THE OREGON HEALTH INSURANCE EXPERIMENT: EVIDENCE FROM THE IN-PERSON INTERVIEWS

Katherine Baicker
Sarah Taubman
Heidi Allen
Mira Bernstein
Jonathan Gruber
Joseph P. Newhouse
Eric Schneider
Bill Wright
Alan Zaslavsky
Amy Finkelstein
And the Oregon Health Study Group*

Analysis Plan
April 16, 2012

In addition to the individuals named above, The Oregon Health Study Group includes Matt Carlson (Portland State University), Tina Edlund (Deputy Directory, Oregon Health Authority), Charles Gallia (Oregon DHS), and Jeanene Smith (Office for Oregon Health Policy and Research).

We are grateful to Chris Afendulis, Josh Angrist, Jack Fowler, Guido Imbens, Larry Katz, Jeff Kling, Ken Langa, Stacy Lindau, Jens Ludwig, Thomas McDade, Ben Olken, and the team from the National Center for Health Statistics for helpful comments and advice, to Brandi Coates, Sara Kwasnick, Nivedhitha Subramanian, and Annetta Zhou for expert research assistance, to our fielding staff for exceptional participant recruitment and data collection, and to numerous Oregon state employees for help acquiring the necessary data and for answering our many questions about the administration of state programs. We gratefully acknowledge funding from the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services, the California HealthCare Foundation, the John D. and Catherine T. MacArthur Foundation, the National Institute on Aging (P30AG012810, RC2AGO36631 and R01AG0345151), the Robert Wood Johnson Foundation, the Sloan Foundation, the Smith Richardson Foundation, and the U.S. Social Security Administration (through grant 5 RRC 08098400-03-00 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium). We also gratefully acknowledge Centers for Medicare and Medicaid Services' matching funds for this evaluation. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, the National Institute on Aging, the National Institutes of Health, any agency of the Federal Government, any of our funders, or the NBER.

Introduction

The goal of the analysis described here is to use the Oregon Health Insurance Experiment and the data we collected through in-person interviews and physical exams to estimate the effects of expanding Medicaid availability to a population of low-income adults. This analysis examines the effects on health care utilization (including preventive care), finances, and health, as well as access to health care, quality of care, and overall wellbeing. There is a particular emphasis on health using the physiologic measures obtained through our in-person data collection.

This analysis plan aims to pre-specify the analysis before comparing outcomes for treatment and control groups. By creating this analysis plan, which serves as a record of our ex ante planned analysis, we hope to minimize issues of data mining and specification searching. We do use the control distributions for all the outcomes and perform treatment-control comparisons that explore the validity of our analysis (such as balance on pre-randomization characteristics and uptake of insurance). This plan was constructed after viewing the findings from a mail survey and administrative data collected approximately one year after the lottery (Finkelstein et al. 2011). The methods and measures proposed here follow those undertaken in that analysis very closely; the primary new outcome measures are physiologic measures of health (Tables 4 and 5), measures of medical diagnoses and treatments (Table 7), as well as some additional self-reported health measures (Table 6). This plan is not intended to include all analysis that will eventually be done using these data – only to capture the analysis that will inform the first paper(s).

Methods

Randomization and intervention

In 2008, Oregon selected roughly 30,000 individuals by lottery from a waiting list of over 80,000 for an otherwise closed Medicaid program. The state conducted eight lottery drawings from March through 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 (OHP Plus); to be eligible, individuals must be adults ages 19 – 64, not otherwise eligible for Medicaid or other public insurance, Oregon residents, U.S. citizens or legal immigrants, have been without health insurance for six months, have income below the federal poverty level, and have assets below $2,000. Among the randomly selected individuals, those who completed the application process and met the eligibility criteria were enrolled in OHP Standard. OHP Standard provides relatively comprehensive medical benefits 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 (Finkelstein et al. 2011)

In-person data collection

Between September 2009 and December 2010, we conducted a large in-person data collection effort to assess a wide variety of outcomes. The 20,745-person sample for the in-person data collection included almost all of the individuals selected in the lottery living in the Portland area and a roughly equal number of unselected controls. We focused on the Portland area because of the logistical constraints of in-person data collection.

Potential participants were released into the field at regular intervals to be recruited for interviews. The recruitment protocol began with a postcard and a phone call attempt; potential participants were then assigned to interviewers, who used a flexible approach (including phone calls, text messages, home visits, and other contacts) to schedule interviews. The interviewers worked closely with our tracking team to locate potential participants if our contact information was out-of-date. As fielding progressed, we would periodically select a random subset of those participants who had not yet completed an interview. Recruitment would stop on the selected random subset, allowing the fielding staff to put additional time and resources into recruiting those remaining. In addition to this "continuous intensive follow-up," in the final months of fielding we expanded our recruitment efforts to encourage those remaining to complete the interview; at this point we increased the compensation, developed new tracking and recruitment procedures, and travelled with a mobile unit to individuals who had moved out of Portland.

Starting in November 2009, Oregon opened a new lottery for OHP Standard coverage. We worked with the state to identify which of our potential participants signed-up for the new lottery and which were selected over the course of subsequent lottery drawings. We did not interview participants who had been selected in the new lottery after they were informed of their new selection. However, we could not simply exclude those on the new lottery list, since signing up was voluntary – and thus not random. Rather, we took advantage of the fact that selection by the state from among those on the new list was random. Those potential participants who had signed up for the new list but were not selected could thus stand in for those on the list who had been selected. We use sample weights (described below) to adjust for dropping subsets either for intensive follow-up or the new lottery. Although we continued collecting data until December 2010, after the state had selected the entire new list, we limit our analysis to interviews completed by October 13, 2010. After that the weights to adjust for the new lottery become substantially more extreme and our estimates are unstable.

The interview generally lasted about 1 hour and 15 minutes, including about half an hour of interview questions, then medication cataloging, anthropometric measurements, blood pressure measurements, and collection of dried blood spots. Oversight of human subjects was performed by multiple IRBs. Written informed consent was obtained at the beginning of the interview; separate written informed consent for the dried blood spots was obtained immediately prior to that component. Interviews were mostly conducted at one of three clinic sites we

operated in Portland, although interviewers traveled to homes if participants preferred. Interviewers were not informed of participants' treatment status prior to the interview, but status sometimes became clear at the end of the interview when asking about insurance coverage. We provided participants with a preliminary "report of findings" about their health (including body mass index (BMI) and blood pressure) at the end of the interview and then with a mailed report of findings (adding the results of the blood tests and depression screening). Participants were compensated for their time and effort: $30 for the interview, an additional $20 for the dried blood spot collection, and $25 for travel if the interview was at a clinic site.

More detail on specific outcome measures is provided in the appendix; the survey instrument is available at www.nber.org/oregon.

Administrative data

In addition to the in-person interview data we collected, the state provided us with detailed data on Medicaid enrollment for every individual on the list (starting prior to the lottery). We use this to construct our primary measure of insurance coverage during study period.

Statistical analysis

Our analytic approach begins with an intent-to-treat model comparing outcomes for all those who were selected in the lottery (the study treatment group) to all those who were on the list but not selected (the study control group). Treatment assignment was done at the level of the household—if an individual was selected, all adult members of the household were offered the chance to apply for coverage—but because the lottery was conducted at the individual level, households with more members on the list were selected at higher rates. We estimate linear probability models for our outcomes with adjustment for the number of household members on the lottery list, which is required for unbiased estimates because the treatment probabilities vary by this. For the analysis of the blood pressure measures, we also adjust for age (in decile bins) and sex. These are not needed to prevent bias, as they are not related to treatment status. As important determinants of these outcomes in particular, however, they may increase the precision of our estimates of the effect of insurance by accounting for some of the variance in the outcome.[1] All standard errors are clustered by household to account for intra-household correlation.

All analysis is weighted using survey weights to account for the sample releases into the field and restriction of the active sample for intensive follow-up and in response to the new

[1] To decide when to add adjustment for additional pre-randomization variables, we examined how much of the variance in the outcome is explained by those variables (limiting to the control respondents). Only age and sex explained considerable variance and only for the blood pressure outcomes. For all other outcomes (and other potential covariates), we decided against additional controls in our baseline specification as we expect the gains in precision to be small and we prefer the more direct treatment and control comparison. In the sensitivity analyses described later in this document we explore how the rest of the results are affected by age and sex adjustment (and how the blood pressure results are affected by their removal).

lottery. The goal of these weights is to ensure that the final active sample is balanced on treatment status and representative of the full sample base. The weights are constructed using the general principle that for any given change to the sample (such as dropping part of the sample to conduct intensive follow-up on the remaining sample), those who were at risk for exclusion but remain in the sample are weighted by the inverse of the probability of being included. The final weights are the product of the weights for each change. The weights are discussed in more detail in the appendix.

In addition to the intent-to-treat analysis, which estimates the effect of *being selected in the lottery*, we also provide a local-average-treatment-effect analysis (sometimes called a complier-average-causal-effect analysis). This analysis, under certain assumptions, provides an unbiased estimate of the effect of *health insurance coverage* for those individuals for whom being selected in the lottery results in insurance coverage that they would not otherwise have obtained (Angrist, Imbens and Rubin 1996). We use being selected in the lottery as an instrument for being covered by health insurance (specifically by OHP Plus or OHP Standard) and estimate a two-stage least squares model with the same adjustments and weights as in the intent-to-treat model.

For each of the outcome domains we examine (e.g., health care utilization, access to health care, or self-reported health), we have multiple measures. Where there is a sensible existing summary measure for a given domain (such as total spending to summarize utilization), we use that; otherwise we present standardized treatment effects, which are a weighted average of the estimates. To derive standardized treatment effects, each individual measure in a domain is defined to have the same direction as the others (e.g. increases imply better health). The effect estimate for each individual measure is divided by the standard deviation of the control group for that measure (converting the effect estimate into standard deviation units). The standardized effect estimates for all the measures in a given domain are then averaged, producing a single standardized treatment effect for the domain. Although standardized treatment effects can provide a convenient summary for closely related measures and can increase power to detect small effects, they have several disadvantages. As used here, the included measures are all given equal weight, ignoring differences in the relative importance of the measures. Furthermore, the standardized treatment units are difficult to interpret.

Results

Preliminaries and initial analysis

The study population

Figure 1 shows the evolution of the study population from submitting names in the lottery to survey response. Of the 89,824 individuals who submitted names to the lottery, a total of 10,405 individuals selected in the lottery and 10,340 individuals not selected were sampled for inclusion in the in-person data collection effort. The in-person data collection effort was limited to the Portland area for logistical reasons. Of those sampled for inclusion, a total of 12,229

individuals responded to the survey by October 13, 2010 for an effective response rate of **73**%. Some of those included in the sample population were not "reachable", however: about **2**% of our sample was either deceased or incarcerated by the time of our fielding and an additional **6**% had moved out of the area; we completed interviews with 80% of the remaining individuals. The average date of survey response was April 24, 2010, approximately 25 months after randomization (standard deviation = 3 months).[2]

The characteristics of the respondent sample are shown in Table 1. Just over half the study participants are women, about a quarter are ages 50-64 (the oldest eligible age group), and about 70 percent are white. We did not see any significant differences between treatment and control groups on these characteristics or on the wide variety of baseline and interview characteristics that we examine in the appendix. There is likewise no significant difference between treatment and control groups in response rates to either the full survey (Table 2) or the components of the survey (see appendix).

Insurance coverage

Table 3 reports the difference in insurance associated with being selected in the lottery. Our primary measure of insurance coverage is whether the individual was ever on Medicaid (which includes both OHP Standard and OHP Plus) during our study period, as measured by the state's Medicaid enrollment files. The results indicate an increase of 24.1 percentage points in the probability of having Medicaid coverage. This is considerably less than 100 percent, reflecting imperfect take-up of Medicaid by those selected in the lottery and (to a lesser extent) those not selected becoming eligible for OHP Plus over our study period (Allen et al. 2010, Finkelstein et al. 2011). Appendix Table A4 reports the effects using alternative definitions of insurance coverage and tests for additional effects of lottery selection. We see no evidence of changes in private insurance rates and only substantively trivial impacts on receipt of other social services.

Health

We hypothesize that health insurance may improve health. Tables 4 and 5 summarize the results for our physiologic health measures and Table 6 summarizes the results for our self-reported health measures. Table 7 combines the measures on health with those on diagnosis and medication.

[2] For the purposes of this analysis plan and subsequent papers, we define the study period as beginning on March 10, 2008, which is the first date that anyone was notified of being selected in the lottery. In Finkelstein et al. 2011, we used a slightly different definition of the study period based on individual notification dates. Using the same definition as in Finkelstein et al., our average survey response occurs 22 months after notification (standard deviation = 4 months) or 20 months after insurance approval (standard deviation = 4 months).

Physiologic health measures

Table 4 summarizes the results for our physiologic measures of health. We consider, as continuous variables, measurements of body mass index (BMI), systolic and diastolic blood pressure, total and HDL cholesterol, and Hemoglobin A1c. In addition, we consider whether subjects have various clinical conditions using cut-points of these measures. Specifically, we consider whether they are overweight or obese, have pre-hypertension or hypertension, have elevated or high cholesterol, or have pre-diabetes or diabetes. Each of these measures is discussed in detail in the appendix. In the first panel of four columns, we examine the physiologic measure for the full sample. In subsequent columns and tables, we add to this primary analysis by considering both diagnosis and treatment (see discussion below). We do not present standardized treatment effects for the physiologic health measures because we use the Framingham risk score to summarize (see discussion of Table 5 below).

Limiting on pre-randomization diagnoses or age

Health insurance may have a much more pronounced impact on health for people with pre-existing poor health than for the general population (Newhouse and the Insurance Experiment Group 1993). For example, we might expect that health insurance, by increasing utilization of health care services, improves the management of blood pressure in people with hypertension and leads to reductions in blood pressure for that group. For individuals with normal blood pressure, however, we might not expect to see much change in blood pressure. Ideally, we would examine this by limiting our study sample on the basis of pre-randomization physiologic measures (looking, for example, at the reduction in blood pressure in those with high pre-randomization blood pressure and so on). We do not have pre-randomization physiologic measures, however, so we proxy for pre-existing conditions using self-reports of diagnoses made before randomization. This is an imperfect proxy; within the set of people who have the condition pre-randomization, those who are not yet diagnosed may be the ones for whom insurance has the largest impact. By limiting to those with pre-randomization diagnoses, we are capturing a population where management of the physiologic measures may be particularly important, but we are missing another population where management (beginning with diagnoses) may be even more important.

Furthermore, although we ask respondents about when they first received a diagnosis (enabling us to construct measures of pre-randomization diagnoses), these self-reports were still collected post-randomization and the possibility of differential post-hoc reporting of diagnosis by treatment and controls cannot be excluded. Appendix Table A3 examines the balance of treatment and control respondents on reports of pre-randomization diagnoses for ten conditions. Participants are considered to have a pre-randomization diagnosis if they reported having a specific diagnosis made before March 2008. The multivariate F-statistic for differences in all these conditions pooled has a p-value of 0.299; the standardized treatment effect for change in diagnosis of all these conditions is -0.003 standard deviations. This suggests that there is no differential reporting of pre-existing diagnoses by treatment or controls. In addition to observing

balance on the individual conditions, we also see no evidence of imbalance on a composite measure for having a pre-randomization diagnosis of diabetes, hypertension, high cholesterol, heart attack or congestive heart failure (estimated average difference is -0.003 (standard error =0.009; p value is 0.77). We use this composite measure to identify a subset of our population that is at increased risk of adverse cardiovascular outcomes. As noted above, this subset does not include those also at increased risk who have not been diagnosed pre-randomization because we have no way to identify them.

