Oregon Health Insurance Experiment
Public Use Data User Guide

Last Updated: January 2, 2014

Introduction

This document provides additional information about the randomization procedure and construction of the analytic sample for the Oregon Health Insurance Experiment. It identifies the sources of administrative data and explains the sampling procedure for mail surveys conducted by the study group.

Contents

Citation for Public Use Data

If obtained through Oregon Health Insurance website: Finkelstein, Amy. Oregon Health Insurance Experiment Public Use Data, 2013. Available at http://www.nber.org/oregon/data.html.

If obtained through ICPSR: Finkelstein, Amy. Oregon Health Insurance Experiment Public Use Data. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013.

Additional Information

Users desiring more detail about sampling procedures, response rates, and data sources as well as previous analyses are encouraged to consult:

Finkelstein, Amy et al "The Oregon Health Insurance Experiment: Evidence from the First Year". *The Quarterly Journal of Economics.* August 2012. Vol 127(3). (Online Appendix also available)

Baicker, Katherine et al "The Oregon Experiment – Effects of Medicaid on Clinical Outcomes". *New England Journal of Medicine.* 2 May 2013. Vol 368(18). (Online Appendix also available)

Taubman, Sarah et al "Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment". *Science* 2 Jan 2014. (Online Appendix also available)

Randomization

In 2008, Oregon selected roughly 30,000 individuals by lottery from a waiting list of about 90,000 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 a Medicaid program called 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. The randomly selected individuals chosen by the lottery who completed the application process and met the eligibility criteria were enrolled in OHP Standard.

Construction of Analytic Sample

The original lottery list provided to the study group by the state included 100,600 records of people who signed up for the lottery. Once duplicate and deactivated records were removed, this data identified 89,824 individuals. The study group then excluded individuals who gave an address outside of Oregon (and are therefore not eligible for state Medicaid), individuals who would be older than 64 by the end of 2009 (and hence eligible for Medicare) or younger than 19 at the beginning of 2008 (not eligible for the lotteried health care program, OHP Standard). Additional exclusions included those who lived at institutional addresses or who were signed up by an unrelated third party, those with multiple active records on the lottery list, and those who died prior to the matched notification date.

Following exclusions we were left with a total of 74,922 individuals to study. The public use data contains only these 74,922 individuals. Of these individuals, 29,834 were selected as treatments (i.e. won the lottery and were given the chance to apply for health insurance); treatment status is indicated by the variable **treatment** in the Descriptive Variables Dataset, oregonhie_descriptive_vars.dta.

Analysis of the Oregon Health Insurance Experiment data allows the comparison between those selected by the lottery (treatments) and those who signed up but not selected (controls). Crucially, the lottery selected individuals, but the opportunity to apply for health insurance was extended to all household members of lottery winners: treatment selection is random only conditional on the number of household members on the waiting list (this is given by the variable **numhh_list** in the Descriptive Variables Dataset, oregonhie_descriptive_vars.dta). For example, an individual could sign up his or herself as well as a spouse for the lottery, and both have equal probability of being chosen. Thus, this person and his or her spouse are twice as likely to win the opportunity to apply for health insurance as someone who only added their own name to the list, without adding other household members.

Merging Datasets

All datasets contain observations at the individual level. Observations can be linked across datasets by the unique identifier **person_id,** which appears in all datasets. No other variable appears across multiple datasets.

Notes on Individual Datasets

This section contains notes on each public use dataset. It includes information on data sources and notes on analysis. It does not provide a comprehensive description of the contents of each dataset: for more detail about individual variables, please refer to the codebooks provided.

Descriptive Variables
oregonhie_descriptive_vars.dta

This dataset includes demographic information that individuals reported about themselves at the time of lottery sign up (January and February 2008), as well as administrative data from several sources that includes information on lottery sign up and selection as well as insurance application and approval. Mortality data are also included here.

Oregon's Department of Human Services' Division of Medical Assistance Programs (DMAP) provided a complete list of individuals who signed up for the lottery with a person identifier (**person_id**), a household identifier (**household_id**), whether an individual was selected in the lottery (**treatment**) and which draw the individual was selected in (**draw_lottery**). For those selected in the lottery, **dt_notify_lottery** is the date on which an individual was notified of his or her selection and ability to apply for Medicaid. For those not selected by the lottery, **dt_notify_lottery** is a "matched notification date" randomly assigned by the Oregon Health Insurance Study Group to facilitate comparison (see Finkelstein et al 2012 online appendix for additional details).

There are three dates relevant to lottery selection: the date of lottery notification when individuals who were selected by the lottery were notified of their ability to apply for Medicaid (**dt_notify_lottery**); for those who applied, the date a decision was made on the application (**dt_app_decision**); and, for those whose applications were approved, the date to which coverage was retroactively applied (**dt_retro_coverage**). Coverage was retroactively applied to the same date for everybody selected in a given lottery draw. For example, those selected in the first lottery draw on March 10, 2008 had coverage retroactively applied to any medical expenses incurred on or after March 11, 2008 even though their applications were not approved until a later date.

