

Final Assignment

PP41600 Survey Research Methodology

Group 5: Greta Ritzenthaler, Lindsay Liebert, Li Liu, Stormy Kim

Nonresponse in ACS Citizenship Question: Characteristics of Respondents

1. Summary	1
2. Policy Focus and Significance	1
3. Data and Relevant Universe	2
4. Survey Methods and Data Collection	4
5. Analysis Plan	9
7. Statistical Analysis of Results	12
6. Analysis Results	13
8. Conclusion	17
Appendix	20
Reference	26

1. Summary

This research project aims to answer the following question: what are the demographic characteristics of the people who do not respond to the citizenship question in the ACS survey? The selected variables for comparison are income levels, family size, industry of employment, and educational attainment level. Our hypothesis is that people with non-responses on the citizenship question are more likely to have lower income, larger family size, work in the service sector, and less education. Most non-responses are likely the vulnerable group of people residing in the United States who fear that their truthful responses to the citizenship question may incur negative legal measures such as deportation.¹ In addition, using the IPUMS data from the American Community Survey between (ACS) 2013 and 2017, we will observe the relationship between the above-mentioned variables and citizenship non-response in five U.S states with a large immigrant population.

2. Policy Focus and Significance

The disparity in socioeconomic statuses between citizens and immigrants in the United States has been a major policy challenge for the government. As the data collected by the Census Bureau are used for all facets of policy design such as welfare, education, employment, and healthcare, an accurate measure of the Americans' socioeconomic characteristics is directly connected to tackling the socioeconomic gap between different groups of people.² Since the non-responses are primarily tied to underrepresented minorities and immigrants, the results from the analysis will serve as an indirect measure of the economic and social variables of

¹ Jordan, Miriam. "If Census Asks About Citizenship, Some Already Have an Answer: No Comment." *The New York Times*, The New York Times, 27 Mar. 2018, www.nytimes.com/2018/03/27/us/census-undocumented-immigrants.html.

² "Read 'Modernizing the U.S. Census' at NAP.edu." National Academies Press: OpenBook, 1995, www.nap.edu/read/4805/chapter/16.

non-citizens.³ Policymakers can incorporate the findings into effective social welfare programs for the immigrant population whose policy needs are often overlooked.

Understanding the socioeconomic information of the non-respondents of the ACS citizenship question is important for two reasons. First, the results from our findings will provide insight on the potential effects of adding the citizenship question in the decennial census of 2020. On March 2018, the Commerce Secretary Wilbur Ross approved plans to add the question, “Is this person a citizen of the United States?”, which is a decision supported by President Trump.⁴ Many fear that asking about citizenship could scare non-U.S citizens from participating in the census, which will affect data quality and under-represent the immigrant population. As the ACS already has a question on citizenship status, the result of examining the relationship between nonresponse rates and demographic characteristics serves as a source to estimate the potential effects of including the citizenship question in the census. Second, descriptive analyses on various characteristics of nonresponse survey participants will shed light on state-level differences in characteristics of non-U.S citizens, assuming non-responses are likely non-U.S citizens (a claim supported by a Georgetown study in 2018).⁵

3. Data and Relevant Universe

The source of data is the Census Bureau’s American Community Survey (ACS), available on the IPUMS website. The ACS is chosen primarily as it is the model for the decennial census. In addition, the ACS has a section on U.S citizenship status and demographic variables of interest:

³ Ahlmark, Nanna, and Maria Algren. “Survey Nonresponse among Ethnic Minorities in a National Health Survey – a Mixed-Method Study of Participation, Barriers, and Potentials.” *Taylor and Francis Online*, Oct. 2014.

⁴ Wang, Hansi Lo. “Trump Officials Prepared For Supreme Court Fight Over Census Question.” NPR, NPR, 12, Sept., 2018, www.npr.org/2018/09/12/646936577/trump-officials-prepared-for-supreme-court-fight-over-census-question.

⁵ O'hare, W. (2018). *Citizenship Question Nonresponse: A Demographic Profile of People Who Do Not Answer the American Community Survey Citizenship Question*. <http://www.georgetownpoverty.org/wp-content/uploads/2018/09/GCPI-ESOI-Demographic-Profile-of-People-Who-Do-Not-Respond-to-the-Citizenship-Question-20180906.pdf>

household income, household size, industry of occupation, and educational attainment level. It collects data through a wide range of methods: mail questionnaire, telephone survey, online questionnaire, and in-person interview. We observe the nonresponse rates between 2013 and 2017 because ACS changed its data collection methodology in 2012. The Census Bureau reduced its use of the Failed Edit Follow-Up (FEFU) operation, which involves a follow-up phone interview with internet and mail respondents whose responses indicate a discrepancy and those who indicated a household size of more than six people. This change led to an increase in blank responses. Moreover, as the ACS implemented an internet response option in 2013, this project will analyze data collected from 2013 for consistency and fair comparison. The different nonresponse rates for the citizenship question in ACS in 2016 are as follows: Internet: 8%; Mail: 6.7%; Personal interview (phone or computer-assisted interview): 3.8%.⁶ Data collected through the internet is associated with a higher non-response rate. Since our variable of interest is nonresponse rates, we plan to look at data collected from 2013 to avoid an upward bias in our results.

ACS covers many harmonized social and economic variables that are relevant to our research question. “CITIZEN” variable indicates the citizenship status of survey respondents who are foreign-born. “QCITIZEN” is the data quality flag for the “CITIZEN” variable which includes the data on nonresponse. The main demographic variables for our analysis include “INCTOT” (Total personal income), “FAMSIZE” (Family unit membership), “IND” (Industry), and “EDUC” (Educational attainment). Five states are chosen based on their relatively high immigration population and nonresponse rates for the citizenship question: Illinois, California, Arizona, Texas, and New York.⁷

⁶

⁷ Krogstad, Jens Manuel, and Michael Keegan. “15 States with the Highest Share of Immigrants in Their Population.” *Pew Research Center*, Pew Research Center, 14 May 2014, www.pewresearch.org/fact-tank/2014/05/14/15-states-with-the-highest-share-of-immigrants-in-their-population/.