The middle panel of Table 4 presents our analysis of physiologic measures limited to this subset of individuals with prior diagnoses of these conditions (about a quarter of our sample). We use this composite measure, including diabetes, hypertension, high cholesterol, heart attack and congestive heart failure, rather than measure-specific diagnoses, in order to have a reasonable sample size in the limited group, and because of the patterns of comorbidity. Individuals with any of these conditions are at increased risk of adverse cardiovascular outcomes and would particularly benefit from care and management. Appendix Table A5 presents additional analysis of the health measures limited to only those with the related pre-randomization diagnosis (considering the blood pressure outcomes in those with pre-randomization diagnosis of hypertension and so on).

We also consider an alternate proxy for poor health based on age. The right-hand panel of Table 4 presents our analysis of physiologic measures limited to the quarter of our sample aged 50-64. There is substantial overlap between those ages 50-64 and those with pre-randomization diagnoses; about half of those aged 50-64 also have a pre-randomization diagnosis and about half of those with a pre-randomization diagnosis are aged 50-64. We expect that the response of physiologic measures to health insurance may be more pronounced in both of these higher-risk groups than in the overall study population.

Framingham risk score

We combine some of our physiologic measures on health with information on gender, age, smoking status and blood pressure medication to calculate the predicted 10-year risk of cardiovascular disease (CVD) using Framingham risk scores (D'Agostino et al. 2008). Table 5 reports the results for each of the components of the Framingham risk score as well as the predicted 10-year risk of CVD. We interpret the Framingham risk score as a standard clinical measure of cardiovascular health and a way of summarizing across multiple measures. For several reasons, we caution against interpreting these results as a change in the probability of CVD in the next 10 years. First, our measurement of Framingham risk score captures only one point in time and any effects of insurance that we observe may not be sustained. Second, the panel relationship between the components of the Framingham risk score and CVD in the Framingham population may not be the same as the experimental relationship in our study population. We therefore interpret the results as evidence of the impact on a useful summary measure of health rather than predictions of 10-year cardiovascular outcomes.

Self-reported health measures

For self-reported health (Table 6) we consider (1) whether subjects report good, very good or excellent health; (2) whether they report fair, good, very good or excellent health; (3) whether their health status has gotten worse over the last twelve months; (4) the physical component summary from the Medical Outcomes Survey Short Form (SF-8); (5) the mental component summary from the SF-8; (6) whether they screen positive for depression; (7) their predicted probability of depression; and (8) whether they report more than very mild pain. As with the physiologic measures of health, we present this analysis for the full sample, those with pre-randomization "high-risk" diagnoses, and those aged 50-64.

Combining health measures, diagnoses, and medications

Where there are medications that can effectively treat a clinical condition related to a physiologic measure (for example, cholesterol-lowering medication for high cholesterol), we combine the health outcome measure with information on diagnoses and medications to categorize respondents with respect to conditions and to create two measures. An abnormal result on a physiologic or self-report measure can be interpreted as reflecting that the individual has the condition (or disease). The two additional measures are:

(1) Undiagnosed condition: an abnormal result on a health measure without a corresponding respondent-reported diagnosis;

(2) Unmedicated condition: an abnormal result on a health measure without a corresponding medication.

We hypothesize that these measures may respond to insurance more than the measured health outcomes themselves, since controlling these diseases is difficult (particularly over a relatively short time horizon) even with the existence of effective medications.

In Table 7, we consider whether the conditions of hypertension, high cholesterol, diabetes, or depression have been diagnosed and whether they are being treated by medication. We consider a person to have an *undiagnosed* condition if the person's self-report or physiologic measures indicate the person has the condition but the person does not report ever having been diagnosed. We consider a person to have an *unmedicated* condition if the person's self-report or physiologic measures indicate the person has the condition but the person is not currently taking condition-specific medications based on the detailed catalogue of current medications.

There are of course treatments for these conditions other than medications, and the clinically effective management of these conditions is far more complicated than this analysis implies. The prototype clinical effectiveness measures are intended to assess the impact of insurance at a general level, and we defer a more complete examination of the management of these diseases—including other treatments and classes of medications—to future analysis.

In Panel C of Table 7, we consider two distinct measures of unmedicated pain: whether the respondent has pain untreated by any prescription pain medication and whether the respondent has pain untreated by any pain medication (prescription or over-the-counter). We

make this distinction because for pain medications in particular the same medication may be available both by prescription and over-the-counter (for example, ibuprofen). Because OHP Standard does not require co-payments for prescription medications, there may be some switching from over-the-counter versions to prescription medications without any change the prevalence of unmedicated pain.

Happiness

We hypothesize that health insurance may increase overall wellbeing. Table 8 reports results for whether the individual is very or pretty happy (as opposed to not too happy). Of course, this question is asked as part of our health-focused survey, so responses are likely influenced by participants' feelings about their health and health care.

Health care utilization

General health care utilization

We hypothesize that health insurance increases health care utilization, although there may be some offsetting effects if care is used in more efficient ways or if health improves enough. Table 9 reports results for our measures of health care utilization; we consider (1) prescription drugs, (2) office visits, (3) outpatient surgery, (4) emergency department visits, and (5) hospital visits. Prescription drugs are measured as those currently taken (collected as a detailed catalog of actual medications) and all other components of utilization are measured as reported use over the past 12 months. We examine both the extensive margin (Panel A) and total utilization (Panel B).

For health care utilization, there is a meaningful way to combine different types of use: their respective costs. In the final row of Table 9, we do a back-of-the-envelope calculation of the mean and change in annual spending using cost data from the Medical Expenditures Panel Survey to convert utilization to spending.

Preventive health care and screening

We hypothesize that health insurance increases use of preventive care and screening tests. Table 10 reports the results for use of preventive care in the last 12 months; we consider (1) cholesterol testing, (2) blood stool testing, (3) colonoscopy, (4) flu shot, (5) pap smear, (6) mammogram, and (7) PSA testing. Except for cholesterol and pap smear, all analysis is limited to those aged 50 and above; analysis for pap smear and mammogram is limited to women, and analysis for PSA testing is limited to men.

Access and quality

We hypothesize that health insurance may increase perceived access to and quality of medical care. Table 11, Panel A reports results for our measures of access to medical care; we consider (1) whether you have a usual place of clinic-based care, and whether the person got all needed (2) medical care, (3) mental health care, (4) substance abuse care, and (5) prescription

medications. Table 11, Panel B reports results for perceived quality of care, conditional on receiving care in the last 12 months (which 78% of our controls did).

Finances

We hypothesize that health insurance may reduce financial strain. Table 12 reports results for our measures of financial strain, all of which have a 12 month look back period except medical debt; we consider (1) whether one has any out-of-pocket spending, (2) the (unconditional) mean dollar amount of out-of-pocket spending, (3) whether one has catastrophic medical expenditures (defined as out-of-pocket spending exceeding 30 percent of reported household income), (4) whether one has any current medical debt, and (5) whether one has skipped other bills or borrowed money to pay medical bills.

Additional analysis

For those outcomes where we have substantively or statistically significant estimates, we plan to explore the results more fully. Rather than produce the full potential set of additional tables in this analysis plan, we instead describe our planned approach more generally..

Sensitivity of results to model specification and adjustment for covariates

As our primary specification we use linear probability models even for rates of binary outcomes. Appendix Tables A6-A14 report the results for Tables 4-12 using an alternate specification of logistic models and estimated marginal effects for all binary outcomes.

We will also investigate the sensitivity of results to adjustment for covariates. We will report our primary specification, as well as a specification either with adjustment for age and sex (where the primary specification did not include it) or without adjustment for age and sex (in the case of our blood pressure measures). We will also report a specification adding controls for a more complete set of pre-randomization characteristics.

Heterogeneity of results

We will investigate whether our treatment effects are heterogeneous along a number of dimensions: gender, age (19-49 and 50-64), race (white and any non-white), pre-randomization access to credit (yes or no), education (more than high school and high school or less), smoking status (ever smoker and never smoker), and signing up for the lottery on the first possible day. This analysis follows Finkelstein et al. and is explained in more detail there (Finkelstein et al. 2011). The measures of access to credit and sign-up date for the lottery use data sources not discussed here but are described fully in Finkelstein et al.

Comparison with prior results

We will compare our estimates, where possible, with parallel measures obtained from our mail survey conducted approximately 18 months after the lottery. Estimates may differ because

of the different mode of data collection (in-person vs. mail), the longer time horizon, or differences in respondent characteristics (the in-person data collection was conducted only for study participants in the Portland area, and also had a higher response rate). To try to shed light on the relative role of these factors, we will make several comparisons. We will replicate the prior mail survey analysis limiting to individuals in the Portland area. We will compare the mail survey results for respondents to the in-person survey with results for non-respondents to the in-person survey. We will compare in-person results for respondents to the mail survey with results for non-respondents to the mail survey. Finally, we will limit both survey analyses to the overlap set of respondents.

Comparison with observational estimates

In Table 13, we compare our results with estimates that would be obtained using simple observational comparisons of people with and without insurance in our study – ignoring the role of the lottery in randomly assigning access to insurance. We present these comparisons for the physiologic measures in Table 4; similar results for self-reported health, health utilization and finances have been previously reported (Finkelstein et al. 2011). The first comparison, done in the full sample, represents an "as treated" analysis (comparing all those covered by Medicaid to all those not covered). We then perform the same analysis in the control group only, which avoids having much of the variation in insurance coming from the lottery; in that group, most of the insurance coverage is OHP Plus, which covers a somewhat different population than OHP Standard. Last, we perform the analysis within the treatment group, comparing those on OHP Standard to those with no Medicaid (and dropping the small percentage of treatment individuals on Plus). Unlike our primary estimates, which use the lottery as a source of random variation in insurance coverage, these observational estimates will be confounded by factors that impact both likelihood of insurance coverage and our outcomes.

Future analysis

This analysis plan is intended to pre-specify an initial set of analyses from the in-person data. There are, however, entire sections of the in-person interview and additional measures that are not included in this plan. These include, for example, rich detail on medication usage and specific conditions (diagnosis and treatment of asthma, diabetes, hypertension, etc.), health behaviors, and labor force participation and work impairment. In addition, there are several physiologic health measures that we did not include here (pulse, C-reactive protein levels, and waist circumference). We intend to develop analysis plans and complete analyses of these outcomes subsequent to this initial set of analyses; our general analytic approach will be similar to that described here.

Figures and tables

Figure 1: Enrollment, treatment assignment, sampling and survey response

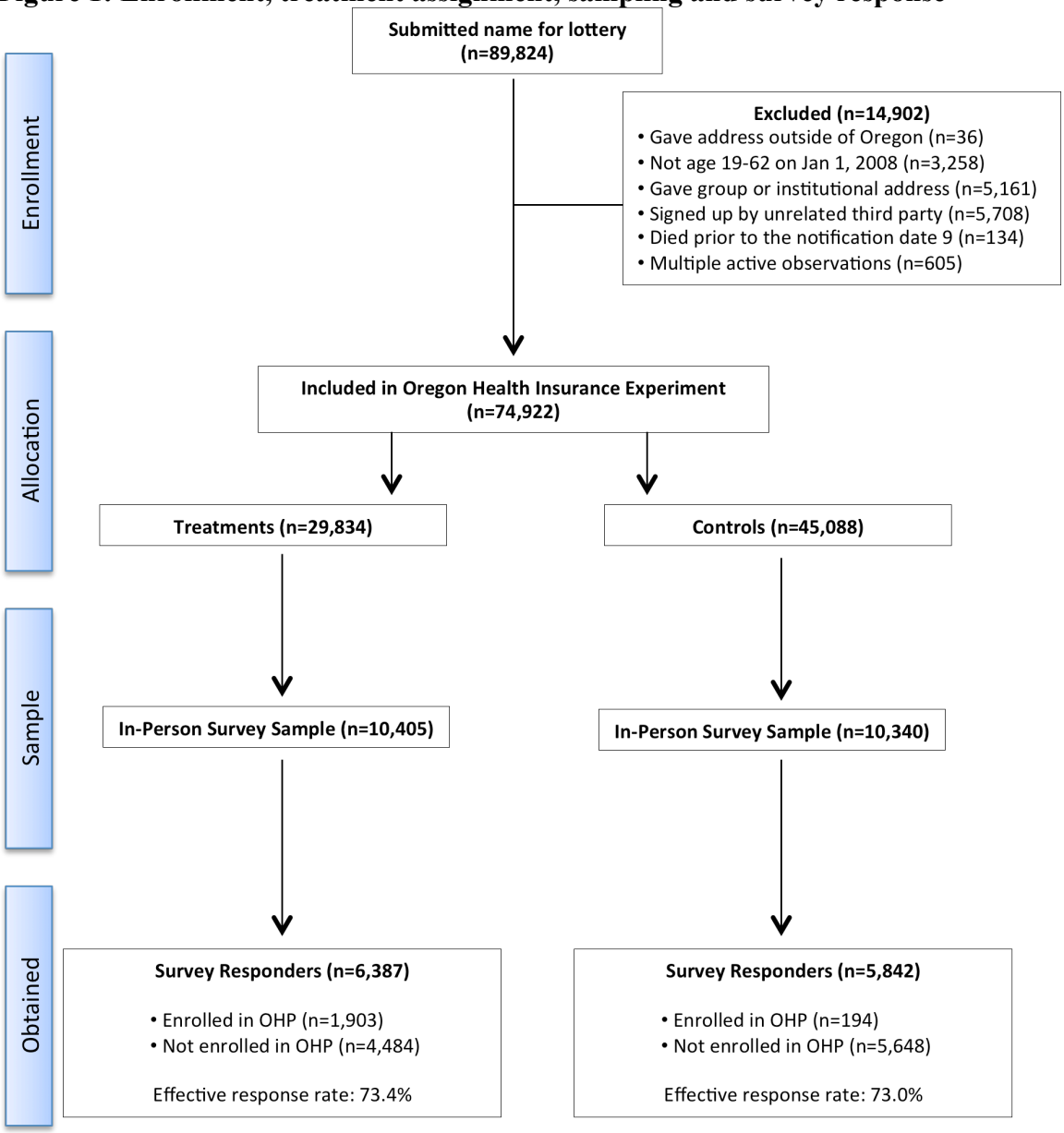


Table 1: Characteristics of the study population

	Mean for controls	Difference between treatment and controls
	(1)	(2)
Female	0.569	-0.0046 (0.0087) [0.597]
Age 19-34	0.360	-0.009 (0.01) [0.382]
Age 35-49	0.364	0.0016 (0.01) [0.873]
Age 50-64	0.276	0.0073 (0.0094) [0.434]
White*	0.688	0.0042 (0.01) [0.68]
Black*	0.105	0.0014 (0.0061) [0.824]
Other race*	0.148	0.00034 (0.008) [0.966]
Hispanic	0.172	-0.0019 (0.0084) [0.818]
Interview conducted in English	0.882	0.0025 (0.0076) [0.737]

Notes: The first column reports the weighted mean for the control respondents. The second column reports the difference between the average outcome for all individuals selected in the lottery and the average outcome for all control individuals, as calculated by ordinary least squares regression; the dependent variable is given in the left hand column. All regressions include indicators for each household size and all standard errors are clustered on household. We report the coefficient, standard error (in parentheses), and per comparison p-value [in square brackets]. All analysis is weighted using survey weights.

Sample consists of survey respondents (N = 12,229).

*Note that people were able to give more than one answer for race.

Table 2: Balance of response rates

	Mean for controls	Difference between treatment and controls
	(1)	(2)
Responded to survey	0.73	0.0028 (0.016) [0.856]

Notes: The first column reports the mean and standard deviation for the control sample. The second column reports the difference between the average response rate for all individuals selected in the lottery to the average response rate for all control individuals, as calculated by ordinary least squares regression. All regressions include indicators for each household size and all standard errors are clustered on household. We report the coefficient, standard error, and per comparison p-value. All analysis is weighted using survey weights. Sample consists of survey sample (N = 20745).

