Pre-analysis Plan for

The Effect of Access to Citizenship on Immigrant Integration and Health[1]

Vasil Yasenov[a,b], Michael Hotard[a], David D. Laitin[a,c], Duncan Lawrence[a], Jens Hainmueller[a,c,d]

[a]Immigration Policy Lab, Stanford University.
[b]IZA Institute of Labor Economics .
[c]Department of Political Science, Stanford University.
[d]Graduate School of Business, Stanford University.

February 9, 2021

1 Introduction

It is commonly claimed that naturalization has positive impacts for immigrants, but it has been challenging to empirically measure these effects due to issues of self-selection. We will quantify the effects of access to naturalization through an experimental study of a public-private program organized by the New York State Office for New Americans (ONA) that helped low-income immigrants to naturalize. Eligible immigrants in New York could register to enter a lottery to win a voucher that would pay their naturalization fee. We will use follow-up surveys to measure the effect that access to citizenship has had on integration and health. The results will be disseminated in studies focusing on the various outcomes specified below.

For more background on the experiment, details on the lottery assignment, and the survey data we refer the readers to our earlier pre-analysis plan AEARCTR-0006790 aimed at examining economic outcomes using credit bureau data. Some of the text in this pre-analysis plan is reused from this earlier plan and the methodology is essentially the same.

2 Follow-up Surveys

Table 1 below provides the timeline of the experiment and follow-up surveys. Because the citizenship process can take up to a year, we focus on outcomes measured in years two and later. The table shows that we have two year outcomes for all three cohorts, but three and four year outcomes for the 2017 and 2016 cohort, respectively.

3 Empirical Specifications

We will use standard methods for the analysis of randomized encouragement designs to measure the impacts of access to naturalization vouchers on downstream outcomes.

[1]This document borrows some material from a companion pre-analysis plan AEARCTR-0006790.

Table 1: Timeline of the Experiment Registration and Follow-up Surveys

Cohort	Registration	Survey Year 1	Survey Year 2	Survey Year 3	Survey Year 4
2016	Sept 2016	Nov 2017	July 2018	Sept 2019	Nov 2020
2017	Aug 2017	–	Sept 2019	Nov 2020	–
2018	July 2018	Sept 2019	Nov 2020	–	–

We will estimate the intent-to-treat (ITT) effect using the following equation:

$$y_i = \alpha_0 + \alpha_1 VOUCHER_i + \alpha_2 X_i + B_i + \epsilon_i, \tag{1}$$

where y_i is the outcome defined below; $VOUCHER_i$ is a dummy variable for whether or not participant i was offered a voucher; X_i is a vector of pre-randomization control variables and outcomes; B_i is a vector of dummy variables that indicate the randomization block, and ϵ_i is the error term.

To estimate the local average treatment effect (LATE) of citizenship for compliers we will estimate the following equation using two stage least squares:

$$y_i = \beta_0 + \beta_1 CITIZENSHIP_i + \beta_2 X_i + B_i + \epsilon_i, \tag{2}$$

where $CITIZENSHIP_i$ is a binary treatment variable for whether or not a participant reported that he or she attained citizenship. In this equation, $CITIZENSHIP_i$ is instrumented by $VOUCHER_i$ to accommodate non-compliance.

Our specification will examine outcomes measured two-years after the intervention (y_i^{2yr}). Robust standard errors will be used. The coefficients of interest are α_1 and β_1 respectively. Both sets of regressions will also employ block level inverse probability weights to account for the unequal probability of treatment assignment.

We will use two versions of the $CITIZENSHIP_i$ indicator. Our primary measure will be coded whether or not the registrants reported having submitted their citizenship application during the first check-in survey. A secondary version of this variable will measure whether or not the registrants reported having submitted their citizenship application during any survey. We prefer the former because it is measured at roughly the same time interval after the lottery during the first check-in. For the secondary measure we have more surveys for the earlier cohorts, so those participants in earlier years have had more time to complete their citizenship application.

The covariate set X_i will follow the variables used Hainmueller et al. (2018) and include:

- Gender, age, education (high school, some college, college dummies), years since green card, country of origin dummies (Dominican Republic, Colombia and Ecuador), language (English, Spanish dummies), marital status (single, married dummies) and (log) household income.

- Additionally, we will control for the number of days between the date of the registration and the date of the voucher lottery as well as the pre-treatment outcomes measured at baseline (when available).

We will conduct the following additional analyses:

- For some outcomes we have both post-treatment data from the follow-up surveys and the same measures pre-treatment from the registration data. For these outcome variables we will estimate models in which the outcome will be measured in changes between pre- and post-treatment (Δy_i) and the lagged outcome will be removed from X_i.

