Assessing the "Year of Naturalization" Data in the American Community Survey: Characteristics of Naturalized Foreign Born Who Report – and Don't Report –the Year They Obtained Citizenship

Elizabeth M. Grieco and David M. Armstrong

Population Division Working Paper No. ##
U.S. Census Bureau
Washington, D.C. 20233

June 2015

Acknowledgements: The authors would like to thank Howard Hogan, Deborah Griffin, Nancy Bates, Joshua Comenetz, Victoria Velkoff, Jason Devine, Mike Starsinic, Amy Lauger, Tom Gryn, Laurie Schwede, Bashir Ahmed, Paul Beatty and Wendie Choudary for their comments and suggestions. The Estimates and Projections Area of the Population Division provided support for this research. An earlier version of this paper was presented at the 2014 Applied Demography Conference in San Antonio, TX.

Assessing the "Year of Naturalization" Data in the American Community Survey: Characteristics of Naturalized Foreign Born Who Report – and Don't Report – the Year They Obtained Citizenship

Elizabeth M. Grieco and David M. Armstrong

Population Division Working Paper
U.S. Census Bureau

KEY WORDS

Nonresponse; Survey mode; Social networks; Naturalization; Foreign Born

ABSTRACT

The American Community Survey (ACS) includes a write-in question asking for "year of naturalization" for all individuals reported as foreign born and naturalized citizens. However, not all of the foreign born who are naturalized provide the year they became citizens. This paper uses logistic regression analysis and data from the 2011 ACS to determine the characteristics associated with those naturalized citizens for whom year of naturalization is – and is not – reported. The research literature demonstrates that various demographic and social characteristics influence reporting behavior. However, this analysis focuses on two variables: 1) survey environment and 2) social proximity to the respondent. Naturalized citizens in households that provide information about their members using a mail-back questionnaire were more likely to have a reported year of naturalization than those in households interviewed by phone or in-person. In addition, they were more likely to have a reported year of naturalization if they were householders (who usually act as primary respondent) or were closely related to householders. The results of the analysis suggest that item nonresponse is likely to be higher for individuals in complex households with weak ties to the householder and little or no direct contact with the survey instrument or interviewer.

1. Introduction

The American Community Survey (ACS) includes a question asking: "Is this person a citizen of the United States?" The Census Bureau considers all individuals who report they were born in the United States, Puerto Rico, a U.S. island area (i.e., Guam, the U.S. Virgin Islands, Northern Marianas, or American Samoa), or abroad of a U.S. citizen parent or parents as native born and all others born outside the United States as foreign born. The foreign born include naturalized U.S. citizens and noncitizens. Persons who are foreign born and are naturalized citizens also are directed to: "Print year of naturalization."

Conducted by the U.S. Census Bureau, the ACS is a nationwide survey designed to provide communities with reliable and timely demographic, social, economic, and housing data every year. According to the 2011 ACS, there were 311.6 million residents in the United States, including 40.4 million foreign born: 18.1 million naturalized citizens and 22.2 million noncitizens. Naturalized citizens comprised about 45 percent of the total foreign-born population.

Not all foreign born who are naturalized provide the year they became citizens. Of the 18.1 million naturalized foreign born estimated by the ACS, 17 percent did not have a year of naturalization reported.[1] As obtaining citizenship is a significant life event for many immigrants – and the date of occurrence likely remembered and recalled – the rate of nonresponse seems

[1] The Census Bureau uses imputation methods to replace missing values with plausible estimates through assignment and allocation (U.S. Census Bureau 2009). The imputation rate is calculated by adding the number of missing responses imputed through allocation and then dividing it by the total number of responses. In 2011, there were 3.1 million naturalized foreign born whose year of naturalization response was either assigned or allocated, representing 17 percent of the total 18.1 million naturalized foreign born.

unusually high, a presumption supported when compared with the nonresponse rates of other items on the ACS. For example, the overall person characteristic imputation rate was 5.8 percent. The rates for other questions that similarly asked respondents to write in or provide a year were also lower: 9.9 percent for the year the respondent came to live in the United States; 7.7 percent for the last year the respondent married; and 3.2 percent for the month and year the householder moved into the housing unit. Additionally, the nonresponse rate for citizenship was lower than for year of naturalization: 2.7 percent for the total population, 2.4 percent for native born and 4.5 percent for foreign born.