4. Survey Methods and Data Collection

The ACS is an annual, cross-sectional survey conducted year round to obtain information on all individuals residing in the United States. The population from which the sample is drawn consists of all housing units and group quarters within the United States. Group quarters include institutionalized and non-institutionalized peoples. Table 1 presents a short list of entities that fall under the definition of group quarters by the Census Bureau.

Table 1. Census Bureau Definition for Group Quarters⁸

Institutional	Non-Institutional
Correctional facilities	College dormitories
Nursing homes	Military barracks
Mental hospitals	Group homes
	Missions
	Shelters

The sample frame referenced for the ACS is the Master Address Frame (MAF), consisting of all residential units initially compiled from the 2000 Census and select non-residential units in the United States and Puerto Rico. Updates made post-2000 include information from 2010 Census field operations and other activities. The final sample frame includes filters and edits to the MAF to reduce the amount of overcoverage and undercoverage in the survey.⁹ This methodology has led to high coverage rates over the past 5 years. It is worth noting that coverage rates for group quarters are lower than housing units over the past 5 years. Furthermore, the coverage rate for group quarters has been declining over the past 5 years, while the coverage rate for housing units has remained constant around 98% (Table 2).

⁸ Census Bureau. "Group Quarters/Residence Rules."

<https://www.census.gov/topics/income-poverty/guidance/group-quarters.html>

⁹ Torrieri, N. "American Community Survey Design and Methodology" *Census Bureau*, June 2014.

Coverage Rates (in percent)

	Housing Units	Group Quarters Population
Year	Total	Total
2017	97.9	88.4
2016	98.3	89.5
2015	98.6	90.6
2014	98.4	91.1
2013	98.8	93.6

Table 2. Coverage Rates for ACS from 2013 - 2017¹⁰

A complex sampling design is implemented to ensure accurate representation of the United States population in the ACS. The sample frame is first divided into housing units and group quarters. Housing units are separated into sixteen different strata and assigned stratum sampling rates. Then, housing units are systematically assigned to one of five sub-frames to ensure that housing units are only sampled once every five years. Finally, housing units to be surveyed are systematically selected from the sub-frame.¹¹

Two strata are created for group quarters which are classified as small and large. Facilities in the small stratum are sampled similar to the housing units as all individuals in selected facilities are surveyed. For the large stratum, residents in group quarters are divided into groups of ten. A systematic sample of these groups is chosen and within the selected groups, one individual from each group is randomly chosen to be surveyed.¹²

Surveys are conducted almost daily. Data collection for each sample lasts nearly three months. Primary modes of data collection are personal visits, phone interviews, and mail/internet interviews.¹³ The response rate from the ACS is consistently above 90% with the exception of

¹⁰ Census Bureau. "American Community Survey - Coverage Rates."

<https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/coverage-rates/>

¹¹ Torrieri, N. "American Community Survey Design and Methodology" *Census Bureau*, June 2014.

¹² Torrieri, N. "American Community Survey Design and Methodology" *Census Bureau*, June 2014.

¹³ Ibid.

2013. This low rate is due to a government shutdown in 2013, prohibiting the ACS from conducting follow ups with nonresponders. Additionally, it is important to note that in 2017 hurricanes impacted data collection and resulted in higher nonresponse rates in Texas, which may impact our findings.¹⁴ Response rates for 2013-2017 appear in Table 3.

	Housing Unit	Group Quarters
Year	Response Rate	Response Rate
2017	93.7	94.7
2016	94.7	95.7
2015	95.8	95.3
2014	96.7	95.9
2013	89.9	95.2

Table 3. Response Rates for ACS from 2013 - 2017¹⁵

Although response rates are high, there are still missing items in the data collected. ACS utilizes two imputation methods: assignment and allocation. Items imputed by assignment use other data from the respondent to infer a missing response. For example, if a respondent answers “yes” to having given birth in the last year, then ACS infers sex as female. The other method is hot-deck allocation, which utilizes a matrix of data from prospective donors (other respondents) to impute a response for the missing item. To identify imputed values, ACS incorporates quality flags into the data.¹⁶

Since the goal of ACS data is to accurately represent the U.S. population, ACS implements a complex weighting scheme to ensure estimates are consistent with official Census Bureau demographic estimates. The basic approach is to assign two sets of weights to each person unit—a weight to each individual sample record and a weight to each housing unit sample record. Weights for housing units and group quarters are calculated differently. Weighting for

¹⁴ Census Bureau. “American Community Survey - Response Rates.”
<https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/>

¹⁵ Ibid.

¹⁶ Torrieri, N. “American Community Survey Design and Methodology” *Census Bureau*, June 2014.

group quarters utilizes a three stage approach. The initial weight is based on the sample rate. An adjustment is then made to ensure substate totals match sampling frame information. Finally a post-stratification weight is included based on state geography. Weighting for housing units also utilizes a three stage approach where the initial weight is based on the sampling rate. An adjustment is made for nonresponding units, and finally weights are updated so that demographic estimates align with the latest Census data.¹⁷

Our dataset focuses on adults aged 16 and older in Illinois, California, Arizona, Texas, and New York from 2013 to 2017. We will utilize person weights, as we are examining individual nonresponse. We restrict the data to persons aged 16 and older to be in line with current DACA regulations.¹⁸ Furthermore, the universe for one of our main variables of interest ("IND" - industry) restricts to persons aged 16 and older. A list of our variables of interest, their questionnaire wording, and final definitions follow. All of these variables will cover our chosen universe, adults aged 16 or older.