- For the outcomes for which we have pre-treatment data we will also utilize a difference-in-differences analysis in which we leverage the entire five-year dataset. In particular, we will estimate:

$$y_{it} = \delta_i + \sigma_t + \gamma VOUCHER_i \times POST_{it} + \epsilon_{it},$$

  where the terms δ_i and σ_t represent individual and year fixed effects, and $Post_{it}$ is an indicator for an observation after the voucher randomization. Standard errors will be clustered at the individual level. The coefficient of interest is γ. The advantage of this model is that it leverages the full data by pooling together the short- and medium-term effects of naturalization.

- We will estimate the model on subgroups based on splits of the pre-treatment covariates that we have:
  - Male and female,
  - Spanish and English speakers,
  - Below and above median age,
  - Low (High School only) and high education level,
  - Below and above median household income.

- We will separately estimate the treatment effects for the different cohorts using the longest possible follow-up period.

- For continuous outcome variables such as the IPL-12 index (described below) we will also use quantile regression to examine distributional effects.

4 Outcome Measures

4.1 Integration

We will use the multidimensional integration tool, IPL-12, developed by Harder et al. (2018). It is comprised of a short survey aimed at measuring six dimensions of immigrant integration – social, political, economic, psychological, navigational and linguistic. The six dimensions are measured separately but can ultimately be combined as an average into a single index referred to as the IPL-12 integration score. It is defined on the 0-1 scale and higher scores correspond to higher levels of integration.

- IPL-12 score

- IPL-10 score - the same as IPL-12 but it excludes economic integration because economic and labor market outcomes are analyzed in a separate study (see AEARCTR-0006790).

Additionally, we will examine each of the six dimensions separately. We will analyze the two integration questions for each integration domain (on the 0-1 scale). The score for each domain will be computed on the 1-5 integer scale where, again, higher values indicate higher integration. The outcomes are the answers to the following questions:

- Social Integration
  - Social IPL-12 integration score (0-1)
  - "In the last 12 months, how often did you eat dinner with Americans who are not part of your family?" (1-5)
  - "Please think about the Americans in your address book or your phone contacts. With how many of them did you have a conversation - either by phone, messenger chat, or text exchange - in the last 4 weeks?" (1-5)
- Political Integration
  - Political IPL-12 integration score (0-1)
  - "How well do you understand the important political issues facing the United States?" (1-5)
  - "In the last 12 months, how often did you typically discuss major political issues facing the United States with others?" (1-5)
  - "People like me don't have any say about what the government does in the US." (1-5)
  - Political knowledge score (0-2) - sum of two indicators for correct answers to the following questions:
    - "Are Republicans to the left or right of the Democrats?" (0,1)
    - "Who is the largest party in the Senate?" (0,1)
  - Political IPL-24 integration score (0-1)

  The last three variables are only available as three year outcomes for the 2016 and 2017 cohorts.
- Economic Integration
  - Economic IPL-12 integration score (0-1) (for the questions and scoring see Harder et al., 2018)

  Note that other economic and labor market variables are pre-registered in our companion pre-analysis plan.
- Psychological Integration

  - Psychological IPL-12 integration score (0-1)
  - "How connected do you feel with the United States?" (1-5)
  - "How often do you feel like an outsider in the United States?" (1-5)

- Navigational Integration
  - Navigational IPL-12 integration score (0-1)
  - "In the United States, how difficult or easy would it be for you to see a doctor?" (1-5)
  - "In the United States, how difficult or easy would it be for you to search for a job?" (1-5)

- Linguistic Integration
  - Linguistic IPL-12 integration score (0-1)
  - "I can read and understand the main points in simple newspaper articles." (1-5)
  - "In a conversation, I can speak about familiar topics and express personal opinions." (1-5)

4.2 Health

We will examine outcomes in three categories - mental health, general health and health insurance coverage.

- Mental Health. These are based on the Kessler Psychological Distress Scale (K6) (Prochaska et al., 2012).
  - K6 Score,
  - K6 Score ≥ 5 indicator,
  - K6 Score ≥ 13 indicator.

- General Health
  - "In general, would you say your health is "? (1-5; 1 - Poor, 2 - Fair, 3 - Good, 4 - Very Good, 5 - Excellent),
  - An indicator that the answer to the previous question was "Good", "Very good" or "Excellent".

- Health Insurance. Indicators for having:
  - Any health insurance,
  - Medicaid,
  - Health insurance through an employer.

References

Hainmueller, Jens, Duncan Lawrence, Justin Gest, Michael Hotard, Rey Koslowski, and David D Laitin, "A randomized controlled design reveals barriers to citizenship for low-income immigrants," *Proceedings of the National Academy of Sciences*, 2018, *115* (5), 939–944.

Harder, Niklas, Lucila Figueroa, Rachel M Gillum, Dominik Hangartner, David D Laitin, and Jens Hainmueller, "Multidimensional measure of immigrant integration," *Proceedings of the National Academy of Sciences*, 2018, *115* (45), 11483–11488.

Prochaska, Judith J, Hai-Yen Sung, Wendy Max, Yanling Shi, and Michael Ong, "Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization," *International journal of methods in psychiatric research*, 2012, *21* (2), 88–97.