Table 3: Insurance coverage

	Control mean	Estimated FS
Ever on Medicaid during study period	0.184	0.241
		(0.0090)
		[<0.0001]

Notes: The first column reports the weighted control mean for the measure of "INSURANCE" defined in the left-hand column; The second column reports the effect on insurance coverage, which compares the average of the insurance measure for all individuals selected in the lottery to the average of the insurance measure for all control individuals, as calculated by ordinary least squares regression. All regressions include indicators for each household size and are weighted using survey weights. All standard errors are clustered on the household.
The insurance measure is taken from the Medicaid enrollment administrative data. The study period is defined as running from March 10, 2008 to the date of survey response; the variable defined as "ever" covers this entire period.

Sample consists of survey respondents (N = 12,229).

(Standard errors in parentheses)
[p-values in square brackets]

Table 4: Physiologic measures of health

	All survey respondents				Pre-existing "high-risk" diagnoses				Ages 50 - 64			
	Control mean (1)	ITT (2)	LATE (3)	p-value (4)	Control mean (5)	ITT (6)	LATE (7)	p-value (8)	Control mean (9)	ITT (10)	LATE (11)	p-value (12)
Body mass index												
BMI	29.822				32.317				30.42			
	(7.581)				(8.129)				(7.759)			
Overweight/obese	0.713				0.830				0.749			
Obese	0.415				0.558				0.448			
Blood pressure												
Systolic BP	119.278				126.142				127.303			
	(16.852)				(19.567)				(19.371)			
Diastolic BP	76.013				80.596				79.557			
	(12.136)				(13.092)				(12.812)			
Prehypertension or hypertension	0.493				0.660				0.667			
Hypertension	0.163				0.293				0.283			
Cholesterol												
Total cholesterol	198.529				200.263				201.438			
	(32.036)				(34.733)				(34.52)			
Elevated total chol	0.444				0.472				0.48			
High total chol	0.102				0.13				0.139			
HDL cholesterol	53.565				52.511				54.174			
	(14.934)				(15.27)				(16.07)			
Low HDL chol	0.184				0.212				0.187			

(continued on next page)

Table 4: Physiologic measures of health (continued)

	All survey respondents				Pre-existing "high-risk" diagnoses				Ages 50 - 64			
	Control mean (1)	ITT (2)	LATE (3)	p-value (4)	Control mean (5)	ITT (6)	LATE (7)	p-value (8)	Control mean (9)	ITT (10)	LATE (11)	p-value (12)
Glycosylated hemoglobin												
Hemoglobin Alc	5.575				6.138				5.899			
	(1.013)				(1.52)				(1.202)			
Pre-diabetic or diabetic	0.252				0.448				0.414			
Diabetic	0.080				0.218				0.147			

Notes: Column 1 reports the weighted mean of the dependent variable in the control sample of survey respondents and standard deviation for continuous outcomes. Column 2 reports intention-to-treat estimates, which compare the average outcome for all individuals selected in the lottery to the average outcome for all control individuals, as calculated by ordinary least squares regression. Column 3 reports the local-average-treatment-effect for insurance coverage as estimated by instrumental variable regression. Column 4 reports the per-comparison p value. All regressions include indicators for each household size, and all standard errors are clustered on the household. All analysis is weighted using survey weights.

For the blood pressure measures, all regressions also include controls for age (in decile bins) and sex.

Columns 1-4 show the full sample of survey respondents (N=12,229). Columns 5-8 limit the sample to those with pre-randomization diagnoses putting them at higher risk for adverse events (diabetes, hypertension, high cholesterol, heart attack, or congestive heart failure) (N=3,314); see text for more details. Columns 9-12 limit the sample to those aged 50-64 (N=3,372).

(Standard errors in parentheses)

Table 5: Framingham risk score

	All survey respondents				Pre-existing "high-risk" diagnoses				Ages 50 - 64			
	Control mean	ITT	LATE	p-value	Control mean	ITT	LATE	p-value	Control mean	ITT	LATE	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	0.569				0.562				0.558			
	(0.496)				(0.496)				(0.498)			
Age	40.723				47.978				55.395			
	(11.688)				(10.042)				(3.937)			
Total cholesterol	198.529				200.263				201.438			
	(32.036)				(34.733)				(34.52)			
HDL cholesterol	53.565				52.511				54.174			
	(14.934)				(15.27)				(16.07)			
Systolic BP	119.278				126.142				127.303			
	(16.852)				(19.567)				(19.371)			
BP medication	0.111				0.335				0.250			
	(0.314)				(0.472)				(0.433)			
Smoker	0.428				0.438				0.438			
	(0.495)				(0.496)				(0.496)			
Diabetic	0.080				0.218				0.147			
	(0.272)				(0.413)				(0.354)			
Framingham Risk Score	0.066				0.090				0.111			
	(0.06)				(0.071)				(0.071)			

Notes: See Table 4 notes. Columns 1-4 show the full sample of survey respondents (N=12,229). Columns 5-8 limit the sample to those with pre-randomization diagnoses putting them at higher risk for adverse events (diabetes, hypertension, high cholesterol, heart attack, or congestive heart failure) (N=3,314); see text for more details. Columns 9-12 limit the sample to those aged 50-64 (N=3,372).

(Standard errors in parentheses)

Table 6: Self-reported measures of health

	All survey respondents				Pre-existing "high-risk" diagnoses				Ages 50 - 64			
	Control mean (1)	ITT (2)	LATE (3)	p-value (4)	Control mean (5)	ITT (6)	LATE (7)	p-value (8)	Control mean (9)	ITT (10)	LATE (11)	p-value (12)
Health g/vg/e	0.596				0.417				0.492			
Health not poor or very poor	0.858				0.743				0.785			
Health same or gotten better	0.804				0.749				0.747			
SF-8 physical subscale score	45.492 (10.495)				41.048 (10.867)				41.402 (11.099)			
SF-8 mental subscale score	44.387 (11.38)				42.285 (11.827)				43.184 (11.62)			
Pos depression screen*	0.30				0.419				0.391			
Probability of depression*	0.257 (0.337)				0.354 (0.369)				0.333 (0.369)			
No or very mild pain	0.564				0.398				0.414			
STE												

Notes: See Table 4 notes. Columns 1-4 show the full sample of survey respondents (N=12,229). Columns 5-8 limit the sample to those with pre-randomization diagnoses putting them at higher risk for adverse events (diabetes, hypertension, high cholesterol, heart attack, or congestive heart failure) (N=3,314); see text for more details. Columns 9-12 limit the sample to those aged 50-64 (N=3,372).

*For the standardized treatment effect, the sign of the positive depression screen and the probability of depression effect is reversed so that a positive effect corresponds to better health as it does for the rest of the measures.

(Standard errors in parentheses)

Table 7: Health, diagnoses and medications

	All survey respondents				Pre-existing "high-risk" diagnoses				Ages 50 - 64			
	Control mean (1)	ITT (2)	LATE (3)	p-value (4)	Control mean (5)	ITT (6)	LATE (7)	p-value (8)	Control mean (9)	ITT (10)	LATE (11)	p-value (12)
Panel A: Undiagnosed conditions												
Undiagnosed hypertension	0.072				0.026				0.102			
Undiagnosed high chl	0.070				0.042				0.079			
Undiagnosed diabetes	0.023				0.026				0.037			
Undiagnosed depression	0.091				0.099				0.120			
STE												
Panel B: Unmedicated conditions												
Unmedicated hypertension	0.123				0.175				0.187			
Unmedicated high chl	0.094				0.109				0.118			
Unmedicated diabetes	0.032				0.055				0.049			
Unmedicated depression	0.211				0.252				0.248			
STE												

(continued on next page)

Table 7: Health, diagnoses and medications (continued)

	All survey respondents				Pre-existing "high-risk" diagnoses				Ages 50 - 64			
	Control mean (1)	ITT (2)	LATE (3)	p-value (4)	Control mean (5)	ITT (6)	LATE (7)	p-value (8)	Control mean (9)	ITT (10)	LATE (11)	p-value (12)
Panel C: Unmedicated pain												
Pain without Rx med	0.296				0.372				0.364			
Pain without any med	0.188				0.225				0.220			

Notes: See Table 4 notes. Columns 1-4 show the full sample of survey respondents (N=12,229). Columns 5-8 limit the sample to those with pre-randomization diagnoses putting them at higher risk for adverse events (diabetes, hypertension, high cholesterol, heart attack, or congestive heart failure) (N=3,314); see text for more details. Columns 9-12 limit the sample to those aged 50-64 (N=3,372).

Undiagnosed conditions are defined as having a health measure indicating the presence condition without a corresponding respondent-reported diagnosis. Unmedicated conditions are defined as having a health measure indicating the presence of the condition without a corresponding medication.

Table 8: Happiness

	Control mean (1)	Reduced form (2)	LATE (3)	p-value (4)
Very happy or pretty happy (vs. not too happy)	0.749			

Notes: See Table 4 notes. Sample is all survey respondents (N=12,229)

(Standard errors in parentheses)

Table 9: Health care utilization

	Control mean (1)	ITT (2)	LATE (3)	p-value (4)
Panel A: Extensive margin (any)				
Rx drugs (currently taking)	0.539			
Office Visits (last 12 months)	0.646			
Outpatient Surgery (last 12 months)	0.078			
ED visits (last 12 months)	0.402			
Hospital visits (last 12 months)	0.127			
Panel B: Total margin (number), unconditional				
Rx drugs (currently taking)	1.832			
	(2.807)			
Office Visits (last 12 months)	6.745			
	(21.476)			
Outpatient Surgery (last 12 months)	0.105			
	(0.435)			
ED visits (last 12 month)	1.027			
	(2.182)			
Hospital visits (last 12 months)	0.29			
	(3.119)			
Annual spending				

Notes: See Table 4 notes. Sample is all survey respondents (N = 12, 229).

(Standard errors in parentheses)

Table 10: Preventive care

	Control mean	Reduced form	LATE	p-value
	(1)	(2)	(3)	(4)
Cholesterol checked	0.272			
Blood stool test (age >=50)	0.191			
Colonoscopy (age >=50)	0.104			
Flu shot (age >=50)	0.355			
Pap smear (women)	0.449			
Mammogram (women >=50)	0.289			
PSA (men >=50)	0.214			
Standardized treatment effect				

Notes: See Table 4 notes. All measures are for preventive care in the last 12 months. Sample is all survey respondents (N=12,229), survey respondents at least 50 years of age (N=3374), female survey respondents (N=6915), female survey respondents at least 50 years of age (N=1864) or male survey respondents at least 50 years of age (N=1509), as indicated in the table.

(Standard errors in parentheses)

Table 11: Access and quality

	Control mean (1)	Reduced form (2)	LATE (3)	p-value (4)
Panel A: Access				
Have usual place of clinic-based care	0.461			
Got all needed medical care	0.610			
Got all needed mental health care	0.756			
Got all needed prescription drugs	0.724			
Standardized treatment effect				
Panel B: Quality				
Quality of care is good, v good or excellent (conditional on any)	0.784			

Notes: See Table 4 notes. Sample is all survey respondents (N=12, 229) except for quality of care which is only defined on those who received care in the last 12 months (N= 9,694). All measures of "got needed care" are over the last 12 months.

(Standard errors in parentheses)

Table 12: Finances

	Control mean (1)	Reduced form (2)	LATE (3)	p-value (4)
Any out-of-pocket spending	0.588			
Amount of out-of-pocket spending	552.839 (1219.49)			
Catastrophic expenditures	0.055			
Any medical debt	0.568			
Borrowed money or skipped bills	0.244			
Standardized treatment effect				

Notes: See Table 4 notes. Sample is all survey respondents (N=12, 229).

(Standard errors in parentheses)

Table 13: Comparison with observational estimates

	Random assignment (1)	Any Medicaid vs. No Medicaid (2)	Any Medicaid vs. No Medicaid (controls only) (3)	OHP Standard vs. No Medicaid (treatment only) (4)
Sample size				
Percent insured				
BMI				
Overweight or obese				
Obese				
Systolic BP				
Diastolic BP				
Prehypertension or hypertension				
Hypertension				
Total cholesterol				
Elevated total cholesterol				
High total cholesterol				
HDL cholesterol				
Low HDL cholesterol				
Hemoglobin Alc				
Pre-diabetic or diabetic				
Diabetic				

Notes: The top two rows report the sample size and percent insured for each sample. Column 1 reports our LATE results from Table 4 ("all survey respondents"). Columns 2 and 3 compare all those with any Medicaid coverage during our study period to those without Medicaid for the full sample and the control group. Column 4 limits to the treatment group and compares those on OHP Standard to those with no Medicaid. All regressions include indicators for each household size and are weighted using the survey weights. All standard errors are clustered on the household.

References

Allen, H., K. Baicker, A. Finkelstein, S. Taubman, B. J. Wright & Oregon Health Study Group (2010) What the Oregon health study can tell us about expanding medicaid. *Health Aff (Millwood),* 29**,** 1498-506.

Angrist, J. D., G. W. Imbens & D. B. Rubin (1996) Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association,* 91**,** 444-455.

D'Agostino, R. B., R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro & W. B. Kannel (2008) General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation,* 117**,** 743-53.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker & The Oregon Health Study Group (2011) The Oregon Health Insurance Experiment: Evidence from the First Year. *NBER Working Paper,* 17190.

Westfall, P. H. & S. S. Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: John Wiley & Sons, Inc.

Appendix

Analytic specifications

The goal of this analysis, as described in the main plan, is to evaluate the effect of lottery selection and subsequent insurance coverage on a range of outcomes. This section describes the empirical specifications.

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

We estimate the Intent-to-Treat (ITT) effect of winning the lottery (i.e. the difference between treatment and controls) by estimating the following OLS equation:

$$y_{ihj} = \beta_0 + \beta_1 LOTTERY_h + X_{ih}\beta_2 + V_{ih}\beta_3 + \varepsilon_{ihj} \quad (1)$$

where i denotes an individual, h denotes a household and $j \in J$ denotes a "domain" of related outcomes (such as health or financial strain). For example y_{ij} might be the self-reported health of individual i, which is one of the health measures in the health "domain" j.

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 larger households.[1]

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 group 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, however, as they are not related to treatment status. Our primary analysis does not control for any V_{ih} covariates; the exception to this is in the analysis of the blood pressure measures, where we add adjustment for age (in decile bins) and sex. As a secondary analysis, we will explore whether our results are sensitive to inclusion of the V_{ih} covariates.

[1] The proportion of treated individuals in household size 1 is 71.40% (78.40% for controls), in household size 2 is 28.36% (21.57% for controls) and in household size 3 is 0.25% (0.03% for controls).

In all of our ITT estimates and in our subsequent instrumental variable estimates (see below), we estimate linear models even though a number of our outcomes are binary. Because we are interested in the difference in conditional means for the treatments and controls, linear probability models pose no concerns in the absence of covariates or in fully saturated models (Angrist 2001, Angrist and Pischke 2009). Our models are not fully saturated, however, so it is possible that this functional form choice could make a difference, especially for outcomes with very low or very high mean probability. We 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 are weighted to account for the sampling design of the survey as described below.

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*. We model this as follows:

$$y_{ihj} = \pi_0 + \pi_1 INSURANCE_{ih} + X_{ih}\pi_2 + V_{ih}\pi_3 + \nu_{ihj} \quad (2)$$

where INSURANCE is a measure of insurance coverage and all other variables are as defined in equation (1). We estimate equation (3) by two stage least squares (2SLS), using the following first stage equation:

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

in which the excluded instrument is the variable *LOTTERY*.[2]

We interpret the coefficient on insurance from instrumental variable estimation of equation (3) as a local average treatment effect of insurance, or LATE (Imbens and Angrist 1994). In other words, our estimate of π_1 identifies the causal impact of insurance among the subset of individuals who obtain insurance on winning the lottery and who would not obtain insurance without winning the lottery (i.e. the compliers).[3]

---

[2] In principle, if we could observe individuals' income in the year before the lottery this could be quite predictive of the first stage and interacting pre-randomization income with the LOTTERY variable in the first stage could improve the precision of our estimates, and thus our power.