Citizenship status: This is a categorical variable which determines whether a person is a citizen, non-citizen, or non-respondent. The universe is "foreign-born persons". We have to re-code the citizenship variable as given by IPUMS to recreate the original non-response rate. In order to do this, we use the citizen quality flag, QCITIZEN, which marks whether IPUMS had to impute these responses to create their final citizenship variable. Their methods are not entirely known; therefore we recreated the process based on our own assumptions. If the response is missing (ie. their observation has a quality flag), but the individual was born in the US, we code them as a citizen. This is a common non-response question due to the order of

¹⁷ Ibid.

¹⁸ Napolitano, J. (June 15, 2012). *Exercising prosecutorial discretion with respect to individuals who came to the United States as children* [Memorandum]. Washington, D.C.: Dept of Homeland Security.
<https://www.dhs.gov/xlibrary/assets/s1-exercising-prosecutorial-discretion-individuals-who-came-to-us-as-children.pdf>

questions in the questionnaire. Respondents are first asked, “Where was this person born?” If the individual was born in the US, it is likely that interviewers and respondents skipped the following question, “Is this person a citizen of the United States?” If the response is missing and they were born abroad, we officially mark them as a non-response for our final analysis. Respondents who do not have a quality flag are coded into citizens and non-citizens.

Total personal income: This continuous variable measures an individual's own pre-tax income or losses. The universe is the group of persons older than 15. The extreme values are bottom and top-coded, according to each state's median. The ACS questionnaire asks the respondent about various sources of income from the past 12 months, such as wages, salary, interest, social security, or welfare; they are then prompted to add up these sources and subtract any losses. Our study does not require any re-coding for this measure.

Family size: This continuous variable counts the number of family members (related by blood, marriage, or adoption) living with the respondent in their household. The universe is “all persons”. Those living by themselves take the value 1. This measure is top-coded at 20. There is no explicit question for this variable; it is inferred by the question “How is this person related to Person 1?” which considers the relationship that person has with the head of the household. There is no re-coding required for this measure.

Educational attainment: This categorical variable measures the highest level of education attained by the respondent. The universe is “all persons”. The ACS questionnaire asks, “What is the highest degree or level of school this person has COMPLETED?” Respondents who are still attending school will answer with the highest grade they've completed so far. Some minor re-coding is required to group grades into categories such as “some high school,” “Bachelor's degree,” or “any graduate degree.”

Industry: This categorical variable designates which industry an individual works in, either currently or at her most recent job. The universe is “persons who are older than 16 and had worked within the previous five years”. Those who work in multiple industries are asked to pick the one in which they spend the most time or make the most money. There is an extensive list of industry codes provided by IPUMS which is too large to draw any conclusions from. We re-coded this variable by condensing the categories into broader industry groups. For example, under the industry category “Agriculture, Forestry, Fishing and Hunting, Mining”, 70+ classifications exist. Instead of using the more detailed level of classification, we aggregated them into one category. Thus, people working in the aforementioned 70+ job categories will be grouped into one large industry category.

A full table of summary statistics can be found as table 1a in the appendix.

5. Analysis Plan

There are four group categories under the citizenship-question:

- 1) Group 1: Noncitizen-response
- 2) Group 2: Noncitizen-nonresponse
- 3) Group 3: Citizen-response
- 4) Group 4: Citizen-nonresponse

The primary group of interest is non-citizens who did not respond to the citizenship question of the ACS. The four groups are filtered by using the quality flag variable “QCITIZEN”. By employing various statistical tests, this study observes whether there are significant differences in means or odds among groups for four major demographic variables:

- 1) Educational attainment level: this variable is a categorical variable as respondents are asked about whether they complete high school or college rather than years of education. We recode each education level into a binary variable and run logistic regressions to observe the odds ratios of the noncitizen-nonresponse group compared

to the two response groups. Our hypothesis is that non-citizen nonresponse individuals will have a higher odds ratio for “less than high school diploma” and a lower odds ratio for educational categories beyond high school completion. In other words, β_1 is less than 1, indicating the “noncitizen-nonresponse” group’s lower odd of having a high school diploma than the comparison group. The general form of the logistic regression model is: $Pr(Edu - level = 1 | X) = \frac{Pr(X)}{1 - Pr(X)} = e^{\beta_0 + \beta_1 X}$, where X is the indicator for whether belonging to the nonresponse group or not.

- 2) Family size: We use linear regression model to test if there are significant differences in the mean values of family size. We hypothesized that “noncitizen nonresponse” group will have the largest family size than the other groups. In other words, to reject the null hypothesis that there are no difference in family size, β_2 has to be larger than the other three coefficients, β_1 , β_3 , and .

$$y = \beta_1 * Group\ 1 + \beta_2 * Group\ 2 + \beta_3 * Group\ 3 + \beta_4 * Group\ 4 + \epsilon$$

Hypothesis 1 (Compare between non-citizens groups): $\beta_2 > \beta_1$

Hypothesis 2 (Compare between non-responses groups): $\beta_2 > \beta_4$

Hypothesis 3 (Compare between group 2 and group 3): $\beta_2 > \beta_3$

- 3) Income: We will use the linear regression model to compare the coefficients of four group categories and observe whether the total income differences are statistically significant across groups.

$$y = \beta_1 * Group\ 1 + \beta_2 * Group\ 2 + \beta_3 * Group\ 3 + \beta_4 * Group\ 4 + \epsilon$$

Hypothesis 1 (Compare between non-citizens groups): $\beta_2 < \beta_1$

Hypothesis 2 (Compare between non-responses groups): $\beta_2 < \beta_4$

Hypothesis 3 (Compare between group 2 and group 3): $\beta_2 < \beta_3$

- 4) Industry of employment: We select the top hiring industries from the industry categories. Then we recode the variable so that each industry becomes a binary variable. For

instance, under the food service industry, an individual working in this sector will take on a value of 1, while respondents employed in other industries are assigned 0. As the citizen group and noncitizen group has different industry employment patterns, we compare the response β_1 and nonresponse groups within the same citizen category. For example, when $X=1$ for nonresponse noncitizens and 0 for response citizen, the β_1 coefficient shows the odds ratio of the nonresponse non-citizen's likelihood of working in a particular industry compared to the noncitizen-response person. We hypothesised that there is no significant difference within the citizen groups in terms of industry employment patterns. For the noncitizen groups, we hypothesised that the nonresponse group will have a higher probability of working in the low-skilled service industry.

