephen R. and Miguel. A. Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology*, 168(6), 656-64.

De La O, Ana. 2013. "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico." *American Journal of Political Science*, **57**(1), 1-14.

Finkelstein, Amy; Sarah Taubman; Bill Wright; Mira Bernstein; Jonathan Gruber; Joseph P. Newhouse; Heidi Allen; Katherine Baicker and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics*, 127(3), 1057-106.

Imbens, Guido W. and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2), 467-75.

Kalton, Graham. 1986. "Handling Wave Nonresponse in Panel Surveys." *Journal of Official Statistics*, 2, 303-14.

Labonne, Julien. 2013. "The Local Electoral Impacts of Conditional Cash Transfers: Evidence from a Field Experiment." *Journal of Development Economics*, 104, 73-88.

Manacorda, Marco; Edward Miguel and Andrea Vigorito. 2011. "Government Transfers and Political Support." *American Economic Journal: Applied Economics*, 3 (3), 1-28.

Pop-Eleches, Christian and Grigore Pop-Eleches. 2012. "Targeted Government Spending and Political Preferences." *Quarterly Journal of Political Science*, 7(3), 285-320.

Taubman, Sarah L.; Heidi L. Allen; Bill J. Wright; Katherine Baicker and Amy N. Finkelstein. 2014. "Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment." *Science*, 343(6168), 263-8.

Zucco, Cesar. 2013. "When Payouts Pay Off: Conditional Cash Transfers and Voting Behavior in Brazil 2002–10." *American Journal of Political Science*, 57(4), 810–22.

VOTING DATA APPENDIX

Voting data were obtained from the Oregon Secretary of State's Election Division in June 2010 and July 2013. Each data set contains a list of active voters and inactive voters and voting history over a series of preceding elections. Each voting data set was merged separately with our lottery list using the usual probabilistic matching techniques in preparation for analysis.

Voting Data

The statewide voter list is maintained by the Office of the Secretary of State, Elections Office, in Oregon. We received a file of "active voters" and a file of "inactive voters." An "active" voter is someone who has voted or re-registered within the last 5 years. In Oregon, people who do not vote for more than 5 years need to re-register to remain active. Only registered individuals may vote, and they must update their voter registration in the case of a move, a name change, or if they wish to register or change an association with a political party (OregonLaws, Chapter 247). People who failed to vote or re-register for 5 years are moved to the "inactive" file; people on the active file can also be moved to the inactive file for a variety of other administrative reasons (such as being incarcerated or having a bounced-back ballot mailing or a signature challenge that the individual did not respond to). Finally, people who die, are found to be registered in another state, are "inactive" for five years, or who ask the state to have their registration cancelled are moved to a third, "cancelled" file not included in this analysis.

We obtained these files on two separate occasions: in June 2010 and another in July 2013. For each individual, each file contains the date of the most recent registration, current political party registration, and voter histories that indicate whether or not an individual voted in a given past election. Elections in the data are both statewide elections like the 2008 general election, and local elections in which certain districts vote on particular measures or elect local politicians (e.g. school board members). The 2013 file contains elections from May 2006 through Nov 2012, while the 2010 file contains elections from May 2008 through May 2010 and also some smaller local elections not included in the 2013 file.

We use these data to analyze:

1. Registration (measured in the June 2010 data pull):
    a. Whether the individual was registered to vote in June 2010. We define individuals as "registered to vote" in June 2010 if they are in the "active file" in the June 2010 data pull.
    b. With what party the individual was registered to vote in June 2010 data pull.
2. Voting
    a. Whether the individual voted in the November 2008 general election. We can measure this in both the June 2010 and the June 2013 data pulls. Our baseline specification measures this in the June 2010 data pull. We can also measure November 2008 voting in the June 2013 file, but we lose information on people who were moved to the "cancelled file." We examine robustness to this alternative measure.)
    b. Whether the individual voted in "any post-lottery election in the data through June 2010 other than the November 2008 general election", as measured in the June 2010 data pull. (As seen in Table A1, included elections in 2008 are May 27, July 15, September 16, October 7, November 18; elections in 2009 on March 10, May 5, May 19, June 23, August 11, September 15, September 29, October 13, November 4, November 17, December 8, December 15, December 29; and elections in 2010 on January 26, March 9, May 15, June 1.)
    c. Whether the individual voted in the November 2010 general election, as measured in the June 2013 data pull.

In addition, as a balance test, we use the 2013 file to test for balance in voting behavior in the 2006 and 2007 elections; the file we received in June 2010 did not contain these pre-lottery elections.