[3] When insurance is defined as "ever on OHP Standard" we can probably be comfortable interpreting the IV estimates of equation (3) as the treatment-on-treated (ToT) rather than a LATE. In practice, there are two small violations of this interpretation. First, if there were no way to get OHP Standard without winning the lottery there would be no "always-takers" in the terminology of Angrist, Imbens and Rubens (1996), but about 2 percent of our controls got onto OHP standard through some limited alternative mechanisms —for example, pregnant women who are on OHP Plus can sometimes stay on OHP Standard after giving birth. Second, it is possible that some compliers were put on OHP Plus rather than Standard, since case workers are instructed to first check applicant eligibility for

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, but it may not be strictly true. First, it is possible that the event of winning (or losing) the lottery may have direct effects on outcomes, although it seems unlikely to us that any such effects persist more than a year after the lottery. Second, individuals who apply for public health insurance may also be encouraged to apply for other public programs, such as food stamps or cash welfare. In particular, if the individual applied for OHP in person (rather than by mail), case workers were instructed to offer assistance to interested applicants in applying for TANF (cash welfare) and the Supplemental Nutrition Assistance Program (food stamps). These other transfer programs could have direct effects on the outcomes we study. This is not an idiosyncratic feature of our setting but a more general feature of the application process for public programs; as such, it may be a relevant component of the impact of attempts to expand Medicaid more generally. However, any direct impact of winning the lottery on receipt of other benefits is a violation of the exclusion restriction for the LATE interpretation of the impact of insurance per se (which is not the case in the ITT analysis estimating the effect of expanded access to Medicaid). This is explored empirically below.

Handling many outcomes

For each of the broad domains, we provide a summary measure in addition to the individual outcomes. For some domains there are natural summaries available (such as total cost for utilization or the Framingham risk score for physiological measures described below). When a natural summary measure is not available, we generate a simple one based on the average observed effect in standardized units:

$$\sum_{j \in J} \frac{1}{J} \frac{\xi_{1j}}{\sigma_j} \qquad (4)$$

where σ_j is the standard deviation of y_j in the control group and ξ_{1j} is the coefficient of interest for outcome j. (Specifically, for the ITT estimates in equation (1), the ξ_{1j} 's correspond to the β_{1j}'s.) In order to account for covariance in the estimates of ξ_{1j} / σ_j we estimate pooled OLS for all outcomes $j \in J$.[4] An important limitation of this ad hoc standardized treatment effect is that it implicitly "weights" each outcome within a domain equally.

Plus; in practice this number is likely to be small since the estimated first stage is very similar for "ever on Medicaid" (which includes Plus and Standard) and "ever on OHP Standard" (see rows 1 and 2 of Table 3).

[4] Specifically, we stacked the data for the individual outcomes within a domain and estimated a single regression equation that allowed the coefficients on each covariate to vary flexibly across the outcomes and for correlation in the error terms across outcomes. The outcomes are defined to point in the same direction (e.g. increasing means better health).

Survey weights

We use weights to adjust for several aspects of our survey fielding in all of our summary statistics and regression analyses. The weights are designed to ensure our recruitment sample (i.e. the sample of people whom we attempted to contact and who ended up with non-zero weights) is balanced on treatment status and representative of our sampling base.

We began recruitment of study participants in groups (called "sample releases"). These rolling releases facilitated efficient use of recruitment, tracking, and interviewing resources. There were 44 such releases, roughly weekly, typically of about 450 individuals. Over the course of the fielding, we routinely dropped individuals from our active recruitment sample. There were two reasons for doing so. First, to promote a high response rate, we regularly took a random subsample of the participants who had been released but had not yet responded and instructed the fielding staff to continue active recruitment only on the selected group. This allowed our staff to devote additional time and effort ("intensive follow-up") to potential participants who were difficult to locate or recruit, without diverting too many resources away from the rest of the potential participants. Second, during our fielding the state was conducting a new lottery for OHP Standard (discussed more below). Following each new lottery drawing, we excluded from our active sample individuals selected in the lottery who had not yet responded ("new lottery drops"). Between the intensive follow-up drops and the new lottery drops, the active sample was restricted roughly every 2 to 4 weeks. A typical drop removed a few hundred individuals from the active sample.

We adjust for both types of drops using weights constructed on the following principle: within any (even non-random) subset of the original sample base, a randomly selected group can be weighted to stand in for the non-selected remainder based on the probability of that random selection without introducing bias. For each of the drops, we construct a weight that corrects for that drop. The final analytic weight is simply the product of all the weights introduced over the course of the study.

Our weighting is roughly analogous to weighting done for censoring or attrition in longitudinal data (Cole and Hernán 2008, Kalton 1986). As in those settings, we weight each observation at each time point by the inverse probability of obtaining data, and we generate overall weights as the product of the weights across all time points. In our setting, the time points correspond to changes in the active sample. We do not need to model the probability of obtaining data since we randomly assigned active sample status and know the probabilities

Table A1 summarizes the distribution of the weights for the entire sampling base, the recruitment base and then survey responders. Over the entire sampling base, the weights have a mean close to 1, and there are relatively few extreme weights. For the survey respondents, which comprise the sample we analyze, the weights are even less extreme. The average weight is 1.24, with the 95^{th} to 5^{th} percentile range of 2.076 to 0.96. The controls are impacted more by the weights than the treatment group, as they were more likely to sign up for the new lottery (see below).

This weighting method allowed us to continue fielding in the face of the new lottery which affected a non-random subsample of our population and to devote resources to an intensive follow up of non-respondents, while preserving balance between treatment and control study participants. The primary cost of such weighting is an increase in variance. One way to quantify the cost in variance is to calculate an "effective sample size", which is the unweighted sample size that would have equivalent precision to our weighted sample. Our effective sample size is 4786 controls and 5406 treatments (compared to our actual sample of 5842 controls and 6387 treatments).

The following sections give more detail on the construction of the weights.

Continuous intensive follow-up of non-respondents

For each "intensive follow-up drop" we constructed weights as follows. Let N_t be the set of individuals in a specified sample release (or a group of releases) who have not yet completed an interview at time *t*. We select a random subsample F_t from N_t with sampling probability p_t=|F_t|/|N_t|. The weights are defined for each individual *i* in the sampling base as:

$$w_t(i) = \begin{cases} \frac{1}{p_t} \; if \; i \; in \; F_t \\ 0 \; if \; i \; in \; N_t \; and \; i \; not \; in \; F_t \\ 1 \; if \; i \; not \; in \; N_t \end{cases}$$

This weighting does not impact any individuals not yet released for fielding or already having completed interviews; it up-weights the subsample selected for intensive follow-up by the inverse of the probability of being selected, so that they stand in for those dropped (who are assigned weight zero).

New state lottery

Early in our fielding period, the state of Oregon began conducting a new lottery for OHP Standard. The state mailed postcards to those on the original list who 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[5] 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

[5] The matching was done using LinkPlus software using name, date of birth, social security number, address and Medicaid client ID where available.

"selected opt-ins"). By December 6, 2010 the state had selected everyone in our original sample who signed up for the new lottery; we limit our analysis to data collected by October 13, 2010 to avoid having extreme weights.

Given the complications in interpreting the "treatment" received by those who were drawn in the new lottery, we chose to drop the selected opt-ins from our recruitment sample. Additional weights are needed 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 were therefore able to use weights to adjust for the sample dropped because of the second lottery using the same principle as above: 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.

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. For those released into active fielding and having already completed an interview, the new lottery does not pose any problems. This whole set is assigned weight 1. For who have not yet completed an interview, we define the weight for individual *i* to be:

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

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*). Additionally, because of complexity in how we chose releases, the probability p_t depended on whether and when an individual was released. Thus we actually assign these weights in groups of releases where p_t was constant.[6]

Final analytic weights

Each weight variable w_t is designed to adjust for the sampling event at time *t* (whether an intensive follow-up drop or a new lottery drop). We define the cumulative weight variable W_T as the product of all w_t for $t \leq T$. Reweighting by W_T ensures that the recruitment sample is representative of the full sampling base. Whenever there is a sampling event, W_T changes appropriately: multiplication by w_t is precisely what is necessary for the recruitment sample to remain representative.

[6] Due to a technical complication in our sampling, releases 4-21 were stratified on opt-in status with different sampling probabilities for opt-ins and non-opt-ins. We use an additional set of sampling weights to correct for this. The net effect of this is small; the range of these corrective weights is 0.85-2.59.

Treatment-control balance

In previous work (Finkelstein et al. 2011), we discuss the random assignment of treatment and control groups. Here we examine treatment and control differences in the subset of the study population who completed interviews. Table A2 reports the results.

Panel A reports the balance on response rates to the survey. Preliminary results suggest that our weighted effective response rate for the controls was 73% and the treatments did not respond at a significantly different rate. In addition to this relatively high response rate, we obtained very high responses to the individual components of the study. We obtained valid anthropometric and blood pressure data on 98% of respondents and valid blood assays (total cholesterol, HDL cholesterol and Hemoglobin A1c) on 99% of respondents. Over 98% of respondents either provided medications to be catalogued or reported no medication use (although 8% indicated that this catalog was incomplete). There is no evidence of differential response rates between treatments and controls on any of these components.

Given any response rate of less than 100%, however, there is the potential for bias even if the overall response rate for treatment and control groups is the same: the controls that respond could have systematically different characteristics from the treatments that respond. In Panel B, we examine respondents' balance on characteristics that are not affected by lottery selection. Some are measured pre-randomization, taken from the information they provided when signing up for the lottery. Some are measured in the survey but are immutable, such as age or race. Others are characteristics of the data collection effort, such as response date (including season, weekend vs. weekday, etc.), response time (days between start of recruitment and completion of the interview), location of the interview, and language of the interview (English, Spanish, or interpreter of another language). All these variables are intended to help identify potential response bias by capturing characteristics of the responders that may be related to outcomes (men may differ from women, those who chose to come in on the weekend may differ from those who chose to come in during the week, and so on) but are not likely to be affected by the lottery itself. The overall F-stat for differences in all the characteristics pooled has a p-value of 0.738.

In Panel C, we test whether there was any evidence of differential sorting across our interviewers or equipment on the basis of treatment status. We do not expect that there will be differences here, as assignment to interviewer or equipment should not be related to treatment status. As such, we do not want to include all these additional tests in our global test of response bias in Panel B because it could mask real differences between respondents in the characteristics in Panel B. However, because the interviewer or equipment used has such a direct effect on the outcome measurement, we might be concerned about differences even arising from chance.[7] The F-stats for the tests on the three pieces of equipment have p-values of 0.291, 0.231, and 0.728 respectively. The F-stat for the test of sorting across our 48 interviewers has a p-value of 0.150.

[7] It is worth noting that the division between Panel B and Panel C is not completely clean. For example, interviewers and equipment were assigned to specific clinics.

As a final check of imbalance between treatment and control respondents, we examined differences in pre-randomization characteristics measured in other, administrative, datasets. We examine whether treatments and controls differed in having any hospital visits or the number of hospital visits in the pre-randomization period (as measured in hospital discharge data) or in having any medical or non-medical collections (as measured in credit report data). These datasets are described by Finkelstein et al. (2011), and this analysis follows Table A13 from that paper. There is no evidence of any difference; the F-stat for the test on these four variables combined has a p-value of 0.187 (not shown).

Although these results are not conclusive—there is still the possibility of differences on other unobserved variables—they are reassuring. To the extent that we are able to examine it, we find no evidence of differential selection into responding between treatment and control groups.

Insurance coverage

Table A4 reports the control means and effects of lottery selection for alternative definitions of insurance coverage. The increase in OHP Standard is slightly greater than the increase in any Medicaid (0.265 percentage points compared to 0.241). This suggests that at least some of the increase in OHP Standard comes from individuals who are on another Medicaid program at some point during the study period. The lottery is associated with an increase of 4.16 months on Medicaid (row 3). Using "current" enrollment (measured on the date of interview) reduces the effect on insurance coverage from 0.241 to 0.111; the increase in Medicaid coverage associated with the lottery attenuates over time as treatments fail to recertify their eligibility for OHP Standard and controls get on OHP Plus. Figure A1 shows the time path of enrollment in OHP Standard and all Medicaid over time.

Unlike the administrative data that capture only Medicaid coverage, the interview data capture all sources of insurance (including private coverage). The difference in Medicaid coverage associated with the lottery as measured in the interviews is similar to the difference in Medicaid coverage as measured in the administrative data on the same date. The increase in any insurance coverage is similar to the increase in Medicaid coverage, suggesting that the lottery had little impact on non-Medicaid insurance coverage. Specifically, we see no evidence of crowd-out of private insurance; private insurance rates are unchanged by the lottery.

Table A4 indicates that selection by the lottery is also not associated with any substantive or statistically significant change in TANF (cash welfare) receipt or benefits. However, lottery selection is associated with a statistically significant but substantively trivial increase in the probability of SNAP (food stamp) receipt (2.3 percentage points), although not in total food stamp benefits. Estimates of the income elasticity of health care use range from a low end of about 0 to a high end of about 1.5 (Getzen 2000, Table 1), suggesting that the income effect of food stamp receipt on health care use would be considerably less than 1 percent. The impact on health seems likely to be negligible as well. Thus, we are comfortable interpreting our IV

estimates as the effect of insurance coverage, or more specifically, the Oregon Medicaid program, on outcomes.

Additional analysis

Table A5 reports analysis of some of our health measures limited to only those with the related pre-randomization diagnosis. We consider the blood pressure measures limited to those with a pre-randomization diagnosis of hypertension, the cholesterol measures limited to those with a pre-randomization diagnosis of high cholesterol, the glycosylated hemoglobin measures limited to those with a pre-randomization diagnosis of diabetes, and the depression measures limited to those with a pre-randomization diagnosis of depression or anxiety.

Tables A6-A14 report analysis of our binary outcome measures using a logistic regression rather than a linear probability regression model. We present the control means and linear probability results (also shown in Tables 4-12 in the main text) as well as average marginal effects from the logistic model estimation. The average marginal effects are calculated by predicting the outcome as a treatment and as a control for each individual, taking the difference in the two predictions, and averaging those differences across the whole sample.

Outcome measures

The outcomes in this analysis are drawn from the physiological measures and in-person questionnaire (available at www.nber.org/oregon). We developed the questionnaire for this study, drawing on existing survey instruments whenever possible. Table A15 provides a summary of the outcome variables and Table A16 provides additional detail on the distribution of some variables. The outcomes fall into several broad domains.

Physiologic measures of health

Our physiologic health measurements were modeled on those done by the National Health and Nutrition Examination Survey (NHANES) and we worked with consultants from the National Center for Health Statistics to develop them.

We measured weight using a Seca 876 portable digital weight scale and height using a Seca 214 portable stadiometer. We examine continuously measured **body mass index (BMI)** defined as a function of height and weight. We define whether you are **overweight or obese** as BMI of at least 25 and whether you are **obese** as BMI of at least 30. These are standard clinical cut-points (Expert Panel on the Identification Evaluation and Treatment of Overweight and Obesity in Adults 1998, World Health Organization 2011) and are used by the NHANES in estimating the prevalence of obesity in the US population (Flegal et al. 2010)

We measured blood pressure using the OMRON IntelliSense unit, model HEM-907XL, which automatically inflates the cuff to the desired level and does not require adjustment by the

interviewer. Our blood pressure measure is the average of three readings taken 30 seconds apart, following a period of sitting quietly for 5 minutes. We examine continuously measured **systolic blood pressure** and continuously measured **diastolic blood pressure**. We define **pre-hypertension or hypertension** using the standard clinical cut-points of systolic blood pressure of at least 120 or diastolic blood pressure of at least 80 and **hypertension** as systolic blood pressure of at least 140 or diastolic blood pressure of at least 90 (Chobanian et al. 2003).

We collected up to 5 drops of whole blood from a finger stick. Samples were collected on Whatman 903 specimen-collection paper and dried and stored following established protocols (McDade, Williams and Snodgrass 2007). The University of Washington Department of Laboratory Medicine performed the assays from the stored blood. We used formulas developed by the National Center for Health Statistics to convert the dried blood spot measurements to the scales used for clinical measures.