The general form of the logistic regression is:

$Pr(\text{Specific industry} = 1 | X) = \frac{Pr(X)}{1-Pr(X)} = e^{\beta_0 + \beta_1 * X}$, where X is the indicator for whether belonging to the group or not.

We use a p-value threshold of 0.01, indicating a 99% confidence level to deem the differences to be statistically significant. Here are some inferential limitations in our analysis:

- 1) The chosen time frame is from 2013 to 2017. Because of the data collection change in 2013, we decided not to include data collected before the change for consistency. By limiting the analysis to the short window of 5-year , we cannot observe the cross-time changes in response rates and characteristics.
- 2) The four groups are not balanced. The noncitizen-nonresponse group only accounts for 2.38% of the whole sample. Thus this group has a much higher standard error than the other three group categories. This group might be underrepresented because many are undocumented.

- 3) For our analysis on employment industry, we only look at the most common industry categories defined by IPUMS. Within one industry category, there can be various occupations that require different skills. However, by dissecting the industry categories into micro-levels, we would lose the precision in our inferences especially for the noncitizen-nonresponse group that is relatively small.

Table A: Number of responses and non responses for noncitizens in Illinois, California, Arizona, Texas, and New York from 2013 to 2017

	2013	2014	2015	2016	2017
Response	83,380	82,525	82,488	80,467	79,224
Nonresponse	15,231	16,487	16,784	18,579	19,408

Source: IPUMS data from the American Community Survey between 2013 and 2017.

7. Statistical Analysis of Results

All the tables are included in the Appendix. Here are some information for the analytic statistics:

1. The data source for all following analyses is IPUMS data from the American Community Survey between 2013 and 2017.
2. The standard errors for all the following analytic statistics are calculated using the first-order Taylor series linear approximation method with Stata version 15.
3. The unweighted observations adds up to 3,648,170 for the selected five states. The counts of unweighted observations for each categorized group are:
 - a. Noncitizen-response : 408,084
 - b. Noncitizen-nonresponse : 86,489
 - c. Citizen-response : 2,976,493
 - d. Citizen-nonresponse : 177,104

Most of our tables have the unweighted observations as stated above and a population size of 366,302,616. Some tables that only look at American citizens or noncitizens have a different number of observations and population size which will be indicated in the upper right side of each table.

6. Analysis Results

State-level Analysis on Response Rates

The five selected states may display different survey response patterns. Table 2a shows that California and New York are the two states that have a noncitizen-nonresponse proportion above 3%. Arizona had the highest citizen-nonresponse proportion at 6.6%, while other four states were below 5%. Table 3a depicts the results of a logistic regression on noncitizen-nonresponse to examine whether the difference in the nonresponse rates are statistically significant. The results indicate that Arizona, having the highest odds ratio, has the odds of having a noncitizen-nonresponse 58% higher than the that of the other four states. A close second, California's odds of having a noncitizen-nonresponse rate is 52% higher than that of other four states. New York's odds of having a noncitizen-nonresponse rate is 47% higher than that of other four states. Texas, had the lowest odds ratio of .058. While a simple tabulation might have suggested that Arizona's nonresponse rate is not as significant as that of California and New York, due to the two states' much larger sample size, the logistic regression has indicated that the probability of noncitizen-nonresponse is actually higher in Arizona.

We found there is a statistically significant positive correlation between immigrant population size and noncitizen nonresponse. However, just because a state has a larger immigrant population proportion does not mean it will have the corresponding increase in nonresponse rates, since

There are other unobservable factors that influence a non-citizen's decision to not respond to the citizenship question, such as non-U.S citizen's perception of risk related to answering the citizenship question. An immigrant residing in Arizona may suffer from higher expected losses of responding to the citizenship question than one residing in California or New York where the

immigrant population are significantly larger and attitudes toward noncitizens tend to be more liberal. This is beyond the scope of this study.

Education

Table 4a suggests that within the citizen group, 21.2% of those who did not respond to the citizenship question didn't even have a high school diploma while the percentage was 13.21% for those who answered the question. Table 5a depicts that within the noncitizen group, 41.31% of those who responded to the citizenship question had less than a high school diploma while only 29.73% of the nonresponse group did not graduate from high school. Contrary to our hypothesis of the noncitizen-nonresponse group having the lowest average educational attainment level, the noncitizen-response group was reported to have the lowest education among the four groups. However, while the noncitizen-nonresponse group has a higher proportion on high school diploma, some college, and bachelor's degree, it had a lower likelihood on having graduate degree by 1.5% than the noncitizen-response group.

The logistic regression results from Table 6a indicate that the odds ratio calculated for each educational level is highly statistically significant. The odds of having less than a high school diploma for nonresponse noncitizens was 1.49 times the odds of citizens and noncitizens who responded to the question. In other words, the odds for the noncitizen-nonresponse group is 49% higher than the odds for the response group. The odds of having a high school diploma for the noncitizen-nonresponse group is 1.11 times the odds of the response group. The odds for the noncitizen-nonresponse group is 12% lower than the odds for the response group for some college. The odds for the noncitizen-nonresponse group is 18% lower than the odds for the response group for a bachelor's degree. Lastly, the odds of having a graduate degree for the noncitizen-nonresponse group is 34% lower than the odds for the response group. As the

education attainment level increases, the likelihood for the noncitizen-nonresponse group to fall under the given education level decreases gradually.