Merge

We probabilistically matched the Oregon Health Insurance Experiment population to the 2010 and 2013 voting files using LinkPlus software. This was done using name, date of birth, and gender. Due to the protected nature of the lottery data, matching of the lottery data to the voting data was done on a secure, non-networked computer, and all identifiers were removed before analysis.

Individuals on the lottery list could thus match to each of the voter files or not. For each election represented in each voter file (e.g. 2008 election as described in the 2010 voter data file), lottery list members who matched could be characterized as having voted, being registered for the election but not having voted, or not having a record for that particular election (e.g. having registered to vote in 2009, so not having a voting record for 2008).

Assessing data quality

Not everyone in the 2010 voter files appears in the 2013 voter files (and vice versa). Table A8 summarizes these results. For example, it shows that of 43,201 people in the 2010 voter files (active or inactive), only 37,310 are in the 2013 voter files. Likewise, of the 40,819 people in the 2013 voter files, only 37,310 are in the 2010 voter files. There are several potential reasons for this. First, there could matching noise introduced by our probabilistic matching techniques. Second, there could be genuine entry into the data between 2010 and 2013, due to new registrations. Third, there could be genuine exit from the data between 2010 and 2013, due to individuals being moved to the "cancelled" file because of death, incarceration, a move out of state, remaining inactive for 5 years, or other administrative reasons.

Reassuringly, Table A8 shows that only 170 people (0.2 percent of the lottery list) enter the data between 2010 and 2013 *and* are recorded as having voted in the 2008 election; these presumably reflect errors in our probabilistic match. Likewise, of the 32,023 people who have a voting record (yes/no) in 2008 recorded in the and 2010 and 2013 data, only 12 (<0.01%) have a different outcome recorded. These checks suggest only a small amount of noise in our measures.

Our primary concern, however, is not with noise (mis-measurement, mis-matching, attrition etc.) per se, but the potential for endogenous selection into the sample based on post-lottery behavior. For example, if the lottery affected voting behavior in 2008, and voting behavior in 2008 affected presence in the 2010 files (i.e. someone who might otherwise have been moved to "cancelled" is maintained), then using information in the 2010 file to infer the effect of the lottery on voting behavior in 2008 would be contaminated by differential selection of treatment and control groups into the sample. Likewise, any impact the lottery had on mortality or moves out of state could also affect our ability to measure 2008 or 2010 voting behavior. Reassuringly, our prior analysis shows no substantial effect of the lottery on mortality (Finkelstein et al., 2012), but the other avenues still have the potential to affect the sample we observe.

We performed two tests for such potential endogenous measurement. First, we looked at whether entry or exit between 2010 and 2013 was correlated with treatment status. Second, using the 2013 file, we analyzed whether pre-lottery (2006 or 2007) voting was correlated with treatment status. In each case we ran the following regression:

$$y_{ih} = \beta_0 + \beta_1 \text{LOTTERY}_h + X_{ih}\beta_2 + \varepsilon_{ih} \quad (1)$$

where i indexes individuals and h indexes households, LOTTERY is an indicator for whether household h was selected in the lottery. X_{ih} includes controls for household size indicators. Standard errors are clustered on the household.

Table A3 shows the results, which are reassuring. The top panel shows that the probability of voting in the pre-lottery period is balanced between treatment and control, and the bottom panel shows that entry into and exit out of the data sets are also balanced.

Table 1: Registration

	Extensive Margin			
	Mean Value in Control Group	Effect of Lottery Selection	Effect of Medicaid Coverage	p-value
	(1)	(2)	(3)	(4)
Registration				
In the June 22, 2010 Active Voter File	0.420			
Party Affiliation in June 2010				
Registered as a Democrat	0.253			
Registered as a Republican	0.103			
Registered with another political party	0.051			
Registered as a non-affiliated voter	0.173			

Notes: All regressions control for household size and adjust standard errors for household clusters. Regressions are weighted to account for a series of new Medicaid lottery draws that began in Fall 2009, using weights that account for lottery selection through June 1, 2010. All registration outcomes are measured in the 2010 data pull.

Table 2: Voting

		Extensive Margin		
	Mean Value in Control Group	Effect of Lottery Selection	Effect of Medicaid Coverage	p-value
	(1)	(2)	(3)	(4)
November 2008 Election	0.336			
November 2010 Election	0.226			
Any other post-lottery election through June 2010 except November 2008 election	0.205			

Notes: All regressions control for household size and adjust standard errors for household clusters. Voting data for the November 2008 election and the "any post lottery election through June 2010 except November 2008" outcomes come from the 2010 data pull. Voting data for the 2010 midterm election come from the 2013 data pull. Analyses of the 2008 election is unweighted; analysis of the 2010 election and the "any other post lottery election" use November 2010 and June 2010 weights respectively. Table A1 lists the "other post lottery elections".

