Oregon Health Insurance Experiment:
Patterns Public Use Data User Guide

Last Updated: October, 2016

Introduction

This document provides an overview of the Public Use Data files associated with analyses of the patterns of health care use over time and across settings (e.g. the timing of emergency department visits and the substitutability of office visits and emergency department visits) from the Oregon Health Insurance Experiment, as well as additional information about the randomization procedure and construction of the analytic samples. It identifies the sources of administrative data and explains the sampling procedure for mail surveys conducted by the study group. A detailed codebook and replication code for portions of Finkelstein et al. (2016) that analyzes the Patterns Public Use Data is also available via http://www.nber.org/oregon/.
Users should consult these guides when using the data.

Contents

Appendix

Project Description

In 2008, a group of uninsured low-income adults in Oregon was selected by lottery to be given the chance to apply for Medicaid. This lottery provides a unique opportunity to gauge the effects of expanding access to public health insurance on the health care use, financial strain, and health of low-income adults using a randomized controlled design. The Oregon Health Insurance Experiment followed and compared those selected in the lottery (treatments) with those not selected (controls).

Public Use Data Files Description

1. "Patterns" (Timing and Joint Outcomes) Data

(File name: **oregonhie_patterns_vars.dta**)

This dataset contains new time-varying variables derived from emergency department (ED) data and Medicaid participation information, and variables from previously published in-person survey and descriptive demographic and lottery list variable datasets that are necessary to replicate select parts of Finkelstein et al. (2016) as a standalone dataset. The unit of observation is the individual with the unique identifier **person_id** (which can be used to link records for each individual to previously published datasets).

a. ED Variables

The Oregon Health Study obtained 2007-2009 visit-level data for all emergency department visits to twelve hospitals in the Portland area. The publicly available ED variables in the Patterns Data are derived from these visit-level data. 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.

The Patterns Data contains only the variables required to replicate the analysis in Finkelstein et al. (2016). Emergency department use for each 180-day period over the 720 days after each individual's lottery notification date is captured by the variable **any_visit_180p_*T*** (where the ***T*** indicates which 180-day window for all variables described here). Note that these variables are measured using slightly different date ranges from those used in our previously-released ED public use data set (which used the initial lottery start date for all individuals, rather than individual-specific draw dates). The analogous variable for the pre-randomization period (defined as December 28, 2006 to March 8, 2008 inclusive) is given by the variable **preperiod_any_visits**. More detail on the date range measures can be found in the codebook of each data set. In addition, for respondents to the in-person interview, we construct a binary variable, **any_inp_match_ed**, indicating "any Emergency Department visit in the last 12 months" based on whether there is an ED discharge record in the 12 months prior to the respondent's in-person interview date in order to study joint outcomes.

All continuous variables (i.e. that capture the total number of ED visits in a given period) and indicator variables (i.e. that capture whether there is any ED visit in a given period) for 60-day increments have been omitted to ensure de-identification. Consequently, any parametric timing analysis and certain non-parametric timing results reported in Finkelstein et al. (2016) cannot be replicated with the publicly available data. For a list of replicable analyses, see note at the top of the replication code: oregonhie_patterns_replication.do.

b. Medicaid Enrollment Variables

Oregon also provided information on Medicaid enrollment histories for each individual on the lottery list. The Patterns Data contains the constructed variables that were based on these data and used to estimate the effect of Medicaid in Finkelstein et al. (2016). In particular, it contains the measures of insurance coverage for 180-day increments used in timing analyses given by **medicaid_all_180p_period_*T*** (where the ***T*** indicates which 180-day window over the 720 days after each individual's lottery notification date), and insurance coverage for in-person survey participants used in joint impact analysis given by **ohp_all_ever_inperson**.

c. In-person Survey Variables

Note that this dataset contains only the variables required to replicate the analysis in Finkelstein et al. (2016). It does not contain all of the information gathered during fielding of the in-person survey. In particular, it includes a binary measure of having any office visit in the last 12 months based on the respondent's self-reported answer to the question "In the last 12 months, about how many times have you seen a doctor or other healthcare professional at a doctor's office, a clinic, or at home?" (**doc_any_incl_probe_inp**), and a binary measure of having any ED visit in the last 12 months based on the respondent's self-reported answer to the question "In the last 12 months, how many times have you gone to an emergency room, or urgent care clinic?" (**ed_any_incl_probe_inp**; used in robustness checks).

The in-person survey was conducted from August 31, 2009 until October 13, 2010. 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). The interviewer script, including skip patterns, is available via ICPSR and also www.nber.org/oregon. The in-person survey variables are only defined for individuals in the Oregon Health Insurance Experiment who were contacted about the in-person survey. Individuals who completed an in-person interview are identified by the variable **sample_inp_resp**, and the date each interview was completed is given by **dt_completed_inp**. Weights that account for the probability of being sampled and survey fielding procedures are given by the variable **weight_total_inp**. These weights are used in all analyses involving the overlap sample (e.g. summary statistics, balance in overlap sample, and joint outcomes analysis). See online appendix on Baicker et al. (2013) for detail on the construction of the weights.

d. Descriptive Variables

The Patterns Data includes demographic information that individuals reported about themselves at the time of lottery sign up (January and February 2008) and information on lottery sign-up and selection.

