

Pre-analysis Plan for  
**The Effect of Citizenship on Financial Outcomes**

Vasil Yasenov<sup>a,b</sup>, Michael Hotard<sup>a</sup>, David D. Laitin<sup>a,c</sup>, Duncan Lawrence<sup>a</sup>,  
Jens Hainmueller<sup>a,c,d</sup>

<sup>a</sup>Immigration Policy Lab, Stanford University.

<sup>b</sup>IZA Institute of Labor Economics .

<sup>c</sup>Department of Political Science, Stanford University.

<sup>d</sup>Graduate School of Business, Stanford University.

November 19, 2020

## 1 Introduction

It is often assumed that naturalization has positive impacts for immigrants, but it has been difficult to measure these effects empirically given the self-selection into naturalization. There exists no experimental evidence on this question. We will examine the effects of naturalization through an experimental study of a public-private program organized by the New York State Office for New Americans (ONA). The ONA created NaturalizeNY to help low-income immigrants to naturalize. NaturalizeNY offered vouchers to low-income immigrants in New York that would pay their naturalization fee. The vouchers were randomly assigned via lotteries in 2016, 2017, and 2018 among eligible immigrants that had registered for the program. We will use follow-up surveys and administrative data linkages to measure the effect that American citizenship has had on their financial outcomes.

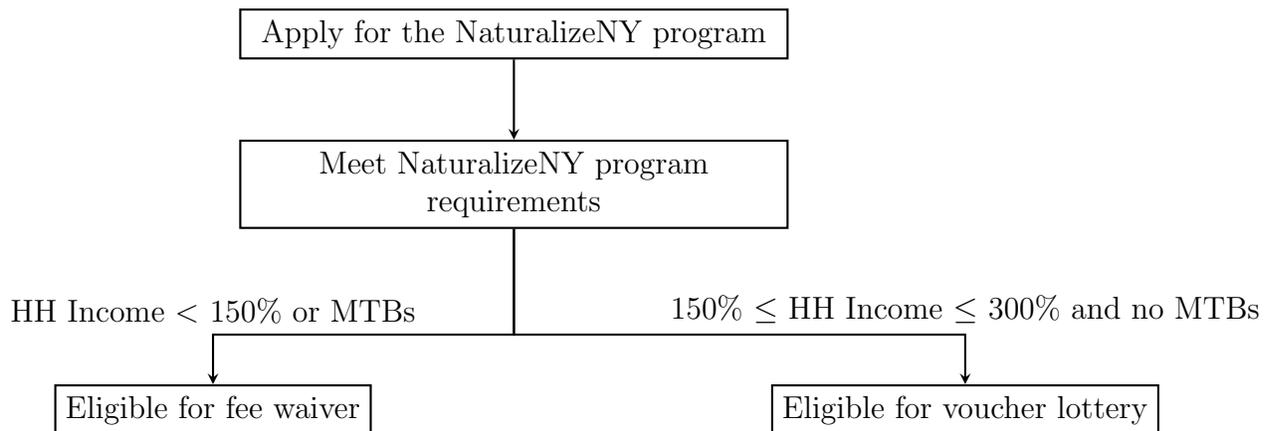
## 2 Study Background

In order to become a US citizen, an immigrant must meet a number of eligibility requirements and complete an application for naturalization. Normally, the application requires a fee. For most of 2016, the cost of submitting a naturalization application was \$680. In December 2016, the cost increased to \$725, although a reduced fee option was introduced that allowed some to pay \$405. For low-income immigrants, these fees can serve as a barrier for seeking naturalization [1]. In addition, very low-income immigrants whose household income falls at or below 150% of the Federal Poverty Guidelines (FPG) or who receive means-tested benefits (MTBs) from the government were eligible for a federal fee waiver that eliminates the application fee [3]. Research has shown that informing low-income immigrants about their eligibility for the fee waiver can increase naturalization rates [2].

NaturalizeNY was a public-private program operating in New York State that promoted naturalization among low-income legal permanent residents who are eligible to naturalize, but may face financial barriers preventing them from doing so. The program provided information on the naturalization process as well as naturalization application assistance through the Office for New Americans' network of over 20 Opportunity Centers (OC) located

throughout the state. The program was offered in 2016, 2017, and 2018. Each year, a publicity campaign in the spring and summer encouraged immigrants to sign up for the program. Eligible immigrants could register for the NaturalizeNY program by completing a registration tool that captured their basic demographic and immigration information. To be eligible for the program immigrants had to: (1) be 18 years or older; (2) reside in New York State; (3) have a household income that falls at or below 300% of the FPG; and (4) demonstrate eligibility for naturalization using an N-400 form. During the registration, the eligible registrants are sorted into two groups:

- If a participant’s household income was below 150% of the FPG or they received MTBs, then the registration system informed them that they potentially qualify for a fee waiver from the federal government. We refer to this group as the fee waiver group. In a separate experiment, different behavioral nudges were randomly assigned among this group [2].
- If a participant’s household income was between 150% and 300% of the FPG and they did not receive MTBs, then the registration system entered them into a lottery for a chance to receive a voucher that paid the fee for their naturalization application. We refer to this group as the voucher lottery group. This group is the focus of this study.