Table 3: Heterogeneity (Voting)

			Voted in November 2008		
	N	First Stage	Control Mean	Effect of Medicaid Coverage	p-value
Baseline Results	74922	0.271	33.645		
Gender					
Female	41249	0.264	36.979		
Male	33673	0.281	29.453		
Age					
Ages 19-49	54814	0.263	30.244		
Ages 50-64	20108	0.294	42.768		
English-language lottery materials					
No	6440	0.189	7.021		
Yes	68482	0.279	35.898		
Zip in a Democratic county (2008)					
No	26139	0.279	32.121		
Yes	48199	0.267	34.428		

Notes: Table reports effects on 2008 voting (from Table 2) for various subsamples. Control means are reported as percentages. Regressions control for household size and adjust standard errors for household clusters. Voting in 2008 is measured in the 2010 data pull.

Table 4: Heterogeneity (Registration)

			Registered (In the June 22, 2010 Active Voter File)		
	N	First Stage	Control Mean	Effect of Medicaid Coverage	p-value
Baseline Results	74922	0.238			
Gender					
Female	41249	0.226			
Male	33673	0.254			
Age					
Ages 19-49	54814	0.231			
Ages 50-64	20108	0.256			
English-language lottery materials					
No	6440	0.164			
Yes	68482	0.245			
Zip in a Democratic county (2008)					
No	26139	0.247			
Yes	48199	0.232			

Notes: Table reports effects on 2010 registration (from Table 1) by various subsamples. All regressions control for household size and adjust standard errors for household clusters. Regressions are weighted to account for a series of new Medicaid lottery draws that began in Fall 2009, using weights that account for lottery selection through June 1, 2010. All registration outcomes are measured in the 2010 data pull.

Table A1: Elections in the Data

Date	Election	Data
5/16/2006	Statewide primary (including: OR Governor, US Congress, OR supreme court judge, OR state legislature)	2013
11/7/2006	State general election (including: Governor, US Congress, OR supreme court judge, US state legislature)	2013
5/15/2007	Off-year primary election	2013
11/6/2007	Special election	2013
5/20/2008	Statewide primary election (including: President, U.S. Congress state legislature)	2013
5/27/2008	Local elections	2010
7/15/2008	Local elections	2010
9/16/2008	Local elections	2010
10/7/2008	Local elections	2010
11/4/2008	Statewide general election (including: President, U.S. Congress, OR state legislature)	2010, 2013
11/18/2008	Local elections	2010
3/10/2009	Local elections	2010
5/5/2009	Local elections	2010
5/19/2009	Primary election	2010, 2013
6/23/2009	Local elections	2010
8/11/2009	Local elections	2010
9/15/2009	Local elections	2010
9/29/2009	Local elections	2010
10/13/2009	Local elections	2010
10/27/2009	Local elections	2010
11/3/2009	Local elections	2010
11/4/2009	Local elections	2010, 2013
11/17/2009	Local elections	2010
12/8/2009	Local elections	2010
12/15/2009	Local elections	2010
12/29/2009	Local elections	2010
1/26/2010	Statewide special election (to vote on two tax measures)	2010, 2013
3/9/2010	Local elections	2010
5/18/2010	Statewide primary election (including US Congress; OR governor; OR state legislature)	2010, 2013
6/1/2010	Local elections	2010
11/2/2010	Statewide general election (including US Congress; OR governor; OR state legislature)	2013

Table A2: Treatment-Control Balance

	Control mean	Treatment-control difference	p-value
	(1)	(2)	(3)
Lottery List Variables			
Year of Birth	1967.998	0.162 (0.10)	0.104
Female	0.557	-0.007 (0.00)	0.039
English as preferred language	0.922	0.002 (0.00)	0.346
Signed up self	0.918	0.000 (0.00)	0.273
Signed up first day of lottery	0.093	0.001 (0.00)	0.627
Gave Phone Number	0.862	-0.003 (0.00)	0.300
Address is a PO Box	0.117	0.000 (0.00)	0.873
Zip code median household income	39265.40 (8463.54)	44.89 (72.89)	0.54
Ever voted in the pre-period	0.215	-0.001 (0.003)	0.700
F statistic for lottery list variables			1.322 0.227
F statistic for lottery list variables and pre-period voting			1.181 0.302

Notes: We report the coefficient on LOTTERY from estimating equation (1) on the dependent variable shown in the first column. All dependent variables are measured based on the lottery sign up, except for "every voted in the pre-period" which is defined as voting in a 2006 or 2007 election, as meaured in the 2013 data pull. All regressions include indicators for the number of household members on the lottery list, adjust standard errors for household clusters, and are unweighted. The final rows report the pooled F-statistics (and p-values) from testing treatment-control balance on sets of variables jointly.