Family Size

Our hypothesis was that the noncitizen-nonresponse group would have the largest family size. According to Table 7a, the mean family size for noncitizen-nonresponse group is 3.69, which is only 0.1 smaller than the mean of the noncitizen-response group. The difference is statistically significant at a 99% confidence level. Interestingly, within the citizen group, the response group had a smaller average family size by 0.4, and the family size difference between the response and nonresponse groups among the citizens was also statistically significant. One issue is that the responses might be underestimated as people might break off on the 4th, 5th or 6th person.

Employment Industry

The most common industry of employment for citizens between 2013 and 2017 was “Educational Services, Health Care, and Social Assistance”. There was no difference in the second and third most common industries for citizens who responded and didn’t respond to the citizenship question. The second most popular industry was “Agriculture, Forestry, Fishing and Hunting, and Mining” and the third was “Retail Trade”.

For non-citizen groups, there was a slight difference in the ranking of the most common hiring industries. The three most common industries of employment for non-citizens who responded to the citizenship question between 2015 and 2017 were “Agriculture, Forestry, and Hunting”, “Professional, Scientific, and Management, and Administrative, and Waste Management Services”, and “Arts, Entertainment, and Recreation, and Accommodation and Food Services”. For non-citizens who did not respond to the citizenship question, the first industry is identical to the previous group while their second and third most popular industries were “Educational

Services, and Healthcare and Social Assistance” and “Professional, Scientific, and Management, and Administrative, and Waste Management Services”.

Table 10a depicts the probability differences for each industry employment between the noncitizen-nonresponse group and the noncitizen-response group are all statistically significant. Noncitizens who did not respond to the citizenship question were 61.7% more likely to be in the “Educational Services and Healthcare and Social Assistance” than the non citizens who did respond. The odds ratio of noncitizen-nonresponse group for the other three industry categories were lower than that of the noncitizen-response group.

Total Income

According to Table 11a, the difference in the average income for four groups are statistically significant. The linear regression analysis entailed that we can reject the null hypothesis of no difference in mean income across groups at a 99% confidence level.

The “citizen-response” group has the highest annual total income of \$41,476. The “citizen-nonresponse” group, while being the second highest, has a total income that is lower by \$10,000 compared to the “citizen-response” group. The total income difference between the two nonresponse groups is only around \$1,000. While the noncitizen-nonresponse group has the second lowest average total income among the four groups, its mean income is \$5,300 higher than the noncitizen-response group. One possible explanation is that those noncitizens with higher income have a higher opportunity cost of losing their livelihoods and facing possible deportation by answering the citizenship question. Thus, a higher income for noncitizens may signal a higher nonresponse pattern for the citizenship question in a questionnaire.

8. Conclusion

Using the ACS survey data from 2013 and 2017, we analyze the demographic characteristics of noncitizens who skipped the citizenship question in the questionnaire. Due to the politicization of immigration and border security issues, foreigners residing in the United States may have become more hesitant to answer the citizenship question over time. The objective of this study is to observe the family size, employment industry, educational attainment level, and total income of non-citizen nonresponse participants from five U.S states with the largest immigrant population.

While California and New York had the highest proportion of nonresponse rates among noncitizens, a logistic regression across five states indicated that Arizona has the highest odds ratio of having a noncitizen-nonresponse. This implies the responding patterns differ across states. However, we only considered five states as a group in this analysis. Our future work would further explore how geographic factors and local policies influence people's response patterns on the citizenship question.

Before the analysis, we made four hypotheses about the "noncitizen-nonresponse" group:

1. It will have the lowest educational attainment level among the four defined groups.
2. It will have the largest family size among the four defined groups.
3. It will have a higher likelihood of employment in the low-skilled service industry.
4. It will have the lowest total income average among the four defined groups.

A logistic regression on educational level indicated that the "noncitizen-response" group has the lowest educational attainment level. The group had a higher likelihood of not having a high school diploma while the "noncitizen-nonresponse" group had a higher likelihood of having a

high school diploma, some college, and a bachelor's degree. On average, U.S. citizens who did not answer the citizenship question had a lower educational attainment level than those who responded.

For family size, the non-citizens had a larger mean than the citizens. Within the non-citizens, the "noncitizen-nonresponse" had an average family size of 3.69 whereas the "noncitizen-response" had 3.78.

There was a statistically significant relationship between non-citizens and their employment in the "Agriculture, Forestry, and Hunting" industry, as the nonresponse people were 61.8% less likely to work in the industry compared to the people who responded to the citizenship question. "Noncitizen-nonresponse" people were 61.7% more likely to be in the "Educational Services and Healthcare and Social Assistance" than the non-citizens who responded to the citizenship question.

Lastly, a linear regression analysis on pre-tax personal income suggested that the "noncitizen-nonresponse" group had an average income that is \$5300 above that of the "noncitizen-response" group. American citizens who did not respond to the citizenship question had an average income of \$10,000 below that of those who did respond. Both of these findings are statistically significant at a 99% confidence interval.

Our results indicated that our hypotheses on demographic characteristics of noncitizens who skipped the citizenship question were more pronounced in the "noncitizen-response" group. "Noncitizen-nonresponse" group's higher educational attainment and total income levels, compared to the "noncitizen-response" group, may be attributed to the fact that noncitizens with higher income and education face a higher opportunity cost of identification or deportation. In

addition, our results indicated a socioeconomic disparity between the “citizen-response” and “citizen-nonresponse” group in terms of education and income. Survey literature suggests that people with higher education tend to have a higher response rate.¹⁹ Our result partially aligns with this, as for the citizen group, the response group has a higher educational attainment level but for noncitizen group, the non-response group is found to have a higher educational attainment level.

Assuming respondents’ the socioeconomic vulnerability influence their behavior of answering the citizenship question, we hypothesised that noncitizen-nonresponse will have the lowest average income, lowest educational attainment level, most likely to be employed in a low-skilled service industry, and largest family size. However, our results suggest that the “noncitizen-response” group displays attributes of a relatively lower socioeconomic status than the “noncitizen-nonresponse” group.