Basic demographic information including gender, birth year, whether English was the preferred language of communication, whether an individual signed up on the first or last day that the lottery list was open, number of people in the household, etc. was provided by individuals on the lottery sign-up form ("lottery list") and is included in this dataset.

Information on Medicaid applications came from Oregon's Office for Health Policy and Research (OHPR) and the Department of Human Services' Children, Adults and Families Division. The data contain information on the status and disposition of any Medicaid applications submitted by individual selected in the lottery through mid-January 2009.

Mortality data came from Oregon's Center for Health Statistics. For individuals who died following their lottery notification date (in 2008) and before 1 January 2010, an indicator variable (**postn_death**) is provided. Individuals who died prior to their notification date are excluded from the public use data provided. This data does not capture deaths which occur outside of Oregon.

State Program Variables
oregonhie_stateprograms_vars.dta

Oregon's DMAP also provided information on Medicaid enrollment histories for each individual on the lottery list. Oregon's Department of Human Services' Children, Adults and Families Division provided, for each individual on the lottery list, data on participation in the Supplemental Nutrition Assistance Program (SNAP) and Temporary Assistance to Needy Families (TANF).

The oregonhie_stateprograms_vars.dta dataset contains the constructed variables that were based on these data and used for analysis in Finkelstein et al 2012 and Baicker et al 2013. In particular, it contains the measures of insurance coverage used in those papers to estimate the effect of Medicaid.

Initial Mail Survey Variables
oregonhie_survey0m_vars.dta

An initial mail survey was fielded between June 2008 and November 2008. The survey sample consisted of 58,405 individuals (29,589 treatments and 28,816 controls). The survey protocol included a screener postcard, 2 survey mailings plus phone follow-up for non-responders. The first of the survey mailings included a $5 cash incentive; in addition, respondents were entered into a lottery to receive an additional $200. Responses were received from 26,423 individuals, a response rate of 45 percent. The average response date of the initial survey was August 29, 2008.

This survey contained questions on health insurance, health care needs and health care use such as whether individuals required medical care, prescription drugs, dental care, etc. Most questions asked individuals to consider the past six months (i.e. "In the last six months, have you needed any dental care?"). The survey itself is available via ICPSR and the website of the Oregon Health Insurance Experiment (www.nber.org/oregon). The survey fielding procedures are described in more detail in the online appendix of Finkelstein et al 2012.

The oregonhie_survey0m_vars.dta dataset has a record for every individual in the Oregon Health Insurance Experiment. Variables capturing survey response are missing for individuals who were not included in the survey sample, who did not respond to the survey, or who did not respond to that specific question.

The 58,405 individuals included in the initial mail survey sample are identified by the variable **sample_0m**. Of the surveyed individuals, 26,423 responded to the survey, identified by the variable **returned_0m.** Note that the follow-up protocol for this survey also included phone calls to a random set of non-respondents; thus, a minority of surveys was completed by phone. This information (method of survey return, phone or mail) is captured by the variable **ret_mode_0m**.

There are no survey weights for the Initial Mail Survey.

Six Month Mail Survey Variables
oregonhie_survey6m_vars.dta

The "six month" mail survey was conducted on a limited subsample (n=11,756) of the initial survey sample, oversampling early respondents to the initial survey. This survey was fielded between January 2009 and May 2009. The survey protocols were the same as for the initial survey, except there was no phone follow-up of non-responders. Responses were received from 6,359 individuals, with an effective (weighted) response rate of 42 percent. The average weighted response date was February 23, 2009.

This survey contained questions on health care needs and health care use such as whether individuals required medical care, prescription drugs, dental care, etc. Most questions asked individuals to consider the past six months (i.e. "In the last six months, have you needed any dental care?"). The survey itself is available via ICPSR and the website of the Oregon Health Insurance Experiment (www.nber.org/oregon). The survey fielding procedures are described in more detail in the online appendix of Finkelstein et al 2012.

The oregonhie_survey6m_vars.dta dataset has a record for every individual in the Oregon Health Insurance Experiment. Variables capturing survey response are missing for individuals who were not included in the survey sample, who did not respond to the survey, or who did not respond to that specific question.

Individuals who were sent a six month survey are identified by the variable **sample_6m.** Of the surveyed individuals, 6,359 responded to the survey, identified by the variable **returned_6m.**

Weights that account for the probability of being sampled are given by the variable **weight_6m**.