Table A3: Tests of Balance for Sample Selection

	Control mean	Treatment-control difference	p-value
	(1)	(2)	(3)
Voted in November 2006 election	0.159	0.004	0.145
		(0.00)	
Voted in November 2007 election	0.145	-0.003	0.359
		(0.00)	
Voted in any 2006 or 2007 election	0.215	-0.001	0.700
		(0.00)	
Entry	0.047	0.000	0.826
		(0.00)	
Exit	0.081	0.000	0.884
		(0.00)	

Notes: We report the coefficient on LOTTERY from estimating equation (1) on the dependent variable shown in the first column. All regressions include indicators for the number of household members on the lottery list, adjust standard errors for household clusters, and are unweighted. The first three rows (analyzing voting in pre-lottery elections) use data from the 2013 data pull. "Voted in any 2006 or 2007 election" includes the November 2006 state elections and the November 2007 special election (including 2 ballot measures) in the previous rows, as well as the May 2006 and May 2007 primaries. "Entry" is an indicator for individuals who appear in the 2013 data pull but not in the 2010 data pull. "Exit" is an indicator for individuals who appeared in the 2010 data pull but not in the 2013 data pull.

Table A4: Distribution of the Weights

	Mean	Standard Deviation	Minimum	Median	75th%ile	95%ile	Max	N	% with zero weight
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
June 2010 weights									
Full Sample	0.9998	0.3748	0.0000	1.0000	1.0000	1.3949	3.8522	74922	0.09
Controls	0.9999	0.4489	0.0000	1.0000	1.2595	1.4888	3.8522	45088	0.13
Treatments	0.9997	0.2199	0.0000	1.0000	1.0000	1.2595	2.7083	29834	0.04
June 2010 weights (non-zero weights)									
Full Sample	1.103	0.202	1.000	1.000	1.112	1.395	3.852	67885	
Controls	1.153	0.236	1.000	1.000	1.395	1.548	3.852	39097	
Treatments	1.036	0.112	1.000	1.000	1.000	1.260	2.708	28788	
November 2010 weights									
Full Sample	0.994	3.173	0.000	1.000	1.000	1.000	190.019	74922	0.35
Controls	0.990	3.568	0.000	1.000	1.000	1.168	190.019	45088	0.44
Treatments	0.998	2.458	0.000	1.000	1.000	1.000	139.360	29834	0.21
November 2010 weights (non-zero weights)									
Full Sample	1.527	3.828	1.000	1.000	1.000	3.305	190.019	48767	
Controls	1.771	4.624	1.000	1.000	1.000	9.498	190.019	25217	
Treatments	1.265	2.705	1.000	1.000	1.000	1.000	139.360	23550	

Notes: Table shows the distribution of weights used to account for the new health insurance lottery that started in the fall of 2009. The top panel (June 2010 weights) displays the distribution of weights used to analyze registration and voting in elections (excluding the 2008 general election and the 2010 midterms), accounting for new lottery selection through June 1, 2010. The bottom panel (November 2010 weights) displays the distribution of weights used to analyze voting in the 2010 Midterms, accounting for new lottery selection through November 4, 2010.

Table A5: Insurance Coverage (First Stage Estimates)

	November 2008		June 2010		November 2010	
	Control mean	Estimated FS	Control mean	Estimated FS	Control mean	Estimated FS
	(1)	(2)	(5)	(6)	(7)	(8)
Ever on Medicaid	8.69	27.11	18.480	23.784	21.551	22.859
		(0.33)		(0.381)		(1.295)
Ever on OHP Standard	1.66	26.21	4.041	26.224	6.041	25.181
		(0.29)		(0.313)		(1.063)
Number of Months on Medicaid	0.53	1.22	2.524	4.349	3.216	4.751
		(0.02)		(0.070)		(0.273)
On Medicaid at the end of the period	8.21	25.24	12.701	10.116	15.047	7.968
		(0.32)		(0.322)		(1.153)

Notes: We report the coefficient on LOTTERY from first stage estimates of equation (3) for different definitions of the dependent variable "INSURANCE." All our analyses in the paper use the definition in the first row ("Ever on Medicaid.") All regressions include indicators for the number of household members on the lottery list, adjust standard errors for household clusters, and include weights that account for the probability of being sampled in the new lottery; specifically, the analysis in Nov 2008 in unweighted, the analysis in June 2010 uses the June 2010 weights, and the analysis in November 2010 uses the November 2010 weights). The study period starts on March 10, 2008 and ends on June 22, 2010.