¹⁹ Tolonen, Hanna, et al. “25-Year Trends and Socio-Demographic Differences in Response Rates: Finnish Adult Health Behaviour Survey.” *European Journal of Epidemiology*, vol. 21, no. 6, 2006, pp. 409–415., doi:10.1007/s10654-006-9019-8.

Appendix

Table 1a: Demographic variable by citizenship and response type from 2013-2017 with 5 selected states

	Non-citizen: Response				Non-citizen: Nonresponse				Citizen: Response				Citizen: Nonresponse			
	Mean	SE	Obs	Count	Mean	SE	Obs	Count	Mean	SE	Obs	Count	Mean	SE	Obs	Count
Family Size																
2013	3.84	0.0156	83,380	10,358,681	3.60	0.0383	15,231	1,477,449	2.83	0.0046	591,993	57,160,773	3.22	0.0218	30,744	2,888,571
2014	3.79	0.0142	82,525	10,367,255	3.65	0.0355	16,487	1,642,466	2.83	0.0046	591,253	57,518,402	3.23	0.0225	33,643	3,216,288
2015	3.78	0.0150	82,488	10,450,812	3.62	0.0382	16,784	1,688,955	2.83	0.0047	592,038	57,590,474	3.23	0.0201	37,767	3,630,871
2016	3.76	0.0155	80,467	10,273,116	3.75	0.0341	18,579	1,868,553	2.83	0.0047	598,096	58,006,087	3.29	0.0231	37,281	3,655,047
2017	3.71	0.0153	79,224	10,059,345	3.75	0.0350	19,408	2,034,181	2.84	0.0048	603,113	58,615,756	3.29	0.0219	37,639	3,799,534
Income (USD)																
2013	22,859	174.68	83,380	10,358,681	27,700	446.79	15,231	1,477,449	39,017	99.40	591,993	57,160,773	29,440	357.49	30,744	2,888,571
2014	23,407	169.60	82,525	10,367,255	29,029	459.78	16,487	1,642,466	39,950	98.68	591,253	57,518,402	30,021	340.73	33,643	3,216,288
2015	24,932	182.01	82,488	10,450,812	29,888	459.78	16,784	1,688,955	41,591	102.49	592,038	57,590,474	31,185	322.89	37,767	3,630,871
2016	26,070	199.74	80,467	10,273,116	31,650	453.06	18,579	1,868,553	42,666	102.97	598,096	58,006,087	32,011	330.32	37,281	3,655,047
2017	28,101	206.39	79,224	10,059,345	32,820	515.71	19,408	2,034,181	44,082	106.08	603,113	58,615,756	33,560	406.41	37,639	3,799,534
Education																
Did not Graduate HS																
2013	45.32%	0.2590	36,981	4,694,370	32.87%	0.5091	5,141	485,614	13.21%	0.0628	78,018	7,551,023	21.20%	0.2965	6,788	612,126
2014	44.46%	0.2553	35,853	4,609,174	31.04%	0.4694	5,259	509,788	12.82%	0.0617	75,703	7,375,086	19.95%	0.2752	7,004	641,126
2015	43.60%	0.2579	35,021	4,556,577	31.54%	0.4470	5,474	532,663	12.61%	0.0612	73,829	7,260,772	20.21%	0.2619	7,924	733,648
2016	42.68%	0.2577	33,515	4,384,988	31.18%	0.4203	5,994	582,616	12.38%	0.0611	72,604	7,180,572	19.81%	0.2585	7,731	724,088
2017	41.31%	0.2590	31,360	4,155,175	29.73%	0.4325	6,033	604,812	12.00%	0.0603	71,026	7,035,135	19.73%	0.2625	7,689	749,672
Graduated HS																
2013	22.47%	0.2053	17,590	2,327,909	24.19%	0.4359	3,605	357,360	24.83%	0.0805	147,180	14,191,685	27.53%	0.3192	8,596	795,357
2014	22.43%	0.2028	17,427	2,325,778	24.04%	0.4184	3,908	394,870	24.64%	0.0801	146,364	14,170,688	26.47%	0.3017	9,047	851,306
2015	22.25%	0.2037	17,426	2,325,744	23.99%	0.4124	3,987	405,240	24.84%	0.0816	146,543	14,304,788	27.53%	0.2876	10,569	999,551
2016	22.96%	0.2100	17,092	2,358,979	24.78%	0.3933	4,611	463,114	24.35%	0.0805	145,299	14,124,738	26.40%	0.2871	10,029	964,814
2017	22.80%	0.2091	16,929	2,293,120	26.51%	0.4186	5,015	539,363	24.49%	0.0805	146,070	14,352,629	26.78%	0.3001	10,709	1,017,542
Some College																
2013	14.76%	0.1662	12,542	1,528,919	22.26%	0.4349	3,374	328,919	33.93%	0.0866	196,109	19,394,560	30.62%	0.3289	9,355	884,420
2014	14.97%	0.1662	12,572	1,552,367	22.85%	0.4111	3,774	375,243	33.91%	0.0863	195,041	19,502,666	31.72%	0.3221	10,505	1,020,124
2015	14.70%	0.1664	12,486	1,536,516	22.51%	0.4120	3,764	380,221	33.56%	0.0865	194,304	19,324,534	30.80%	0.2970	11,606	1,121,652
2016	14.55%	0.1653	11,989	1,495,002	21.71%	0.3838	3,999	405,578	33.50%	0.0854	195,784	19,434,923	31.49%	0.3092	11,529	1,150,918
2017	14.79%	0.1736	11,943	1,487,491	21.40%	0.3789	4,135	435,358	33.24%	0.0851	196,136	19,483,365	31.02%	0.3112	11,566	1,178,760
Bachelor's Degree																
2013	10.31%	0.1465	9,252	1,067,647	13.72%	0.3627	2,047	202,715	17.98%	0.0705	106,520	10,274,787	14.28%	0.2589	4,173	412,577
2014	10.73%	0.1475	9,532	1,112,522	14.53%	0.3608	2,323	238,723	18.28%	0.0705	108,270	10,516,137	14.77%	0.2472	4,787	475,121
2015	11.52%	0.1543	10,050	1,536,516	14.48%	0.3471	2,374	244,529	18.59%	0.0707	110,506	10,706,853	14.91%	0.2385	5,287	541,228
2016	11.52%	0.1557	10,026	1,183,449	14.92%	0.3384	2,656	278,788	19.04%	0.0707	114,982	11,042,178	15.73%	0.2496	5,588	574,983
2017	12.20%	0.1565	10,776	1,227,370	14.91%	0.3394	2,792	303,213	19.33%	0.0706	118,096	11,330,195	15.60%	0.2598	5,647	592,836
Graduate Degree																
2013	7.14%	0.1196	7,015	739,836	6.96%	0.2595	1,064	102,841	10.06%	0.0516	64,166	5,748,718	6.36%	0.1802	1,862	183,729
2014	7.40%	0.1216	7,141	767,614	7.54%	0.2615	1,223	123,842	10.35%	0.0523	65,875	5,953,825	7.09%	0.1820	2,300	228,009
2015	7.93%	0.1278	7,505	828,351	7.48%	0.2792	1,185	126,302	10.41%	0.0518	66,856	5,993,577	6.07%	0.1625	2,381	234,795
2016	8.28%	0.1309	7,845	850,698	7.41%	0.2561	1,319	138,457	10.73%	0.0529	69,427	6,221,176	6.57%	0.1668	2,404	240,244
2017	8.91%	0.1349	8,416	896,189	7.45%	0.2443	1,433	151,435	10.94%	0.0527	71,785	6,414,432	6.86%	0.1709	2,528	260,724
								</								