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_treat** 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, **draw_lottery** is a "matched draw" randomly assigned by the Oregon Health Insurance Study Group to facilitate comparison (see Finkelstein et al. 2012 online appendix for additional details).

Basic demographic information including gender, birth year, whether English was the preferred language of communication, whether an individual signed up on the first 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.

e. Treatment

The treatment of the OHIE was selection by the lottery. Lottery selection gave an individual the ability to apply for Medicaid. The treatment variable is called **treatment** and can be found in the Patterns Data. This variable can be used to estimate intent to treat estimates of the effect of lottery selection.

In addition to estimates of the effect of lottery selection, estimates of the effect of Medicaid coverage may also be of interest. Variables summarizing Medicaid enrollment information include time-varying variables for 180-day increments (**medicaid_all_180p_period_*T***) used in timing analyses, as well as the variable for in-person survey participants (**ohp_all_ever_inperson**) used in joint impact analysis. These variables can be used to estimate local average treatment effect (two-stage least-squares instrumental variable) estimates of the effect of Medicaid coverage.

Selection by the lottery was random conditional on the number of household members on the lottery list (**numhh_list**). It is necessary to control for this in all experimental analyses using indicator variables or another approach**.** Because treatment occurs at the household level, all analyses from the Oregon Health Insurance Experiment cluster standard errors at the household (identified by **household_id**). Both **numhh_list** and **household_id** are available in the Patterns Data.

f. Weights

Like much of our prior OHIE analysis, weights were used in the study of patterns of ED use. For timing analyses, time-varying weights that account for the probability of being selected in the new lottery are given by the variable **weight_*T*days** for analyses at each point in time *T* days after the individual's lottery notification date. See the online appendix for Finkelstein et al. (2016) for detail on the construction of the weights.

For all analyses involving the overlap sample (e.g. summary statistics, balance in overlap sample, and joint outcomes analysis), the "in-person interview weights" are used. The in-person interview weights account for the probability of being sampled and fielding procedures and are given by the variable **weight_total_inp** as in the previously published In-Person Survey dataset. See the online appendix for Baicker et al. (2013) for detail on the construction of the weights.

2. Other Public Use Data

- Descriptive (demographic and lottery list) Variables
- State Program Variables
- Initial Mail Survey
- Six Month Mail Survey
- Twelve Month Mail Survey
- In-person Survey
- Emergency Department Data

Seven separate datasets from the OHIE have previously been made publicly available via the Inter-university Consortium for Political and Social Research (ICPSR 34314) and the study web page (http://www.nber.org/oregon/): Descriptive Variables, State Program, Initial Mail Survey, Six Month

Mail Survey, Twelve Month Mail Survey, In-person Survey, and Emergency Department Data. The unit of observation is the individual: records for each individual can be linked across datasets, including the new patterns dataset, using the unique identifier **person_id** (found in every dataset). No other variables are repeated across these seven datasets.

In addition, we have previously made available code that replicates the portions of Finkelstein et al. (2012), Baicker et al. (2013), and Taubman et al. (2014) that analyze these publicly available data.

References and Replication Code

Finkelstein, Amy et al. "The Effect of Medicaid Coverage on ED Use – Further Evidence from Oregon's Experiment". *New England Journal of Medicine*. 20 October 2016. [VOL No]. (Online Appendix also available)

Code has been made available that replicates the analyses in Finkelstein et al. (2016) which are performed on the publicly available data. The code includes a Stata .do file (oregonhie_patterns_replication.do) which calls on several subprograms. A list of the replicated analyses along with instructions for running the .do file can be found at the top of this master .do file.

Citation for Public Use Data

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

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

Appendix

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). Treatment status is indicated by the variable **treatment**. 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 individuals randomly selected in the lottery who completed the application process and met the eligibility criteria were enrolled in OHP Standard.

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**). For example, individuals could sign up both themselves and their spouses, and both would have equal probability of being chosen. They would thus be twice as likely to win the opportunity to apply for health insurance as someone who signed up without listing a spouse.

Construction of Analytic Samples

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 dataset identified 89,824 individuals. The study group then excluded individuals who gave an address outside of Oregon (and were 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 previously released public use data contains only these 74,922 individuals.

To study Medicaid's impact on patterns of ED use, we use two primary analytic samples: the ED sample and the overlap sample. First, we limited our analytic sample to the individuals residing in zip codes that rely almost exclusively on one of the twelve Portland-area hospitals represented in our emergency department data; specifically, we restrict our analysis sample to individuals residing in zip codes where 98 percent or more of hospital admissions originating in the emergency department were to one of our twelve hospitals. The resulting "ED sample" that we analyze includes 70 zip codes in the Portland metropolitan area in which 24,646 individuals in our study population (or about one-third of the full study population) reside.

Second, we analyze the subset of our study population that is in the intersection of in-person survey respondents (N=12,229) and the ED sample (N=24,646). This intersection is N=10,178. When we restrict the sample to those providing a non-missing answer to the office visit survey question, our final "overlap sample" is N=10,156. We conducted interviews between September 2009 and December 2010. In our overlap sample, the average interview date was April 22, 2010, or about 22 months after notification for individuals in the treatment group. The effective response rate was 73 percent.