We examine continuously measured **total cholesterol**. We define **elevated cholesterol** as total cholesterol greater than or equal to 200 mg/dL and **high cholesterol** as total cholesterol greater than or equal to 240 mg/dL (Expert Panel on Detection Evaluation And Treatment of High Blood Cholesterol In Adults 2001). We also examine continuously measured ("good") **HDL cholesterol**. We define **low HDL cholesterol** as HDL cholesterol below 40 mg/dL.

We examine continuously measured **Hemoglobin A1c**. We defined **elevated risk of diabetes** as Hemoglobin A1c of at least 5.7% and **diabetes** as Hemoglobin A1c of at least 6.5% (International Expert Committee 2009, American Diabetes Association 2010).

Framingham risk score

We use a sex-specific multivariable point-mapping system to calculate the probability of specific atherosclerotic cardiovascular disease (CVD) events, i.e., coronary heart disease, cerebrovascular disease, peripheral vascular disease, and heart failure (D'Agostino et al. 2008). This system, derived using data from the Framingham Heart Study, incorporates age, total and high-density lipoprotein (HDL) cholesterol, systolic blood pressure, treatment for hypertension, smoking, and diabetes status. For each of these variables, a number of points between -3 and 12 is allotted. Total points are then aggregated across all variables and mapped to a probabilistic risk of CVD events in the next 10 years. To calculate the CVD risk score, we used gender and age variables from survey responses and total and HDL cholesterol, systolic blood pressure and diabetes as described in *physiologic measures of health* above. A person is considered "treated for hypertension" if one or more medication from her medication survey was classified as a hypertension medication (more detail in *undiagnosed and unmedicated conditions* below). Individuals answering "yes" to the survey question "Are you currently smoking?" are considered smokers.

Self-reported health

We use a question about self-reported health to construct two binary measures: self-reported **health good, very good or excellent** (vs. fair, poor, or very poor) and self-reported **health fair, good, very good, or excellent** (vs. poor or very poor). These differ in the handling of the **26** percent of participants reporting fair health. The next measure is whether **health status has gotten worse over last twelve months** (vs. stayed the same or gotten worse).

Our survey included the Medical Outcomes Survey Short Form (SF-8) and we examine both the **SF-8 physical component summary** and **SF-8 mental component summary**. The SF-8 is a short form (8-item version) of the Medical Outcomes Survey designed to measure health-related quality of life (Ware et al. 2001). The eight questions ask about general health, work, physical and social limitations, pain, energy levels and emotional problems. Each response is assigned a score, and the physical component summary and mental component summary are both sums of those scores using different weightings. The scoring is designed so that the summary scores will be comparable to scores obtained using the validated SF-36 (McHorney, Ware and Raczek 1993). The scores range from 0 to 100, with higher scores indicating better health, and are normalized to a mean of 50 and standard deviation of 10 in a general population sample.

The next two measures are based on the 8-question version of the Patient Health Questionnaire (PHQ-8). The Patient Health Questionnaire is a standard scale for measuring depression (Kroenke et al. 2009) and is used for measuring depression prevalence in the US population in both the NHANES (Shim et al. 2011) the BRFSS (Kroenke et al. 2009). We include both your **probability of depression** and whether you **screened positive for depression**. The PHQ-8 asks about the frequency of eight depression symptoms. The summary score is calculated by assigning a score of 0 – 3 for each question of the questionnaire (0 for not at all; 3 for nearly every day) and then summing those scores, so higher scores indicate more severe depression symptoms. We map the score to a probability of depression (Kroenke and Spitzer 2002). The positive depression screen is based on a cut-point of PHQ-8 summary score of 10 or above. Using a cut-point of 10 or above for depression in a 9-question version of the PHQ has been shown to correlate highly with clinician diagnosis of major depressive disorder (Kroenke, Spitzer and Williams 2001). The PHQ-8 is a modified version of the 9-question version differing only in excluding the question about suicidal ideation (which is rarely answered in the affirmative, and thus makes little substantive difference in scores (Huang et al. 2006)).

Finally, we use one of the SF-8 questions separately to capture whether you **had no or only very mild pain** in the past 4 weeks.

Undiagnosed and unmedicated conditions

In addition to direct measurement of the presence of several health conditions, we also examine the prevalence of unmedicated or undiagnosed health conditions. We construct these variables using measures of hypertension, high cholesterol, diabetes, depression and pain measures (derived from the combination of survey and physiologic health measures, as described

above) paired with survey data on respondent-reported diagnoses and our cataloguing of current medication usage, described in more detail below.

For "undiagnosed" measure, we consider an individual to have **undiagnosed hypertension** if our blood pressure reading meets the clinical criteria for hypertension, yet the individual reports never having been diagnosed with hypertension. We consider an individual to have **undiagnosed high cholesterol** if our dried blood spot cholesterol reading meets the clinical criteria for high cholesterol, yet the individual reports never having been diagnosed with high cholesterol. We consider an individual to have **undiagnosed diabetes** if our dried blood spot Hemoglobin A1c reading meets the clinical criteria for diabetes, yet the individual reports never having been diagnosed diabetes. We consider an individual to have **undiagnosed depression** if the individual has a PHQ-8 score of 10 or above, yet the individual reports never having been diagnosed with depression.

For "unmedicated" measures, we use information on medications for hypertension, high cholesterol, diabetes, depression, and pain from our medication cataloging. If an individual reported any medication use in the last 4 weeks, we took a detailed inventory of the actual medications. Participants were asked to bring all current medications (prescription and over-the-counter) to their interview. The interviewers entered information (including medication name, dosage, frequency and route) on each medication through an interface that looked up records in a drug database obtained from First DataBank. This drug database codes medications into classes, with drugs with multiple uses having multiple class codes. We use these classes to identify indications with input from a physician. For example, we considered anyone taking a medication classified as an antidepressant to be taking medication for depression (even though that drug may have been prescribed for a different indication). Table A17 lists the names of medications included in each group. For each medication, interviewers asked whether the medication was prescribed or over the counter. If the participant had not brought all current medications to the interview, a phone follow-up was attempted to obtain any remaining medications. Of the 68% of participants who reported any medication use, 11% said that did not provide all medications at the interview and did not complete a phone follow-up.

We consider an individual to have **unmedicated hypertension** if our blood pressure reading meets the clinical criteria for hypertension, yet the individual does not report anti-hypertensive medication. We consider an individual to have **unmedicated high cholesterol** if our dried blood spot cholesterol assay meets the clinical criteria for high cholesterol, yet the individual does not report cholesterol medication. We consider an individual to have **unmedicated diabetes** if our dried blood spot Hemoglobin A1c reading meets the clinical criteria for diabetes, yet the individual does not report diabetes medication. We consider an individual to have **unmedicated depression** if the individual has a PHQ-8 score of 10 or above, yet the individual does not report antidepressant medication. We note that this is a very restrictive definition of treatment based only on medication; future work will explore disease management more broadly.

We consider two measures of unmedicated pain. We consider an individual to have **pain without prescription pain medication** if the individual reports more than mild pain, yet the

individual does not report using prescription pain medication. We consider an individual to have **pain without any pain medication** if the individual reports more than mild pain, yet the individual does not report using any pain medication (prescription or over-the-counter). For both we use analgesics as pain medications, and do not consider other classes of medications that are sometimes used in the context of pain (e.g. corticosteroids).

Happiness

We also asked about how individuals were feeling in general, and we construct a measure of being "**very happy" or "pretty happy"** as compared too "not so happy".

Health care use

We consider both the extensive and the total margins of five utilization categories: prescription drugs, doctor's office visits, outpatient surgery, emergency department visits, and hospital visits. Our survey module on utilization was based on the Health and Retirement Survey questionnaires (Health and Retirement Study 2000). We asked about each kind of health care visit separately. In cases where the participant could not give a close estimate of how many visits, we asked for the best guess between zero, one, and more than one visit. If the answer to the probe was "more than once," we coded it as if the individual had 2 visits in the last 12 months. Less that 0.2% of answers were imputed from probes for each variable. We truncated each of the number of visits measures (office, outpatient surgery, emergency room, and hospital) at $2*99^{th}$ percentile, recoding outliers to missing. The cutpoints for truncation and percent truncated are shown in Table A16.

If an individual reported any medication use in the last 4 weeks, we took a detailed inventory of the actual medications (as described in the *undiagnosed and unmedicated conditions* section above). For **any current prescription drugs** and **number of current prescription drugs**, we counted up all medication records that could be identified as prescription drugs from the medication survey, after removing duplicates. We note that the number of prescription drugs is likely an underestimate because for 8% of respondents the medication catalogue was incomplete.

We define **number of office visits in the last 12 months** by the individual's response to the question "In the last 12 months, about how many times have you seen a doctor or other health care professional at a doctor's office, a clinic, or at home?" We define **any office visits in the last 12 months** as whether the individual provided a positive answer to this question.

We define **number of outpatient surgery visits in the last 12 months** by the individual's response to the question "In the last 12 months, how many times have you had outpatient surgery?" We define **any outpatient surgery in the last 12 months** as whether the individual provided a positive answer to this question. Almost everyone who reports an outpatient surgery visit also reports an office visit, as we would expect since such surgery would likely require associated office visits (for diagnosis, pre-operative consultation or post-operative follow-up).

We define **number of ED visits in the last 12 months** by the individual's response to the question "In the last 12 months, about how many times have you gone to an emergency room or urgent care clinic?" We define **any ED visits in the last 12 months** as whether the individual provided a positive answer to this question.

We define **number of hospital visits in the last 12 months** by the individual's response to the question "In the last 12 months, how many times have you had to stay in a hospital at least overnight?" We define **any hospital visits in the last 12 months** as whether the individual provided a positive answer to this question.

Annual spending estimation

To calculate the implied annual spending effects associated with the estimated utilization effects we use data from the 2002-2007 (pooled) Medical Expenditure Panel Survey (MEPS) on expenditures of all nonelderly (19-64) adults below 100 percent of poverty who are publicly insured. This gives us a total sample of over 7,500 individuals. We use their expenditures (all inflated with the CPI-U to 2007 dollars) to calculate average expenditures per outpatient visit (including doctor visits, outpatient surgery, and outpatient visits to any other facilities), average expenditures per ED visit, average expenditures per inpatient visit (for visits not related to childbirth). For medications, we calculate average spending per prescription drug by dividing total annual prescription drug costs by the total number of prescription drugs taken over the course of the year. All spending numbers are bases on total expenditures (i.e. not just expenditures in the insured or insurance expenditures). The underlying costs are $150 per outpatient visit, $435 per ED visit, $7,523 per inpatient visit, and $312 expenditure per prescription drug. For each type of utilization we observe (doctor visit, outpatient surgery, ED visit, inpatient visit and prescription drug), we multiply the estimated change in number by the cost per visit estimated in the MEPS. For both doctor visits and outpatient surgery, we use the $150 outpatient visit estimate which is calculated across all outpatient visits (including doctor office visits, outpatient surgery and other outpatient visits).

Preventive care and screening

Our module on preventive care used questions from the BRFSS (Centers for Disease Control and Prevention (CDC) 2000) for blood stool tests, colonoscopy, pap smear, mammogram and PSA tests. It also included questions from the NHANES (Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS) 2000) on cholesterol checks. We only asked individuals about their use of preventive care if they reported having used medical care in the past 12 months. This means we do not know about use of preventive care prior to the past 12 months in individuals who did not report medical care in the past 12 months. To avoid potential bias from these missing data, we have adopted a 12-month time frame for all types of preventive care, even when the recommended interval is longer.

We consider an individual as having had a **cholesterol check in the last 12 months** (1) if the individual answered "yes" to the survey question "Has a doctor or other health professional ever told you that you had high cholesterol?" and "within the last year" to the survey question "When were you first told that you had high cholesterol?", or (2) for individuals who have not been diagnosed with high cholesterol, if the individual answered "within the last year" when asked "long has it been since you last had your cholesterol checked?" Cholesterol testing is recommended every 5 years starting at age 20 (Expert Panel on Detection Evaluation And Treatment of High Blood Cholesterol In Adults 2001).

We consider an individual as having had a **blood stool test in the last 12 months** if the individual answered "Yes" to the survey question "In the last 12 months, has a doctor asked you to do a blood stool test?" We consider an individual as having had a **colonoscopy in the last 12 months** if the individual answered "Yes" to the survey question "In the last 12 months, have you had a sigmoidoscopy or a colonoscopy?" We do not look at blood stool tests or colonoscopies for individuals younger than 50. The U.S. Preventive Services Task Force recommends screening using blood stool test, colonoscopy, or sigmoidoscopy for colorectal cancer in all adults beginning at age 50 years and continuing until age 75 years. The Task Force recommended annual screening with high-sensitivity blood stool test, or sigmoidoscopy every 5 years coupled with high-sensitivity blood stool test every 3 years, or screening colonoscopy every 10 years (U.S. Preventive Services Task Force 2008).

We consider an individual as having had a **flu shot in the last 12 months** if the individual answered "Yes" to the survey question "Have you had a flu shot in the last 12 months?" We do not look at flu shots for individuals younger than 50. Although the Advisory Committee on Immunization Practices (ACIP) in the Center for Disease Control (CDC) recommends annual flu shots for everyone older than 6-months, it also recommend priorities to be given to young children and those 50 or older in cases of limited supply (Fiore et al. 2010).

We consider a woman as having had a **pap smear in the last 12 months** if she answered "Yes" to the survey question "In the last 12 months, have you had a Pap test or Pap smear?" This variable is not applicable to men. The U.S. Preventive Services Task Force recommends initial screening for cervical cancer with Pap smear or liquid-based cytology starting by age 21 years or approximately 3 years after the first sexual intercourse. Future screenings should occur every year with a traditional Pap smear or every 2 years with liquid-based cytology. At or after age 30 years and with three normal test results, intervals can be decreased to every 2 to 3 years with traditional Pap smear or every 3 years with HPV assay testing plus cervical cytology (U.S. Preventive Services Task Force 2003).

We consider a woman as **having had a mammogram in the last 12 months** if she answered "Yes" to the survey question "In the last 12 months, have you had a Mammogram?" This variable is not applicable to men, and we limit to women 50 or older. According to updated guidelines in 2009 from the U.S. Preventive Services Task Force, biennial screening mammography is recommended for women aged 50 to 74 years (U.S. Preventive Services Task Force 2009).

We consider a man **having had a PSA test in the last 12 months** if he answered "Yes" to the survey question "In the last 12 months, have you had a blood test to check for prostate cancer?" This variable is not applicable to women, and we limit to men 50 or older. PSA screening may not be beneficial, and the US Preventive Task Force recently circulated draft recommendations against such screening (Chou et al. 2011). It is, however, quite common, with 54% of the U.S. men aged 50-64 reporting have received a test in the last year,[8] and as of 2009, American Urological Association and American Cancer Society have recommended that early detection begin at age 50 years for men at average risk of prostate cancer (Greene et al. 2009). We include it because access to health insurance may increase use of commonly used tests, even if those tests are of limited value.

Access and quality

We considered a number of questions on individuals' access to health care. We asked if they **had a usual place of clinic care**. We exclude emergency rooms but include all doctors' offices in a hospital, a private clinic, or a community health center. We also asked if individuals needed medical care in the last 12 months, and if so, whether they received all needed care. These questions focused on care for a physical illness, injury, or condition and excluded dental care or routine vision services. We consider people having **gotten all needed medical care in the last 12 months** if they reported needing care and receiving all needed care or if they reported not needing care. We asked separate questions about whether an individual **got all needed mental health care** and **got all needed prescription drugs.** For medical care, 23% of controls reported not needing any care; for mental health care 60% reported not needing any care and for prescription drugs 32% reported not needing any care.

We asked individuals to rate the **quality of care they received in the past 12 months,** conditional on receiving care, and analyzed if it was **good, very good, or excellent**, vs. **fair or not so good**. This measure of quality of care is defined for the 78% of participants who reported receiving any medical care (include office visits, outpatient surgery, emergency room visits, hospital stays, and other care).