A. State-level nonresponse analysis**Table 2a: State-level Data Tabulation>**

Number of observations= 3,648,170

Population size= 366,302,616

	Noncitizen -response	Noncitizen -nonresponse	Citizen -response	Citizen -nonresponse	Total
Arizona	9.322	1.809	82.23	6.636	100
California	18.02	3.202	73.82	4.953	100
Illinois	8.664	1.297	86.02	4.104	100
New York	12.68	3.068	79.42	4.828	100
Texas	13.58	1.454	80.89	4.073	100
Total	14.06	2.378	78.87	4.693	100

<Table 3a: Nonresponse by State: Logistic Regression>

Number of observations= 3,648,170

Population size= 366,302,616

	Odds Ratio	Standard Error	t-statistic	P > t
Arizona	1.576798	.0261678	27.44	0.000
California	1.517908	.0152295	41.60	0.000
Illinois	.9588217	.0132043	-3.05	0.002
New York	1.465499	.0168775	33.19	0.000
Texas	.0584996	.0004745	-349.95	0.000

B. Education**<Table 4a: Tabulation of education: citizen-response and citizen-nonresponse>**

Number of observations= 3,153,597

Population size= 306,081,803

Educational Attainment	Citizen-response Proportion (SE)	Citizen-nonresponse Proportion (SE)	Total
Did not graduate high school	13.21 (.0628)	21.2 (.2965)	18.98
Highschool Diploma	24.83 (.0805)	27.53 (.3192)	24.07

Some College	33.93 (.0866)	30.62 (.3289)	30.03
Bachelor's Degree	17.98 (.0705)	14.28 (.2589)	17.11
Graduate Degree	10.06 (.0516)	6.361 (.1802)	9.816
	100	100	100

<Table 5a: Tabulation of education: noncitizen-response and noncitizen-nonresponse>

Educational Attainment	Noncitizen-response Proportion (SE)	Noncitizen-nonresponse Proportion (SE)	Total
Did not graduate high school	41.31 (.259)	29.73 (.4325)	17.37
Highschool Diploma	22.8 (.2091)	26.51 (.4186)	23.89
Some College	14.79 (.1736)	21.4 (.3789)	29.53
Bachelor's Degree	12.2 (.1565)	14.91 (.3394)	18.43
Graduate Degree	8.909 (.1349)	7.445 (.2443)	10.78
	100	100	100

<Table 6a: Education of "noncitizen-nonresponse": Logistic Regression>

Number of observations= 3,648,170

Population size= 366,302,616

Group	Odds-Ratio	Standard Error	t-statistic	P > t
Did not graduate high school	1.49961**	.0095974	63.31	0.000
Highschool Diploma	1.105411**	.0065451	16.93	0.000
Some College	.8797891**	.0050865	-22.15	0.000
Bachelor's Degree	.8251652**	.0061927	-25.61	0.000
Graduate Degree	.66021**	.0068193	-40.20	0.000

** for P < 0.01

C. Family Size

<Table 7a: Family size>

Number of observations= 3,648,170

Population size= 366,302,616

Group	Mean	Standard Error	t-statistic	P > t
Citizen-response	2.832448**	.002089	1355.88	0.000
Citizen-nonresponse	3.255914**	.0098353	331.04	0.000
Noncitizen-response	3.777885**	.0067551	559.26	0.000
Noncitizen-nonresponse	3.678077**	.0159424	230.71	0.000

** for P < 0.01

<Two-tailed t-test: Difference in Family Size between “noncitizen-response” and “noncitizen-nonresponse”>

Coefficient	Standard error	t-statistic	P > t
-.0998075**	.0169919	-5.87	0.000

** for P < 0.01

<Two-tailed t-test: Difference in Family Size between “citizen-nonresponse” and “citizen-nonresponse”>

Coefficient	Standard error	t-statistic	P > t
.4221629**	.0172924	24.41	0.000

** for P < 0.01

D. Employment Industry

<Table 8a: Tabulation of Common Industries: citizen-response and citizen-nonresponse>