### 3 Experimental Design

The NaturalizeNY lottery was run in 2016, 2017, and 2018. In each year, immigrants registered for the program and during the registration process their eligibility for the voucher lottery was determined. Registrants who met the program requirements were deemed eligible for the voucher lottery. After the registration period closed, these registrants were randomly assigned to one of two groups. If a participant was assigned to the treatment group, he or she received a voucher that paid the full cost of applying for citizenship (\$680 in 2016, \$405 or \$725 in 2017 and 2018 depending on the immigrant’s household (HH) income). The voucher was processed by a specific Office for New Americans OC in New York. The fee voucher was directly paid to the United States Citizenship and Immigration Services (USCIS) by the OC

and could not be used for any other purpose than to pay for the naturalization application. Participants assigned to the control group did not receive a fee voucher.

Each year, the voucher randomization was conducted within blocks. In particular, eligible registrants were assigned to one randomization block based on their geocoded street address provided during the registration and the type of voucher that was needed (whether a person would likely need a full or partial voucher). The geographic blocking was conducted to minimize the distance that lottery winners would have to travel to get their vouchers processed at an OC. A lottery was conducted in all blocks where the demand for the vouchers exceeded the number of available vouchers. The blocks that met this criterion were New York City and Long Island. Only registrants from these blocks are included in the experiment. Table 1 below shows the number of immigrants randomly assigned into the treatment and control groups for each block and year.

The experiments were approved by the Stanford University IRB protocol #34554.

Table 1: Sample Size by Randomization Group

Cohort	Block ID	Control (N)	Treatment (N)
2016	1	6	23
2016	2	11	18
2016	3	8	20
2016	4	502	248
2017	5	93	168
2017	6	145	235
2017	7	8	38
2017	8	10	49
2018	9	217	244
2018	10	295	305
2018	11	20	41
2018	12	45	53

## 4 Timeline and Data

### Registration

The study uses various data sets. The first one is the registration data. During registration for the program, participants provided their background and demographic characteristics, contact information, immigration history, language abilities, and the data necessary to determine their program eligibility. They also consented to participate in the research study and have their information linked with administrative data for research purposes.

In 2016, the registration window was from July 11, 2016, to September 23, 2016. The vouchers were randomly assigned after the closing of the registration in the following week. Voucher winners were notified the week of October 23, 2016. In 2017, the registration window was from September 24, 2016, to July 28, 2017. The vouchers were randomly assigned after

the closing of the registration in the following week. Voucher winners were notified the week of August 7, 2017. In 2018, the registration window was from July 29, 2017, to July 3, 2018. The vouchers were randomly assigned after the closing of the registration in the following week. Voucher winners were notified the week of July 10, 2018.

## Surveys

Each cohort was scheduled to receive two check-ins and two in-depth follow-up surveys. (Additional follow-up surveys may be conducted if funding is available.) The check-ins were used to measure the first stage effect of the intervention – whether the financial assistance increased naturalization rates, and maintain up-to-date contact information. The in-depth surveys were scheduled for one and two years after the voucher winners will likely have naturalized. Because naturalization can take an additional six to twelve months after an application is submitted, the timing for the surveys does not align perfectly with the closing of the registration for a given cohort.

Table 2 provides information about the timeline of the surveys.

Table 2: Timeline of NaturalizeNY Registrations and Surveys

Cohort	Event	Date
Cohort 2016	Voucher winners notified	September 2016
Cohort 2016	Check-in 1	March 2017
Cohort 2016	Check-in 2	November 2017
Cohort 2016	Survey 1	July 2018
Cohort 2016	Survey 2	September 2019
Cohort 2017	Voucher winners notified	August 2017
Cohort 2017	Check-in 1	July 2018
Cohort 2017	Check-in 2	May 2019
Cohort 2017	Survey 1	September 2019
Cohort 2017	Survey 2	November 2020
Cohort 2018	Voucher winners notified	July 2018
Cohort 2018	Check-in 1	May 2019
Cohort 2018	Check-in 2	September 2019
Cohort 2018	Survey 1	November 2020
Cohort 2018	Survey 2	September 2021