Financial strain from health care costs

We considered several measures of financial strain based on out-of-pocket spending, medical debt, and borrowing money or skipping paying bills because of medical debt.

Our module on health care use and costs was based on the Health and Retirement Survey questionnaires (Health and Retirement Study 2000). We asked survey participants about out-of-pocket spending in the last year for their own doctor visits, ED visits, outpatient surgeries,

[8] Estimated from 2008 BRFSS data, N= 28380.

hospital visits, dental care, and "other" medical care not included in the first five categories. The survey also asked about monthly out-of-pocket prescription medication costs, which we converted to estimated yearly costs by multiplying by 12. Participants were only asked about a given category of spending if they reported use of that category of medical care.

In cases where participants could not give a close estimate of how much they spent in a given category, they were asked follow-up probes that broke spending into nine possible intervals. We incorporated answers to probes into total spending estimates using the midpoint of each probe interval, except for top-coded intervals, which we coded as their lower bounds.

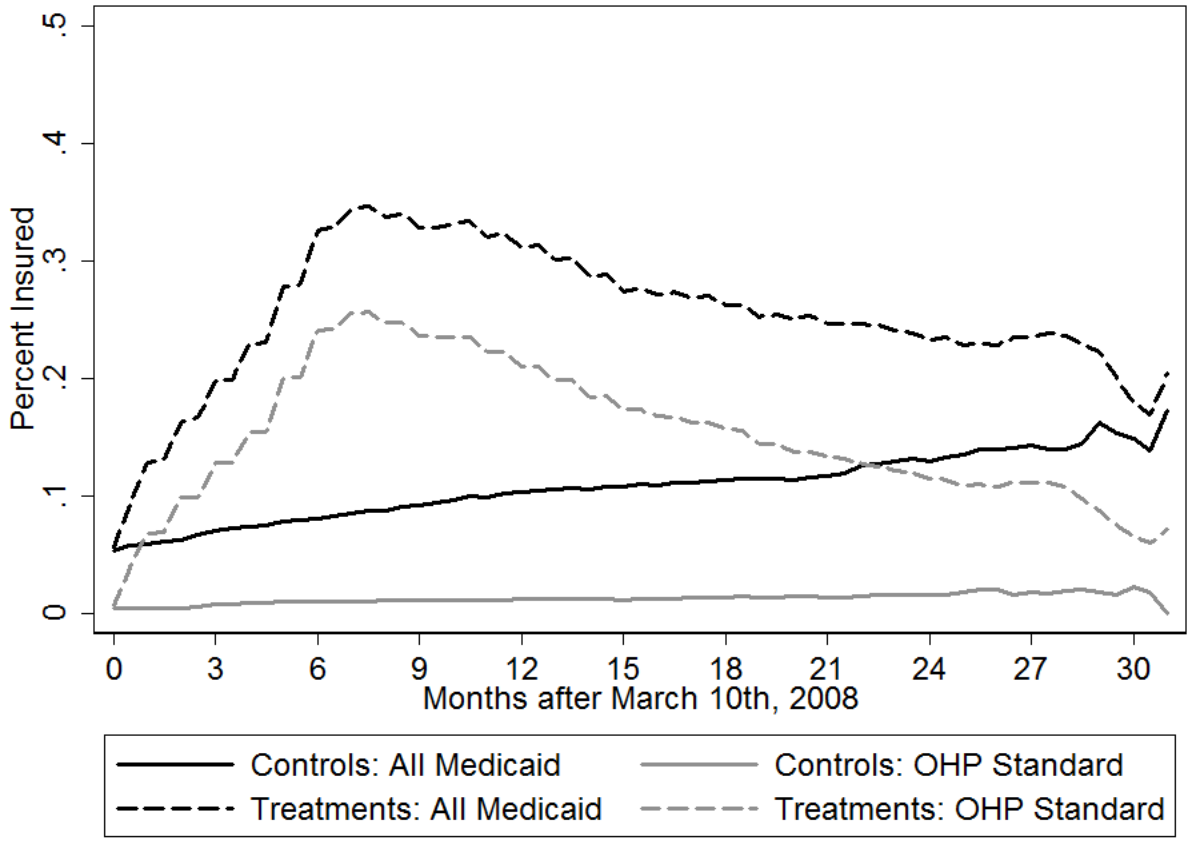
We defined **any out-of-pocket spending** as occurring when the individual reported non-zero spending in at least one of the following: doctor visits, ED visits, outpatient surgery, hospital visits, or other medical care. We did not include out-of-pocket spending on dental care because such care in not covered by OHP. We defined the **amount of out-of-pocket spending** as the sum of reported spending for the same categories (doctor visits, ED visits, outpatient surgery, hospital visits, prescription medications, and other medical care). We treated the sum as missing if any of the component measures was missing. We truncated the amount of out-of-pocket spending at $2*99^{th}$ percentile, recoding outliers to missing.

We defined **catastrophic expenditures** as occurring when the amount of an individual's reported out-of-pocket spending on himself exceeded 30 percent of reported household income. Household income was reported in brackets; for this calculation we used the midpoint of each bracket and the lower bound of the top bracket ($50,000). Different studies use different definitions for catastrophic expenditures based on the share of total income or post-subsistence income (Xu et al. 2003, King et al. 2009). We used a cut-point of 30 percent of income following King et al., but we use total income (whereas they used post-subsistence income), as there was no clear way to separate subsistence and post-subsistence income in our data.

The questions on medical debt were taken from our 12-month mail survey (Finkelstein et al. 2011). We defined **any medical debt** by the individual's response to the question "Do you currently owe money to a health care provider, credit card company, or anyone else for medical expenses?" We defined **borrowed money or skipped bills** by the individual's response to the question "In the last 12 months, have you had to borrow money, skip paying other bills, or pay other bills late in order to pay health care bills?"

Figures

Figure A1: Enrollment in OHP Standard and all Medicaid



Notes: Figure shows the weighted percent with public insurance coverage over time. Weighted percent with insurance is shown separately for treatments and controls, and both all Medicaid coverage and OHP Standard coverage percentages are given. Time is measured in months from March 10, 2008; percent enrolled is observed twice a month. Individuals are censored following interview date and no longer contribute to the weighted percent. This mimics how we define the study period for each individual (from March 10, 2008 to interview date). The numbers closest to the end of the time period (28 months or more from March 10, 2008) are thus based on small numbers for respondents and not estimated precisely. Sample consists of survey respondents (N=**12,229**)

Tables

Table A1: Distribution of analytic weights

	Mean	SD	Min	5th %tile	25th %tile	Median	75th %tile	95th %tile	Max	N
Sampling base	0.998	1.282	0	0	0	1	1.159	3.46	52.211	20745
Recruitment base	1.511	1.31	0.671	0.978	1	1.15	1.378	3.491	52.211	13707
Survey respondents	1.24	0.57	0.681	0.96	1	1.07	1.212	2.076	13.634	12229
Control respondents	1.301	0.663	0.681	0.95	1	1.14	1.307	2.361	13.634	5842
Treatment respondents	1.178	0.461	0.862	1	1	1.003	1.152	1.815	10.872	6387

Notes: Zero weights are the result of being dropped from active follow-up. The recruitment base is the sampling based limited to those with non-zero weights. Respondents all have non-zero weights.

Table A2: Balance of treatment and controls for responders to the in-person survey

	Mean for controls	Difference between treatment and controls
	(1)	(2)
Panel A: Response Rates		
Responded to survey	0.73	0.0028 (0.016) [0.856]
Had Anthropometric Measures	0.977	0.00037 (0.0029) [0.9]
Had at least one DBS measure	0.995	0.0003 (0.0014) [0.828]
Had all DBS measures	0.991	0.0021 (0.0024) [0.367]
Had valid medication data (or does not need medication data)	0.984	0.00027 (0.0025) [0.912]
Pooled F-Stat ***p-value*** ***N***		0.243 0.914 12229
Panel B: Responder and interview characteristics, limited to responders		
Age	40.723 (11.688)	0.205 (0.247) [0.408]
Female	0.569	-0.0046 (0.0087) [0.597]
Black	0.105	0.0014 (0.0061) [0.824]
Other race	0.148	0.00034 -0.008 [0.966]
Hispanic	0.172	-0.0019 -0.0084 [0.818]
English as preferred language	0.907	-0.0017 (0.0069) [0.805]
Signed self up	0.895	0.0011

		(0.002) [0.572]
Signed up first day of list	0.096	0.0064 (0.0068) [0.344]
Gave phone number	0.876	-0.00029 (0.0075) [0.969]
Address a PO Box	0.03	0.0046 (0.0039) [0.245]
In MSA	0.996	0.0019 (0.0011) [0.096]
Median hh income of zip code	44097.99 (9563.062)	-12.417 (210.951) [0.953]
Interview date (difference in days)	24April2010 (102.965)	-1.405 (2.286) [0.539]
Response time (days)	43.71 (52.338)	1.216 (1.617) [0.452]
Winter interview	0.202	0.0011 (0.0072) [0.876]
Spring interview	0.28	-0.011 (0.0093) [0.217]
Summer interview	0.194	0.016 (0.0092) [0.081]
Weekend interview	0.11	-0.0037 (0.0064) [0.561]
In-home interview	0.092	-0.0055 (0.0066) [0.408]
East side clinic interview	0.395	0.013 (0.011) [0.215]
South side clinic interview	0.209	0.0031 (0.0092) [0.74]
Spanish language instrument	0.092	0.00097 (0.0065)

		[0.881]
Interviewed with interpreter	0.028	-0.0038 (0.0049) [0.433]
Pooled F stat ***p-value*** ***N***		0.798 0.738 12229
Panel C: Measurement variables, limited to responders		
Interviewer (pooled F-stat) *p-value* *N*		1.211 0.15 12224
Scale (pooled F-stat) *p-value* *N*		1.108 0.291 12202
Stadiometer (pooled F-stat) *p-value* *N*		1.152 0.231 12211
Sphygmomanometer (pooled F) *p-value* *N*		0.855 0.728 12189

Notes:

Panel A shows the response rate to the survey.

Panel B variables are pre-randomization "demographics" taken from the lottery list (from January and February 2008) and characteristics of the interview itself. Age, sex, race and ethnicity are taken from information reported in the interview; respondents were allowed to report multiple races. The next set of variables is from the lottery list. "English as preferred language" indicates whether you did not check a box requesting materials in a language other than English. "Signed up self?" is an indicator for whether you signed yourself up (as opposed to a household member including your name when they signed up). "Signed upon first day of list?" is an indicator variable for whether you signed up the first day the list was open. "Gave phone number" is an indicator variable for whether you provided a phone number when you signed up. "Interview date" is the time when the interview was conducted. The unit for the mean and the standard deviation are in days. "Response time" indicates the number of days between when the study participant was first released to an interviewer for recruitment and when the interview took place. "Response time" is missing for 885 survey responders because we could not accurately identify their release date. "Winter interview", "Spring interview", "Summer interview" are indicators for whether the interview was conducted in the corresponding season; the omitted category is "Fall interview." "Spring" is defined as March, April, and May, and all other seasons are defined accordingly. "Weekend interview" indicates if the interview took place on a weekend. "In-home interview", "East side clinic interview", "South side clinic interview" are indicators for whether the interview was conducted in the corresponding location; the omitted category is "West side clinic interview." "Spanish language instrument" is an indicator for whether the survey instrument is in Spanish. "Interviewed with interpreter" is an indicator for whether an interpreter was present during the interview.

All analysis is weighted using survey weights. The first column reports the mean of these variables for the control sample and standard deviation for continuous variables. Column (2) reports estimated differences between treatments and controls in the survey responders for the outcome shown in the left hand (except in Panel A where the whole survey sample is used). Specifically it reports the coefficient on LOTTERY based on estimating equation (1); the dependent variable is given in the left hand column. All regressions include household fixed effects and all standard errors are clustered on household. We report the coefficient, standard error, and per comparison p-value. The last row of panel B reports the pooled F-stat from estimating for all the variables in that panel jointly.

Panel C reports global tests for if there is any evidence of sorting across interviewers or equipment used. The scales are equipment used for measuring weight, the stadiometers are equipment used for measuring height and the

sphygmomanometers are equipment used for measuring blood pressure. There are 49 interviewers, and we have interviewer information for all but 5 observations. We could identify the scale, stadiometers, sphygmomanometers used for 12202, 12211, 12189 observations, respectively. A few equipments are only used once or twice. To increase power, for each category, we grouped all equipments used for 10 or fewer observations into an "other" category. After this grouping, there are 44 different scale groups, 44 stadiometer groups, and 43 sphygmomanometer groups in our analysis. The global test for sorting across interviewers (scales, stadiometers, sphygmomanometers) calculated by estimating equation (1) with each of the 49 interviewers (44 scales, 44 stadiometers, 43 sphygmomanometers) as the outcome, then testing whether the 49 coefficients on LOTTERY are equal.

Table A3: Balance of pre-randomization diagnoses

	Control Mean	Difference between treatment and controls
	(1)	(2)
Asthma	0.199	-0.0068 (0.008) [0.396]
Diabetes	0.072	-0.0016 (0.005) [0.755]
Hypertension	0.181	0.0021 (0.0076) [0.788]
High cholesterol	0.127	-0.0015 (0.0067) [0.828]
Heart attack	0.020	-0.0012 (0.0027) [0.663]
Congestive heart failure	0.0097	0.0016 (0.0019) [0.39]
Emphysema/COPD	0.023	0.00018 (0.003) [0.952]
Failing kidneys	0.018	-0.00049 (0.0024) [0.84]
Cancer	0.043	0.0015 (0.0041) [0.716]
Depression/anxiety	0.35	-0.008 (0.0095) [0.401]
Pooled F – stat p-value N		0.299 0.982 12229
Standardized treatment effect p-value N		-0.0026 .761 12229
Diabetes, hypertension, high cholesterol, heart attack, or congestive heart failure	0.273	-0.0026 (0.009) [0.768]

Notes: All analysis weighted using survey weights. The first column reports the mean of these variables for the control sample. Column (2) reports estimated differences between treatments and controls in the survey responders for the outcome shown in the left hand. Specifically it reports the coefficient on LOTTERY based on estimating equation (1); the dependent variable is given in the left hand column. All regressions include household fixed effects. All standard errors are clustered on household. We report the coefficient, standard error, and per comparison p-value. We report the pooled F-stat and the standardized treatment effect from estimating for all the variables (except the composite measure) jointly.

We asked individuals about whether they were ever diagnosed with the following conditions: asthma, diabetes, hypertension, high cholesterol, heart attack, congestive heart failure, emphysema/COPD, kidney failure, cancer, and depression, and when they were diagnosed. If an individual was interviewed in 2010 and answered "more than 3 years ago" to the question "when were you first diagnosed", or if the individual was interviewed in 2009 and answered "more than 2 years ago" to the question "when were you first diagnosed", we knew that the diagnosis was before the lottery. In other cases we asked explicitly about the month and year of diagnosis to determine whether the diagnosis was before or after the lottery. For each of these conditions, we consider the individual to have a **pre-randomization diagnosis of asthma, diabetes, hypertension, high cholesterol, heart attack, congestive heart failure, emphysema/COPD, kidney failure, cancer, or depression** if we could identify that the diagnosis of the specific condition happened before March 10, 2008 (the earliest possible selection date for the lottery).

We consider an individual **high risk** if there was a pre-randomization diagnosis of diabetes, high blood pressure, high cholesterol, heart attack, or congestive heart failure (although there are likely other high risk individuals with such conditions who were never diagnosed). The last row reports the results for this composite measure, which is the one we use to select the sub-sample for the middle panel of Tables 4 through 7.