The U.S. Census Bureau first added year of naturalization to the 2008 ACS questionnaire, motivated primarily by opportunities to improve editing of U.S. citizenship status (Harris et al. 2007). Researchers have also used the data to assess its quality by comparing it to administrative data on naturalizations collected by the Department of Homeland Security U.S. Citizenship and Immigration Services (Baker 2012, Van Hook and Bachmeier 2013). However, high imputation rates have remained an issue since the year of naturalization data were first released, ranging between 14.7 percent and 18.2 percent for 2008 to 2011 (Table 1). This can negatively affect the accuracy of research results, as high imputation rates increase the risk of biased estimates, especially if nonrespondents are not similar to respondents but are imputed as such (Groves et al. 2004).

To investigate possible reasons for the high year of naturalization nonresponse rate, this paper uses logistic regression analysis and data from the 2011 ACS to determine the characteristics associated with those naturalized foreign born who have – and do not have – a reported year they

became citizens. We develop a statistical model that incorporates both respondent- and household-level characteristics but focuses on two primary explanatory variables: 1) survey environment and 2) social proximity to the primary respondent. While many of the individual and household characteristics included do significantly influence reporting behavior, two variables – survey mode and relationship to householder – exhibited greater explanatory power. Naturalized citizens in households that provide their information using a mail-back questionnaire were more likely to have a reported year they obtained citizenship than those interviewed by phone or in person. In addition, they were more likely to have their year of naturalization reported if they were householders – i.e., the people who most commonly act as respondents for households – or closely related to householders. The results of the analysis suggest that, in general, item nonresponse is likely to be higher for individuals in complex households with weak ties to the householder and little or no direct contact with the survey instrument or interviewer.

We begin by discussing the individual, household, and contextual variables that, according to the literature, have significant influence on survey response behavior, incorporating many of the same variables into our model. Next, we develop a conceptual framework that outlines how the survey process works within households to either enable/encourage or prevent/discourage information from being passed by a household member either 1) directly to the survey taker or 2) indirectly to the survey taker through a primary respondent. We then describe the American Community Survey data, the statistical methods used, and the dependent, primary, and control variables included in the model. Unlike other studies, which usually exclude cases with missing data, we retain them and control for nonresponse through the use of dummy variables. Finally,

we review the results of the logistic regression models. We conclude by discussing how the conceptual model presented in this paper can be used to explain differential item nonresponse.

2. Literature Review

Survey methodologists divide nonresponse behavior into two types: 1) unit nonresponse and 2) item nonresponse. Unit nonresponse occurs when the sample unit does not respond to the request to be surveyed, while item nonresponse is the failure to obtain answers to individual survey questions when unit response is obtained (Dillman et al. 2002).

Research has demonstrated a number of individual characteristics that are associated with higher rates of item nonresponse. For example, age is positively associated with item nonresponse (Craig and McCann 1978, Dillman et al. 2002, Francis and Robbins 1995, Kaldenberg, Koenig, and Becker 1994, Messer, Edwards, and Dillman 2012, Messmer and Seymour 1982). Women are more likely than men to exhibit item nonresponse (Francis and Busch 1975, Goyder 1982, McDaniel, Madden, and Verille 1987). However, most of the item nonresponse studies have focused on the effect of education and socioeconomic status on nonvalid responses (Shoemaker, Eichholz, and Skewes 2002). For example, research has found educational attainment to be a significant predictor of item nonresponse; that is, people with less education are less likely to provide answers in many surveys (Craig and McCann 1978, Dillman et al. 2002, Dodge et al. 1993, Francis and Busch 1975, Messer, Edwards, and Dillman 2012). In addition, socioeconomic status is negatively associated with item nonresponse (Bell 1984, Bishop, Tuchfarber, and Oldendick 1986, Schuman and Presser 1996).

While Goyder (1987) stressed that surveys are inherently social activities, fewer studies have analyzed the influence of the social context – as operationalized by household or neighborhood characteristics – on individual response rates. Groves and Couper (1998) analyzed a range of social indicators, including population density, crime rates, and household composition, that might logically be thought to correlate with nonresponse rates in surveys. Johnson et al. (2002) discussed how culture could explain intergroup differences in response rates, for example, across certain subgroups or even within a population. However, they concluded that correlations between culture and nonresponse rates are hard to measure because there are few good quantitative indicators of cultural climate.