## Credit Data

Data from a credit bureau will be used to measure the financial situation of registrants. The credit report data will include information on consumer credit and borrowing, such as the number and type of credit lines that they have, the number and types of delinquencies reported, and their average credit balance over a period of time. In addition to their credit history, we will also obtain historical credit scores and an income estimate from a proprietary

prediction algorithm. In the research consent form, participants agreed to allow the research team to link to administrative records. The credit bureau matches the records based on the participant’s reported name, address, and date of birth using a proprietary matching algorithm. The bureau only returns a match if it deems that there is a high probability that the record provided matched their records. A record may not match if the information provided by the participant during registration was incorrect or if the participant does not have a credit file. Because the data match is done on historical snapshots and not current data, it will not have any impact on participants’ credit scores or history.

The data files returned by the credit bureau will be fully de-identified so that the research team is unable to link any record to a study participant. It will include a sparse set of treatment indicators and covariates that will be used in statistical models, but that will not allow us to identify any individual participant.

The credit data includes five yearly snapshots of the financial situation for each registrant, including one month before the lottery, and approximately one and two years after the lottery.

## 5 Model Specifications

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

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

$$y_i = \alpha_0 + \alpha_1 VOUCHER_i + \alpha_2 X_i + B_i + \epsilon_i, \quad (1)$$

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  $\epsilon_i$  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:

$$y_i = \beta_0 + \beta_1 CITIZENSHIP_i + \beta_2 X_i + B_i + \epsilon_i, \quad (2)$$

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.

We will estimate models using both two-year ( $y_i^{2yr}$ ; primary specification) and one-year ( $y_i^{1yr}$  secondary) outcomes. Robust standard errors will be used. The coefficient of interest are  $\alpha_1$  and  $\beta_1$  respectively. Both sets of regressions will also employ block level inverse probability weights to account for the unequal probability of treatment assignment. Note that, in a previous study, we found that the voucher increased the likelihood of submitting a citizenship application by approximately 40% [1].

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 include a variable for English-language ability at the time of registration, a coarsened household income variable (above or below the median household income per capita reported during registration), and the lagged outcomes that we will construct from the credit data. We will use multiple imputation to deal with missing data.

In addition, we will conduct the following secondary analyses:

- We will estimate models in which the outcome will be measured in changes between pre- and post-treatment ( $\Delta y_i$ ) and the lagged outcome will be removed from  $X_i$ .
- We will estimate the model on subgroups based on splits of the pre-treatment covariates that we have - HH income (above/below median) and English language ability (yes/no).
- We will 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  $\delta_i$  and  $\sigma_t$  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  $\gamma$ . Note that this model pools together the short- and long-term effects of naturalization.

- We will separately estimate the treatment effects for the different cohorts using the longest possible follow-up period for each to differentiate long- from short-term effects.
- We will use quantile regression to examine distributional effects on credit score and income.

## Outcomes

We will use a set of outcomes ( $y_i$ ) that measure the following key concepts:

- Credit score. The credit bureau provides both the Vantage (primary measure) and the FICO (secondary) credit scores. We will use the scores individually as well as create an index of both.
- Income. This is an estimate also provided by credit bureau. We will use a logarithmic transformation of this variable.
- Access to credit. We will use two types of variables. The first type measures whether the person has access to credit, and includes indicators for having a credit score, at least one open line of credit and a “thick” file (3 or more open lines of credit). We will use these variables individually and also build a composite index. The second type measures the amount of credit to which a person has access using the log of the total credit line of open revolving trades.

- Financial distress. This includes variables for adverse financial events, such as delinquency (number of trades 30+ days past due) and collection (number of third party collections). The delinquency variable will be top-coded at 5 or more, while the collection variable will be top-coded at 3 or more. We will examine these events separately as well as combine them into a composite index. We will not use measures of bankruptcy or liens because they are extremely rare in this sample.

## References

- [1] Jens Hainmueller, 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*, 115(5):939–944, 2018.
- [2] Michael Hotard, Duncan Lawrence, David D Laitin, and Jens Hainmueller. A low-cost information nudge increases citizenship application rates among low-income immigrants. *Nature human behaviour*, 3(7):678–683, 2019.
- [3] Vasil Yassenov, Michael Hotard, Duncan Lawrence, Jens Hainmueller, and David D Laitin. Standardizing the fee-waiver application increased naturalization rates of low-income immigrants. *Proceedings of the National Academy of Sciences*, 116(34):16768–16772, 2019.