(Standard errors in parentheses)
[P-values in square brackets]

Table A4: Insurance coverage

	Control mean	Estimated FS
(1) Ever on Medicaid during study period	0.184	0.241
		(0.0090)
		[<0.0001]
(2) Ever on OHP Standard during study period	0.033	0.265
		(0.007)
		[<0.0001]
(3) # of Months on Medicaid during study period	2.558	4.16
		(0.164)
		[<0.0001]
(4) Currently on Medicaid	0.133	0.113
		(0.008)
		[<0.0001]
(5) Currently have any insurance (self-report)	0.358	0.111
		(0.01)
		[<0.0001]
(6) Currently have Medicaid (self-report)	0.128	0.123
		(0.008)
		[<0.0001]
(7) Currently have private insurance (self-report)	0.147	-0.004
		(0.007)
		[0.583]
(8) Ever on TANF	0.043	-0.003
		(0.004)
		[0.487]
(9) TANF Benefits ($)	276.368	-20.135
		(26.706)
		[0.451]
(10) Ever on Food Stamps	0.610	0.023
		(0.008)
		[0.004]
(11) Food Stamp Benefits ($)	3141.772	17.587
		(66.946)
		[0.793]

Notes: The first column reports the weighted control mean for the measure of "INSURANCE" defined in the left-hand column; The second column reports the effect on insurance coverage, which compares the average of the insurance measure for all individuals selected in the lottery to the average of the insurance measure for all control individuals, as calculated by ordinary least squares regression. All regressions include household size fixed effects and are weighted using survey weights. All standard errors are clustered on the household.

The insurance measures are taken from the Medicaid enrollment administrative data except for those labeled "self-report" (rows 5-7) which are taken from the survey. The measures of TANF and food stamps are taken from state administrative records for those progems. In the survey, respondents could report various types of insurance; we define "private insurance" as employer or private insurance and "any insurance" as Medicaid, Medicare, employer, private or other insurance. The study period is defined as running from March 10, 2008 to the date of interview; variables defined as "ever" (rows 1-3 and 8-11) cover this entire period; variables defined as "currently" are current for the interview date. All outcomes are measured for the individual except that in Row 9 (11) where TANF (Food stamp) benefits measure the average monthly benefit amount across all months that the household was covered by TANF during the study period.

Sample consists of survey responders (N = 12,229).

(Standard errors in parentheses)

[p-values in square brackets]

Table A5: Health measures limited to specific pre-randomization conditions

	Control mean (1)	ITT (2)	LATE (3)	p-value (4)
Blood pressure in those with pre-randomization hypertension diagnosis (N=2225)				
Systolic BP	129.83			
	(20.746)			
Diastolic BP	82.877			
	(13.665)			
Prehypertension or hypertension	0.742			
Hypertension	0.38			
Cholesterol in those with pre-randomization high cholesterol diagnosis (N=1549)				
Total cholesterol	205.549			
	(37.087)			
Elevated total chol	0.552			
High total chol	0.172			
HDL cholesterol	53.39			
	(15.367)			
Low HDL chol	0.188			
Glycosylated hemoglobin in those with pre-randomization diabetes diagnosis (N=872)				
Hemoglobin Alc	7.772			
	(1.884)			
Pre-diabetic or diabetic	0.873			
Diabetic	0.698			
Depression in those with pre-randomization depression or anxiety diagnosis (N=4166)				
Positive depression screen	0.521			
Probability of depression	0.432			
	(0.373)			

Notes: See Table 4 notes. For each set of outcomes, the analysis is limited to those with the related pre-randomization diagnosis.

Table A6: Logistic specification: physiologic measures of health

	Control mean	ITT primary spec	ITT logistic spec
	(1)	(2)	(3)
Body mass index			
Overweight/obese	0.713		
Obese	0.415		
Blood pressure			
Prehypertension or hypertension	0.493		
Hypertension	0.163		
Cholesterol			
Elevated total chol	0.444		
High total chol	0.102		
Low HDL chol	0.184		
Glycosylated hemoglobin			
Pre-diabetic or diabetic	0.252		
Diabetic	0.08		

Notes: Column 1 reports the weighted mean of the dependent variable in the control sample of survey responders. Column 2 reports intention-to-treat estimates, which compare the average outcome for all individuals selected in the lottery to the average outcome for all control individuals, as calculated by ordinary least squares regression. Column 3 reports intention-to-treat estimates as calculated by logistic regression. We present average marginal effects for the logistic specification. All regressions include indicators for each household size, and all standard errors are clustered on the household. All analysis is weighted using survey weights. Sample is all survey respondents (N=12,229).

For the blood pressure measures, all regressions also include controls for age (in decile bins) and sex.

(Standard errors in parentheses)
[p-values in square brackets]

Table A7: Logistic specification: Framingham risk score

	Control mean	ITT primary spec	ITT logistic spec
	(1)	(2)	(3)
Female	0.569		
	(0.496)		
BP medication	0.111		
	(0.314)		
Smoker	0.428		
	(0.495)		
Diabetic	0.08		
	(0.272)		

Notes: See Table A6 notes. Sample is all survey respondents (N=12,229).

(Standard errors in parentheses)
[p-values in square brackets]

Table A8: Logistic specification: self-reported measures of health

	Control mean	ITT primary spec	ITT logistic spec
	(1)	(2)	(3)
Health g/vg/e	0.596		
Health not poor or very poor	0.858		
Health same or gotten better	0.804		
Positive depression screen	0.3		
No or very mild pain	0.564		

Notes: See Table A6 notes. Sample is all survey respondents (N=12,229).

(Standard errors in parentheses)
[p-values in square brackets]

Table A9: Logistic specification: health, diagnoses and medications

	Control mean	ITT primary spec	ITT logistic spec
	(1)	(2)	(3)
Panel A: Undiagnosed diseases			
Undiagnosed hypertension	0.072		
Undiagnosed high chl	0.07		
Undiagnosed diabetes	0.023		
Undiagnosed depression	0.091		
Panel B: Unmedicated diseases			
Unmedicated hypertension	0.123		
Unmedicated high chl	0.094		
Unmedicated diabetes	0.032		
Unmedicated depression	0.211		
Panel C: Unmedicated pain			
Pain unmedicated by Rx meds	0.296		
Pain unmedicated by any meds	0.188		

Notes: See Table A6 notes. Sample is all survey respondents (N=12,229).

(Standard errors in parentheses)
[p-values in square brackets]

Table A10: Logistic specification: happiness

	Control mean	ITT primary spec	ITT logistic spec
	(1)	(2)	(3)
Very happy or pretty happy (vs. not too happy)	0.749		

Notes: See Table A6 notes. Sample is all survey respondents (N=12,229).

(Standard errors in parentheses)
[p-values in square brackets]

Table A11: Logistic specification: health care utilization

	Control mean (1)	ITT primary spec (2)	ITT logistic spec (3)
Extensive margin (any)			
Rx drugs currently taking	0.539		
Office Visits (last 12 months)	0.646		
Outpatient Surgery (last 12 months)	0.078		
ED visits (last 12 months)	0.402		
Hospital visits (last 12 months)	0.127		

Notes: See Table A6 notes. Sample is all survey respondents (N=12,229).

(Standard errors in parentheses)
[p-values in square brackets]

Table A12: Logistic specification: preventive care

	Control mean (1)	ITT primary spec (2)	ITT logistic spec (3)
Cholesterol checked	0.272		
Blood stool test (age >=50)	0.191		
Colonoscopy (age >=50)	0.104		
Flu shot (age >=50)	0.355		
Pap smear (women)	0.449		
Mammogram (women >=50)	0.289		
PSA (men >=50)	0.214		

Notes: See Table A6 notes. Sample is all survey respondents (N=12,229), survey respondents at least 50 years of age (N=3374), female survey respondents (N=6915), female survey respondents at least 50 years of age (N=1864) or male survey respondents at least 50 years of age (N=1509), as indicated in the table.

(Standard errors in parentheses)
[p-values in square brackets]

Table A13: Logistic specification: access and quality

	Control mean (1)	ITT primary spec (2)	ITT logistic spec (3)
Panel A: Access			
Have usual place of clinic based care	0.461		
Got all needed medical care	0.61		
Got all needed mental health care	0.756		
Got all needed prescription drugs	0.724		
Panel B: Quality			
Quality of care is good, v good or excellent (condl on any)	0.784		

Notes: See Table A6 notes. Sample is all survey respondents (N=12, 229) except as indicated. All measures of "got needed care" are over the last 12 months.

(Standard errors in parentheses)
[p-values in square brackets]

Table A14: Logistic specification: finances

	Control mean (1)	ITT primary spec (2)	ITT logistic spec (3)
Any out-of-pocket spending	0.588		
Catastrophic expenditures	0.055		
Any medical debt	0.568		
Borrowed money or skipped bills	0.244		

Notes: See Table A6 notes. Sample is all survey respondents (N=12,229).

(Standard errors in parentheses)
[p-values in square brackets]

Table A15: Summary of analytic variables

	Time frame of question	Survey question name(s)*	Non-missing data (N)	Non-missing data (%)
Physiologic measures of health				
BMI	Current	physical measures	12175	99.6
Systolic blood pressure	Current	physical measures	12188	99.7
Diastolic blood pressure	Current	physical measures	12188	99.7
Total cholesterol	Current	physical measures	12174	99.6
HDL cholesterol	Current	physical measures	12172	100
Hemoglobin A1C	Current	physical measures	12140	99
Framingham risk score				
Female	Current	gender_inp	12229	100
Age	Current	ageconf_raw	12229	100
On bp medication	Current	hbp_rx_use_inp	12229	100
Smoker	Current	smk_now_inp	12225	100
Framingham risk score	Current	from the above + physical measures	4134	33.8
Self-reported measures of health				
Health status	Last 12 months	health_last12_inp	12222	99.9
Health change not worse	Last 12 months	health_change_inp	12226	100
SF-8 physical subscale score	Last 4 weeks	sf1_inp to sf8_inp	12204	99.8
SF-8 mental subscale score	Last 4 weeks	sf1_inp to sf8_inp	12204	100
PHQ total severity score	Last 4 weeks	phq1_inp to phq8_inp	12161	99
No or mild pain?	Last 4 weeks	sf4_inp	12225	100
Health, diagnoses and medications				
Undiagnosed hypertension	Current	hbp_dx_inp + phyiscal measures	12151	99.4
Undiagnosed high chol	Current	chl_dx_inp + physical measures	11986	98
Undiagnosed diabetes	Current	dia_dx_inp + physical measures	12,137	99.2
Undiagnosed depression	Current	dep_dx_inp + physical measures	12188	99.7
Unmedicated hypertension	Current	medication survey + physical measures	12198	99.7
Unmedicated high cholesterol	Current	medication survey + physical measures	12176	99.6
Unmedicated diabetes	Current	medication survey + physical measures	12143	99.3
Unmedicated depression	Current	medication survey + physical measures	12175	100
Unmedicated pain	Current	medication survey + physical measures	12226	100
Happiness				
Very happy or pretty happy	Current	happy_inp	12206	99.8
Health care utilization				
Any Rx drugs taken	Last 4 weeks	rx_any_inp + medication survey responses	12226	100
Any office visits	Last 12 months	doc_use_inp and doc_use_probe_inp	12205	99.8
Any outpatient surgery	Last 12 months	surg_use_inp and surg_use_probe_inp	12,204	99.8
Any emergency dept. visits	Last 12 months	ed_use_inp and ed_use_probe_inp	12204	99.8
Any hospital visits	Last 12 months	hosp_use_inp and hosp_use_probe_inp	12205	99.8
Number of Rx drugs taking	Current	rx_any_inp + medication survey responses	11912	97.4
Number of office visits	Last 12 months	doc_use_inp and doc_use_probe_inp	12158	99.4
Number of outpatient surg	Last 12 months	surg_use_inp and surg_use_probe_inp	12188	100
Number of ED visits	Last 12 months	ed_use_inp and ed_use_probe_inp	12175	99.6
Number of hospital visits	Last 12 months	hosp_use_inp and hosp_use_probe_inp	12175	99.6
Preventive care in last 12 months (all use care_any_inp and probe)				
Cholesterol checked	Last 12 months	chl_dx_inp, chl_test_when_inp	11382	93.1
Had a blood stool test (age≥50)**	Last 12 months	fobt_ever_inp	3358	99.5
Had a colonoscopy (age≥50)**	Last 12 months	col_ever_inp	3,361	99.6
Had a flu shot (age≥50)**	Last 12 months	did_flu_inp	3364	99.7
Had a pap smear (women)**	Last 12 months	pap_inp	6673	96.9
Had mammogram (women, age≥50)**	Last 12 months	mam_inp	1858	100
Had a PSA test (men, age≥50)**	Last 12 months	psa_inp	1381	91.1

Access and quality				
Have usual place of clinic-based care	Current	usual_place_inp, usual_place_where_inp	12219	99.9
Got all needed medical care	Last 12 months	got_care_phs_inp, needed_care_phys_inp	12216	99.9
Got all needed mental health care	Last 12 months	got_care_ment_inp, needed_care_med_inp	12192	99.7
Got all needed drugs	Last 12 months	rx_delay_inp, rx_inp	12215	99.9
Quality of care (cond. on any)	Last 12 months	satisfaction_inp	9694	99.3
Finances				
Any out-of-pocket spending	Last 12 months	doc/surg/ed/hosp/other_cost_inp plus	12194	99.7
Amount of out-of-pocket spending	Last 12 months	probes	12145	99.3
Had catastrophic expenditures	Last 12 months	from the above + hh_income_inp	11795	96.5
Any medical debt	Current	owe_inp	12108	99
Borrow money or skipped bills	Last 12 months	borrow_inp	12212	99.9

*The survey question names are provided to identify the questions in the survey instrument, which is available online at www.nber.org/oregon.

**The count of non-missing observations is restricted when the question only applies to a particular subgroup (e.g., we would only expect responses for mammogram, pap test questions from women).

Table A16: Distribution of variables (control sample only)

Panel A: Physiologic Health

	Mean	SD	5th %tile	25 %tile	Median	75th %tile	95th %tile
BMI	29.82	7.58	20.23	24.44	28.56	33.6	44.27
Systolic Blood Pressure	119.28	16.85	97	107	117	128	149
Diastolic Blood Pressure	76.01	12.14	58	67	75	83	98
Total Cholesterol	198.53	32.04	151.71	176.29	195.65	218.13	254.74
HDL Cholesterol	53.56	14.93	31.44	43.37	52.74	62.11	80.85
Hemoglobin A1C	5.58	1.01	4.69	5.09	5.35	5.75	7.47

Panel B: Health Status

	N	%
General health, last 12 mo.		
1: Very poor	134.72	2.31
2: Poor	692.40	11.86
3: Fair	1533.75	26.27
4: Good	2100.89	35.99
5: Very good	1024.71	17.55
6: Excellent	351.52	6.02
Health status compared to 12 mo. ago		
0: Better	1461.07	25.02
1: Worse	1144.10	19.59
2: About the same	3233.83	55.38

Note: All numbers are weighted

Panel C: Self Reported Health

	Mean	SD	5th %tile	25 %tile	Median	75th %tile	95th %tile
SF-8 physical subscale score	45.49	10.5	25.65	37.83	47.71	54.16	57.32
SF-8 mental subscale score	44.39	11.38	22.62	36.48	46.76	53.31	57.67
Probability of Depression	0.26	0.34	0	0	0.129	0.58	0.92

Panel D: Health Care Use

	Percent reporting any	Mean	SD	Median	75th %tile	95th %tile	Cutpoint for truncation	% of data truncated
Doctor office visits	64.55	8.61	13.49	4	10	30	164	0.5
Outpatient surgery visits	7.81	1.25	.58	1	1	2	4	0.2
ED visits	40.23	2.48	2.5	2	3	7	20	0.1
Inpatient hospital visit	12.66	1.5	1.04	1	2	4	6	0.3
Number of Rx drugs	53.89	3.48	3.04	2	5	10	n/a	n/a

Note: In Panels D and E, the mean, standard deviation, median, 75th and 95th percentile values reflect non-zero observations only, after truncating at 2*99% based on the unweighted distribution. "Number of Rx drugs" is not truncated. Percent reporting any use, cutpoint for censoring and percent of data censored reflect all valid non-missing data, including observations with zero values. The value for “percent reporting any use” for “Number of Rx drugs” includes 157 control respondents who reported taking Rx drugs but for whom we could not accurately count the number of Rx drugs. All numbers in table except “Cutpoint for truncation” are weighted.