Table A6: Robustness (Voting)

	Extensive Margin			
	Mean Value in Control Group	Effect of Lottery Selection	Effect of Medicaid Coverage	p-value
	(1)	(2)	(3)	(4)
Panel A: Using 2013 data				
November 2008 Election	0.316			
Panel B: Additional controls				
November 2008 Election	0.336			
November 2010 Election	0.226			
Any other election in the data (excluding the November 2008 and 2010 elections)	0.205			
Panel C: Logit estimates				
November 2008 Election	0.336			
November 2010 Election	0.226			
Any other post-lottery election through June 2010 except November 2008 election	0.205			
Panel D: Controlling for pre-period voting (using 2013 data)				
November 2008 Election	0.316			
November 2010 Election	0.226			

Notes: All regressions control for household size and adjust standard errors for household clusters. Analyses of 2008 voting are unweighted, analyses of 2010 voting are weighted using November 2010 weights, and analyses of "other elections in the data" are weighted using June 2010 weights.

Panel A: Panel shows results for the November 2008 election using the 2013 data.

Panel B: In addition to the standard controls, we also control for some other observable characteristics: gender, birth year, whether English is the individual's first language, whether the individual is the head of the household, whether the individual signed up on the first day, whether the the individual gave their phone number, and the zipcode's median household income. This panel uses 2010 data.

Panel C: Effect of lottery selection estimated by logistic regression and reported as average marginal effect. This panel uses 2010 data.

Panel D: In addition to the standard controls, we also control for whether an individual voted in the pre-period (defined as having voted in at least one of the 2006 or 2007 elections, as measured in the 2013 data). Panel shows results for the November 2008 and November 2010 elections using the 2013 data.

Table A7: Robustness (Registration)

	Extensive Margin			
	Mean Value in Control Group	Effect of Lottery Selection	Effect of Medicaid Coverage	p-value
	(1)	(2)	(3)	(4)
Panel A: Using 2013 data				
Registration				
In the June 22, 2010 Active Voter File	0.409			
Party Affiliation in June 2010				
Registered as a Democrat	0.228			
Registered as a Republican	0.095			
Registered with another political party	0.050			
Panel B: Additional controls				
Registration				
In the June 22, 2010 Active Voter File	0.420			
Party Affiliation in June 2010				
Registered as a Democrat	0.253			
Registered as a Republican	0.103			
Registered with another political party	0.051			
Registered as a non-affiliated voter	0.173			
Panel C: Logit estimates				
Registration				
In the June 22, 2010 Active Voter File	0.420			
Party Affiliation in June 2010				
Registered as a Democrat	0.253			
Registered as a Republican	0.103			
Registered with another political party	0.051			

Registered as a non-affiliated voter	0.173

Notes: All regressions control for household size and adjust standard errors for household clusters. Regressions are weighted to account for a series of new Medicaid lottery draws that began in Fall 2009. For the registration outcomes analyzed in this table, we use weights that account for lottery selection through June 1, 2010. The Analysis Plan provides details on the construction of weights and Table A4 shows the distribution of weights.

Panel A: Panel shows results for registration outcomes using the 2013 data.
Panel B: In addition to the standard controls, we also control for some other observable characteristics: gender, birth year, whether English is the individual's first language, whether the individual is the head of the household, whether the individual signed up on the first day, whether the the individual gave their phone number, and the zipcode's median household income. This panel uses 2010 data.

Panel C: Effect of lottery selection estimated by logistic regression and reported as average marginal effect. This panel uses 2010 data.

Table A8: 2010 and 2013 Data Files

			2013 data				
				Matched			
			Not matched	Missing Nov 2008 Voting Data	Registered but Did Not vote in November 2008	Voted in November 2008	Total
2010 data	**Not matched**		28,212	3,245	94	170	31,721
	Matched	Missing Nov 2008 Voting Data	3,029	4,818	3	7	7,857
		Registered but Did Not Vote in Nov	1,155	46	8,731	7	9,939
		Voted in Nov 2008	1,707	53	5	23,640	25,405
	Total		34,103	8,162	8,833	23,824	74,922

Note: "Missing" from the 2008 November voting data means that the individual is matched to the voting data but we have no record of whether she voted in that election. That could be, for example, because she registered to vote after that election (but before the data pull).