Number of observations= 4,143,163

Population size= 418,013,219

Industry Category	Citizen-response Proportion (standard error)	Citizen-nonresponse Proportion (standard error)	Total
Educational Services, Health Care, and Social Assistance	17.06 (.0285)	14.83 (.106)	16.08 (.0237)
Agriculture, Forestry, Fishing and Hunting, and Mining	11.59 (.0245)	11.72(.096)	13.01 (.0225)
Retail Trade	8.728 (.0229)	9.798 (.0907)	8.269 (.0188)

Table 9a: Tabulation of Common Industries: “noncitizen-response” and “noncitizen-nonresponse”

Number of observations= 4,143,163

Population size= 418,013,219

Industry Category	Noncitizen-response Proportion (standard error)	Noncitizen-nonresponse Proportion (standard error)	Total
Agriculture, Forestry, Fishing and Hunting, and Mining	21.07 (.0866)	14.16 (.151)	13.01 (.0225)
Professional, Scientific, and Management, and Administrative, and Waste Management Services	10.09 (.0623)	8.451 (.1231)	8.726 (.0186)
Arts, Entertainment, and Recreation, and Accommodation and Food Services	10.49 (.0707)	8.496 (.1243)	7.595 (.0192)
Educational Services, and Health Care and Social Assistance	9.164 (.0566)	14.03 (.1464)	16.08 (.0237)

Table 10a: Common Industries: Logistic Regression: “noncitizen-nonresponse” and “noncitizen-response”

Number of observations = 494,573

Population size = 60,220,813

Group	Odds-Ratio	Standard Error	t-statistic	P > t
Agriculture, Forestry, Fishing and Hunting, and Mining	0.61784**	0.0082982	-35.85	0.000
Educational Services, and Health Care and Social Assistance	1.6176**	0.0224482	34.66	0.000
Arts, Entertainment, and and Food Services	0.7924**	0.0139389	-13.22	0.000
Professional, Scientific, and Management, and Administrative, and Waste Management Services	0.82267**	0.0142255	-11.29	0.000

** for P < 0.01

E. Total Income

Table 11a: Pre-tax Total Income: Linear Regression

Number of observations= 3,648,170

Population size= 366,302,616

Group	Mean	Standard Error	t-statistic	P > t
Citizen-response	41476.41**	44.47546	932.57	0.000
Citizen-nonresponse	31374.71**	159.1675	197.12	0.000
Noncitizen-response	25054.05**	82.99071	301.89	0.000
Noncitizen-nonresponse	30417.42**	212.9094	142.87	0.000

** for P < 0.01

<Two-tailed t-test: Difference in Pre-tax Total income between “noncitizen-response” and “noncitizen-nonresponse”>

Coefficient	Standard error	t-statistic	P > t
5363.37**	228.503	23.47	0.000

** for P < 0.01

<Two-tailed t-test: Difference in Pre-tax Total income between “citizen-response” and “citizen-nonresponse”>

Coefficient	Standard error	t-statistic	P > t
10101.7**	165.265	61.124	0.000

** for P < 0.01

Reference

1. Jordan, Miriam. "If Census Asks About Citizenship, Some Already Have an Answer: No Comment." *The New York Times*, The New York Times, 27 Mar. 2018, www.nytimes.com/2018/03/27/us/census-undocumented-immigrants.html.
2. "Read 'Modernizing the U.S. Census' at NAP.edu." National Academies Press: OpenBook, 1995, www.nap.edu/read/4805/chapter/16.
3. Ahlmark, Nanna, and Maria Algren. "Survey Nonresponse among Ethnic Minorities in a National Health Survey – a Mixed-Method Study of Participation, Barriers, and Potentials." *Taylor and Francis Online*, Oct. 2014.
4. Wang, Hansi Lo. "Trump Officials Prepared For Supreme Court Fight Over Census Question." NPR, NPR, 12, Sept., 2018
5. O'hare, W. (2018). *Citizenship Question Nonresponse: A Demographic Profile of People Who Do Not Answer the American Community Survey Citizenship Question*. <http://www.georgetownpoverty.org/wp-content/uploads/2018/09/GCPI-ESOI-Demographic-Profile-of-People-Who-Do-Not-Respond-to-the-Citizenship-Question-20180906.pdf>
6. Ibid.
7. Krogstad, Jens Manuel, and Michael Keegan. "15 States with the Highest Share of Immigrants in Their Population." *Pew Research Center*, Pew Research Center, 14 May 2014, www.pewresearch.org/fact-tank/2014/05/14/15-states-with-the-highest-share-of-immigrants-in-their-population/.
8. Census Bureau. "Group Quarters/Residence Rules." <https://www.census.gov/topics/income-poverty/guidance/group-quarters.html>
9. Torrieri, N. "American Community Survey Design and Methodology" *Census Bureau*, June 2014.
10. Census Bureau. "American Community Survey - Coverage Rates." <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/coverage-rates/>
11. Torrieri, N. "American Community Survey Design and Methodology" *Census Bureau*, June 2014.
12. Torrieri, N. "American Community Survey Design and Methodology" *Census Bureau*, June 2014.
13. Ibid.
14. Census Bureau. "American Community Survey - Response Rates." <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/>
15. Ibid.
16. Torrieri, N. "American Community Survey Design and Methodology" *Census Bureau*, June 2014.
17. Ibid.
18. Napolitano, J. (June 15, 2012). *Exercising prosecutorial discretion with respect to individuals who came to the United States as children* [Memorandum]. Washington, D.C.: Dept of Homeland Security. <https://www.dhs.gov/xlibrary/assets/s1-exercising-prosecutorial-discretion-individuals-who-came-to-us-as-children.pdf>
19. Tolonen, Hanna, et al. "25-Year Trends and Socio-Demographic Differences in Response Rates: Finnish Adult Health Behaviour Survey." *European Journal of Epidemiology*, vol. 21, no. 6, 2006, pp. 409–415., doi:10.1007/s10654-006-9019-8.