Twelve Month Mail Survey Variables
oregonhie_survey12m_vars.dta

The "twelve month" mail survey was conducted on the same sample (n=58,405) as the initial survey. This survey was fielded in seven waves starting in July 2009; extended follow-up lasted until March 2010. The basic survey protocol included a screener postcard and 3 survey mailings (the 3rd including a URL to complete the survey online if desired). The first of the survey mailings included a $5 cash incentive; in addition, respondents were entered into a lottery to receive an additional $200. We selected a subsample of 30 percent of non-respondents for a more intensive follow-up protocol, which included multiple phone follow-up attempts, two additional mailings, additional tracking and monetary incentives. Responses were received from 23,777 individuals. The response rate to the basic protocol for the 12-month survey was 36 percent; following the intensive protocol (described below), the overall effective (weighted) response rate was 50 percent. The average response date was September 19, 2009.

This survey contained questions on health care needs and health care use such as whether individuals required medical care, prescription drugs, dental care, etc. Most questions asked individuals to consider the past six months (i.e. "In the last six months, have you needed any dental care?"). The survey itself is available via ICPSR and the website of the Oregon Health Insurance Experiment (www.nber.org/oregon). The survey fielding procedures are described in more detail in the Online Appendix of Finkelstein et al 2012.

The oregon_survey12m_vars.dta dataset has a record for every individual in the Oregon Health Insurance Experiment. Variables capturing survey response are missing for individuals who were not included in the survey sample, who did not respond to the survey, or who did not respond to that specific question.

Individuals who were sent a twelve month survey are identified by the variable **sample_12m.** The twelve month survey had both a basic and intensive follow-up protocol (the protocol to which an individual was assigned) is given by the variable **protocol_12m**. Method of survey return (mail, phone, or web) is identified by the variable **ret_mode_12m**.

Weights that account for the probability of being sampled and survey fielding procedures are given by the variable **weight_12m**; see online appendix of Finkelstein et al 2012 for detail on the construction of the weights.

In-person Survey Variables
oregonhie_inperson_vars.dta

Note that this dataset contains only the variables required to replicate the analysis in Baicker et al (2013). It does not contain all of the information gathered during fielding of the in-person survey.

The "in-person survey" was conducted from August 31 2009 until October 13, 2010. For logistical reasons, the study sample was limited to the Portland area and included 10,405 individuals selected in the lottery and 10,340 individuals not selected. The survey fielding procedures are described in detail in the Online Appendix of Baicker et al (2013) and in the protocol document available on the website of the Oregon Health Insurance Experiment (www.nber.org/oregon). A total of 12,229 individuals completed the interview. The effective (weighted) response rate for the controls was 73 percent.

The computer-assisted personal interview included questions on health care needs and health care use such as whether individuals required medical care, prescription drugs, dental care, etc. Most questions asked individuals to consider the past year (e.g. "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?). Not every interview subject was asked every question: for example, there were in-depth questions about selected conditions (e.g. asthma, diabetes, hypertension) that were applicable only to individual diagnosed with those conditions. The interviewer script, including skip patterns, is available via ICPSR and the website of the Oregon Health Insurance Experiment (www.nber.org/oregon). In addition to the questions, the interview included blood pressure and anthropometric measurements (weight, height, etc), a catalogue of medications and collection of dried blood spots. The dried blood spots provide measurements of total cholesterol, HDL cholesterol, and glycated hemoglobin.

The dataset has a record for every individual in the Oregon Health Insurance Experiment who was contacted about the in-person survey (N=20,745). Individuals who completed an in-person interview are identified by the variable **sample_inp_resp.**

Weights that account for the probability of being sampled and survey fielding procedures are given by the variable **weight_total_inp;** see online appendix on Baicker et al (2013) for detail on the construction of the weights.

Emergency Department Variables
oregonhie_ed_vars.dta

The Oregon Health Study obtained 2007-20009 visit-level data for all emergency department visits to twelve hospitals in the Portland area. The publicly available data are derived from these visit-level data. The data are at the individual level and the dataset contains only the variables required to replicate the analysis in Taubman et al (2014).

The data contain variables with the number of emergency department visits of different types (and indicators for any emergency department use of that type) over the study period, defined as March 10, 2008 to September 30, 2009. These variables capture any emergency department use, as well as emergency department use categorized by inpatient admission, time of visit (e.g. nights and weekends), and reason for visit. Analogous variables for the pre-randomization period (defined as January 1, 2007 to March 9, 2008) are also included. Most continuous variables (e.g. that capture the total number of ED visits in a given period) have been censored to ensure de-identification. Variables were censored so that no individual value has fewer than ten observations. This results in the right tail of the distribution of these variables being grouped into one large upper bin. Consequently, certain results reported in Taubman et al (2014) cannot be replicated with the publicly available data (for a list of replicable analyses, see note at the top of the replication code: oregonhie_science_replication.do).

The dataset has a record for every individual in the Oregon Health Insurance Experiment who resides in an area that relies primarily on these twelve hospitals for emergency department care (N=24,646). The Online Appendix of Taubman et al (2014) provided details on how that catchment area was defined.