Panel E: Financial Strain

	Percent reporting any	Mean	SD	Median	75th %tile	95th %tile	Cutpoint for truncation	% of data truncated
Total out of pocket expense	58.8	942.18	1472.37	440	1075	3661	15200	0.3
Total medical expense	57.68	1066.43	2916.92	425	1040	3816	n/a	n/a
Total other expense	8.3	312.83	491.4	150	350	1200	n/a	n/a

Note: "Total out of pocket expense" is the sum of "total medical expense" and "total other expense." In Panels D and E, the mean, standard deviation, median, 75th and 95th percentile values reflect non-zero observations only, after truncating at 2*99% based on the unweighted distribution. "Total medical expense" and "total other expense" are not truncated. Percent reporting any expenses, cutpoint for censoring and percent of data censored reflect all valid non-missing data, including observations with zero values. Missing values are largely due to individuals answering "don't know" or "prefer not to answer" to the survey question. All numbers except “Cutpoint for truncation” are weighted.

Table A17: Classification of medications

Panel A: Distribution of Anti-hypertensives

Medication Name	Frequency	Percent	Cumulative
Lisinopril	950	38.28	38.28
Hydrochlorothiazide	531	21.39	59.67
Furosemide	151	6.08	65.75
Clonidine	137	5.52	71.27
Lisinopril-Hydrochlorothiazide	106	4.27	75.54
Cozaar	66	2.66	78.2
Spironolactone	64	2.58	80.78
Triamterene-Hydrochlorothiazid	47	1.89	82.68
Doxazosin	36	1.45	84.13
Enalapril Maleate	32	1.29	85.41
Losartan	24	0.97	86.38
Terazosin	23	0.93	87.31
Accupril	21	0.85	88.15
Prazosin	17	0.68	88.84
Benicar	16	0.64	89.48
Diovan	16	0.64	90.13
Lasix	16	0.64	90.77
Chlorthalidone	16	0.64	91.42
Hydralazine	15	0.6	92.02
Amlodipine-Benazepril	14	0.56	92.59
Quinapril	14	0.56	93.15
Accuretic	10	0.4	93.55
Aldactone	10	0.4	93.96
Cardura	10	0.4	94.36
Atenolol-Chlorthalidone	10	0.4	94.76
Other	130	5.2	100

Panel B: Distribution of Antihyperlipidemics

Medication Name	Frequency	Percent	Cumulative
Simvastatin	411	31.79	31.79
Lipitor	228	17.63	49.42
Lovastatin	214	16.55	65.97
Pravastatin	87	6.73	72.7
Gemfibrozil	69	5.34	78.04
Fish Oil	68	5.26	83.29
Crestor	32	2.47	85.77
Atorvastatin	23	1.78	87.55
Omega 3 Fish Oil	14	1.08	88.63
Niacin	13	1.01	89.64
Niaspan Extended-Release	12	0.93	90.56

Lovaza	10	0.77	91.34
Tricor	10	0.77	92.11
Zocor	10	0.77	92.88
Other	92	7.13	100

Panel C: Distribution of Diabetes Medicine

Medication Name	Frequency	Percent	Cumulative
Metformin	657	42.22	42.22
Glipizide	174	11.18	53.41
Lantus	140	9	62.4
Glyburide	96	6.17	68.57
Humalog	53	3.41	71.98
Novolog	39	2.51	74.49
Novolin R	35	2.25	76.74
Actos	34	2.19	78.92
Novolin 70/30	33	2.12	81.04
Novolin N	27	1.74	82.78
Glucotrol Xl	23	1.48	84.25
Glimepiride	20	1.29	85.54
Glucophage	15	0.96	86.5
Humulin N	15	0.96	87.47
Humulin R	15	0.96	88.43
Lantus Solostar	14	0.9	89.33
Glucotrol	11	0.71	90.04
Novolog Flexpen	11	0.71	90.75
Other	144	9.19	100

Panel D: Distribution of Antidepressants

Medication Name	Frequency	Percent	Cumulative
Trazodone	430	16.39	16.39
Citalopram	384	14.64	31.03
Amitriptyline	215	8.2	39.23
Sertraline	200	7.62	46.85
Cymbalta	196	7.47	54.33
Bupropion Hcl	137	5.22	59.55
Paroxetine Hcl	121	4.61	64.16
Lexapro	113	4.31	68.47
Effexor Xr	96	3.66	72.13
Zoloft	95	3.62	75.75
Prozac	68	2.59	78.35
Celexa	63	2.4	80.75
Venlafaxine	52	1.98	82.73
Bupropion (Bulk)	49	1.87	84.6
Nortriptyline	49	1.87	86.47
Wellbutrin	46	1.75	88.22

Wellbutrin Sr	36	1.37	89.59
Mirtazapine	35	1.33	90.93
Doxepin	33	1.26	92.18
Paxil	32	1.22	93.4
Wellbutrin Xl	25	0.95	94.36
Pristiq	23	0.88	95.23
Effexor	20	0.76	96
Budeprion Xl	19	0.72	96.72
Budeprion Sr	18	0.69	97.41
Escitalopram	10	0.38	97.79
Other	58	2.25	100

Panel E: Distribution of Pain Medications (Rx Analgesics)

Medication Name	**Frequency**	**Percent**	**Cumulative**
Ibuprofen	564	17.57	17.57
Hydrocodone-Acetaminophen	378	11.78	29.35
Oxycodone	356	11.09	40.44
Vicodin	340	10.59	51.03
Naproxen	197	6.14	57.17
Aspirin	145	4.52	61.68
Clonidine	136	4.24	65.92
Percocet	106	3.3	69.22
Morphine	89	2.77	71.99
Oxycodone-Acetaminophen	88	2.74	74.74
Oxycontin	52	1.62	76.36
Meloxicam	41	1.28	77.63
Aspirin Low Dose	31	0.97	78.6
Acetaminophen	31	0.97	79.56
Tylenol	30	0.93	80.5
Piroxicam	30	0.93	81.43
Advil	24	0.75	82.18
Acetaminophen-Codeine	24	0.75	82.93
Norco	23	0.72	83.64
Naproxen Sodium	23	0.72	84.36
Diclofenac Sodium	21	0.65	85.02
Etodolac	18	0.56	85.58
Salsalate	17	0.53	86.11
Enteric Coated Aspirin	15	0.47	86.57
Motrin	15	0.47	87.04
Tylenol Extra Strength	15	0.47	87.51
Indomethacin	15	0.47	87.98
Nabumetone	15	0.47	88.44
Aleve	14	0.44	88.88
Tylenol-Codeine #3	13	0.4	89.28
Ultram	12	0.37	89.66

Vicodin Es	12	0.37	90.03
Fentanyl	12	0.37	90.4
Hydrocodone-Ibuprofen	12	0.37	90.78
Hydromorphone	12	0.37	91.15
Celebrex	11	0.34	91.5
Endocet	11	0.34	91.84
Diclofenac Potassium	11	0.34	92.18
Propoxyphene N-Acetaminophen	11	0.34	92.52
Oxycodone Hcl-Oxycodone-Asa	10	0.31	92.83
Other	230	7.05	100

Notes: Some of these medications are available either over the counter or by prescription. They are included here if the participant reported that the medication had be prescribed.

Panel F: Distribution of Pain Medications (Prescription and OTC Analgesics)

Medication Name	**Freq.**	**Percent**	**Cum.**
Ibuprofen	1,709	23.98	23.98
Aspirin	603	8.46	32.44
Hydrocodone-Acetaminophen	378	5.3	37.74
Oxycodone	356	4.99	42.73
Vicodin	340	4.77	47.5
Advil	331	4.64	52.15
Tylenol	324	4.55	56.69
Tylenol Extra Strength	317	4.45	61.14
Aleve	255	3.58	64.72
Naproxen	215	3.02	67.73
Aspirin Low Dose	163	2.29	70.02
Clonidine	137	1.92	71.94
Excedrin Migraine	109	1.53	73.47
Acetaminophen	109	1.53	75
Excedrin Extra Strength	108	1.52	76.52
Percocet	106	1.49	78
Morphine	89	1.25	79.25
Oxycodone-Acetaminophen	88	1.23	80.49
Advil Liqui-Gel	73	1.02	81.51
Naproxen Sodium	71	1	82.51
Motrin	55	0.77	83.28
Oxycontin	52	0.73	84.01
Acetaminophen Extra Strength	50	0.7	84.71
Baby Aspirin	50	0.7	85.41
Bayer Aspirin	49	0.69	86.1
Meloxicam	41	0.58	86.67
Piroxicam	30	0.42	87.09
Ibuprofen Ib	26	0.36	87.46

Acetaminophen Pain Relief	25	0.35	87.81
Aspirin Child	24	0.34	88.15
Low-Dose Aspirin	24	0.34	88.48
Acetaminophen-Codeine	24	0.34	88.82
Advil Pm	23	0.32	89.14
Norco	23	0.32	89.46
Aspirin Low-Strength	21	0.29	89.76
Diclofenac Sodium	21	0.29	90.05
Aspirin Extra Strength	19	0.27	90.32
Enteric Coated Aspirin	18	0.25	90.57
Etodolac	18	0.25	90.82
Ecotrin	17	0.24	91.06
Salsalate	17	0.24	91.3
Excedrin Ib	16	0.22	91.53
Tylenol Arthritis Pain	16	0.22	91.75
Excedrin Back & Body	15	0.21	91.96
Indomethacin	15	0.21	92.17
Nabumetone	15	0.21	92.38
Acetaminophen Non Aspirin	14	0.2	92.58
Bayer Childrens Aspirin	14	0.2	92.77
Excedrin Tension Headache	14	0.2	92.97
Tylenol 8 Hour	14	0.2	93.17
Motrin Ib	13	0.18	93.35
Tylenol-Codeine #3	13	0.18	93.53
Ultram	12	0.17	93.7
Vicodin Es	12	0.17	93.87
Fentanyl	12	0.17	94.04
Hydrocodone-Ibuprofen	12	0.17	94.21
Hydromorphone	12	0.17	94.37
Celebrex	11	0.15	94.53
Endocet	11	0.15	94.68
Diclofenac Potassium	11	0.15	94.84
Propoxyphene N-Acetaminophen	11	0.15	94.99
Aspirin Maximum Strength	10	0.14	95.13
Oxycodone Hcl-Oxycodone-Asa	10	0.14	95.27
Other	337	4.52	100

References

American Diabetes Association (2010) Standards of medical care in diabetes--2010. *Diabetes Care,* 33 Suppl 1**,** S11-61.

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

Angrist, J. D. & J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Centers for Disease Control and Prevention (CDC). 2000. Behavioral Risk Factor Surveillance System Survey Questionnaire. Atlanta, GA: U.S Department of Health and Human Services, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS). 2000. National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Chobanian, A. V., G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright, E. J. Roccella, National Heart Lung and Blood Institute Joint National Committee on Prevention Detection Evaluation and Treatment of High Blood Pressure & National High Blood Pressure Education Program Coordinating Committee (2003) The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA,* 289**,** 2560-72.

Chou, R., J. M. Croswell, T. Dana, C. Bougatsos, I. Blazina, R. Fu, K. Gleitsmann, H. C. Koenig, C. Lam, A. Maltz, J. B. Rugge & K. Lin (2011) Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med,* 155**,** 762-71.

Cole, S. R. & M. A. Hernán (2008) Constructing inverse probability weights for marginal structural models. *Am J Epidemiol,* 168**,** 656-64.

D'Agostino, R. B., R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro & W. B. Kannel (2008) General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation,* 117**,** 743-53.

Expert Panel on Detection Evaluation And Treatment of High Blood Cholesterol In Adults (2001) Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP). *JAMA,* 285**,** 2486-97.

Expert Panel on the Identification Evaluation and Treatment of Overweight and Obesity in Adults (1998) Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: executive summary. *Am J Clin Nutr,* 68**,** 899-917.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker & The Oregon Health Study Group (2011) The Oregon Health Insurance Experiment: Evidence from the First Year. *NBER Working Paper,* 17190.

Fiore, A. E., T. M. Uyeki, K. Broder, L. Finelli, G. L. Euler, J. A. Singleton, J. K. Iskander, P. M. Wortley, D. K. Shay, J. S. Bresee, N. J. Cox & Centers for Disease Control and Prevention (CDC) (2010) Prevention and control of influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP), 2010. *MMWR Recomm Rep,* 59**,** 1-62.

Flegal, K. M., M. D. Carroll, C. L. Ogden & L. R. Curtin (2010) Prevalence and trends in obesity among US adults, 1999-2008. *JAMA,* 303**,** 235-41.

Greene, K. L., P. C. Albertsen, R. J. Babaian, H. B. Carter, P. H. Gann, M. Han, D. A. Kuban, A. O. Sartor, J. L. Stanford, A. Zietman & P. Carroll (2009) Prostate specific antigen best practice statement: 2009 update. *J Urol,* 182**,** 2232-41.

Health and Retirement Study. 2000. Health and Retirement Study 2000 Questionnaire. Ann Arbor, MI: University of Michigan.
Huang, F. Y., H. Chung, K. Kroenke, K. L. Delucchi & R. L. Spitzer (2006) Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med,* 21**,** 547-52.
Imbens, G. W. & J. D. Angrist (1994) Identification and Estimation of Local Average Treatment Effects. *Econometrica,* 62**,** 467-475.
International Expert Committee (2009) International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care,* 32**,** 1327-34.
Kalton, G. (1986) Handling wave nonresponse in panel surveys. *Journal of Official Statistics,* 2**,** 303-314.
King, G., E. Gakidou, K. Imai, J. Lakin, R. T. Moore, C. Nall, N. Ravishankar, M. Vargas, M. M. Tellez-Rojo, J. E. Hernandez Avila, M. Hernandez Avila & H. Hernandez Llamas (2009) Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme. *The Lancet,* 373**,** 1447-1454.
Kroenke, K. & R. L. Spitzer (2002) The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatric Annals,* 32**,** 509-521.
Kroenke, K., R. L. Spitzer & J. B. Williams (2001) The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med,* 16**,** 606-13.
Kroenke, K., T. W. Strine, R. L. Spitzer, J. B. Williams, J. T. Berry & A. H. Mokdad (2009) The PHQ-8 as a measure of current depression in the general population. *J Affect Disord,* 114**,** 163-73.
McDade, T. W., S. Williams & J. J. Snodgrass (2007) What a drop can do: dried blood spots as a minimally invasive method for integrating biomarkers into population-based research. *Demography,* 44**,** 899-925.
McHorney, C. A., J. E. Ware & A. E. Raczek (1993) The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care,* 31**,** 247-63.
Shim, R. S., P. Baltrus, J. Ye & G. Rust (2011) Prevalence, treatment, and control of depressive symptoms in the United States: results from the National Health and Nutrition Examination Survey (NHANES), 2005-2008. *J Am Board Fam Med,* 24**,** 33-8.
U.S. Preventive Services Task Force (2003) Screening for cervical cancer: recommendations and rationale. *Am J Nurs,* 103**,** 101-2, 105-6, 108-9.
--- (2008) Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med,* 149**,** 627-37.
--- (2009) Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med,* 151**,** 716-26, W-236.
Ware, J. E., M. Kosinki, J. Dewey & B. Gandek. 2001. How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8 Health Survey. Boston: QualityMetric Inc.
World Health Organization. 2011. World Health Organization Global Database on Body Mass Index.
Xu, K., D. B. Evans, K. Kawabata, R. Zeramdini, J. Klavus & C. J. Murray (2003) Household catastrophic health expenditure: a multicountry analysis. *Lancet,* 362**,** 111-7.