The results of a study on coverage error by Fein and West (1988) underscore the need to consider household context when considering reasons for nonresponse behavior. The authors use data from the 1986 Los Angeles Test Census, fielded by the U.S. Census Bureau as part of the preparations for the 1990 Census, to study causes of undercount in a hard-to-count, largely Hispanic community. They found that partial household omissions accounted for most of the census undercount in hard-to-enumerate urban areas as compared to omissions of entire households. However, they argue that undercount is a problem that occurs primarily at the household rather than individual level, especially given a census is based on contacting households rather than every individual in a household.

Even if a household agrees to participate in a census or survey, there are two reasons why information on all its members may not be provided to an enumerator, according to Fein and

West (1988). First, there may be perceived costs to reporting all members of a household: for example, the presence of adult males may jeopardize welfare eligibility; persons illegally in the country may be deported; and reporting more people than allowed on a lease may result in landlord troubles, etc. Second, respondents may not understand who is supposed to be included on the questionnaire. Definitional errors (i.e., omissions from the household roster) arise when respondents fail to understand and apply census enumeration and residence rules, especially in large and complex households. For example, definitions of kinship may influence who is or is not perceived as a household member, which could increase the likelihood of omission of distant kin and nonrelatives. Size may work to lower the social visibility of individual members, resulting in an increased risk that they will be forgotten by respondents.

Fein and West (1988) found that respondents in households that were large, that contained more distant relatives and persons unrelated to the household head, whose primary language was not English, and where the census respondent's education level was less than high school exhibited greater risks of making definitional errors. Within-household omission also was found to be correlated with the presence of immigrants, welfare recipiency, and crowding. The authors suggest the results of their study support the hypothesis that some household members with these characteristics may be concealed by census respondents.

More recent research by Schwede and Terry (2013) also found that complex households exhibited greater possible coverage error than those considered noncomplex. Using comparative ethnographic methodology, the goals of the analysis were to explore types and sources of possible census coverage error and identify the characteristics of households and of persons

affected by them through observation of 2010 Census interviews. Households were classified into a complex/noncomplex typology by defining 1) nuclear families consisting of married parents and their joint biological children, single parents and their biological children, and persons living alone as noncomplex and 2) all others as complex. Schwede and Terry (2013) found that complex rather than noncomplex households exhibited more instances of possible coverage error, results supported by past qualitative research (de la Puente 1993, Schwede 2003, Schwede, Blumberg, and Chan 2006). In addition, co-residents more socially distant from the reference person – including aunts/uncles, roomers/boarders, and housemates/roommates – were the relationship types that had the greatest proportions of possible coverage error. In their dichotomy, complex households would include these distant relatives or nonrelatives.

In addition to individual and household characteristics, survey methodologists also recognize the influence of mode on response rates. When compared to telephone or face-to-face interviewing, item nonresponse is a larger problem in mail surveys (de Leeuw 1992, de Leeuw, Hox, and Huisman 2003). According to Tourangeau, Rips, and Rasinski (2000), item nonresponse is more likely to occur in self-administered questionnaires than those conducted by interviewers because self-administered formats allow the respondent to determine which questions are read, in what order, and whether or not an answer is recorded.

3. Conceptual Framework

Most large surveys conducted by the U.S. Census Bureau are household surveys.[2] The housing unit – not the individual – is selected into the sample, although information about each household member is collected. In most cases, one person – the primary respondent – provides the requested information for all members of the household. While the primary respondent provides his or her own information, the remaining household members have their information provided by proxy through the primary respondent. Research analyzing 2004 American Community Survey mail-mode data demonstrated that the primary respondent is usually the person listed as the "householder" (Hill et al. 2008), that is, the person (or one of the people) who owns or rents the housing unit.[3] According to the analysis, approximately 81 percent of people listed as the householder were respondents. In other words, for most households, the householder – who should have proxy knowledge about each member of the household – acts as the primary respondent, supplying the requested information to the Census Bureau.

However, the influence a primary respondent has on how and what information moves through the household – from each household member to the householder and then to the Census Bureau – will depend on two additional factors. The first is survey environment. The American Community Survey is a multi-mode survey and household members respond in one of three ways: 1) by mail, using a paper questionnaire; 2) by phone, through a computer assisted

[2] Note that the American Community Survey is a survey of the resident population and includes both households and group quarters (e.g., dormitories, nursing homes, and prisons). The results from the 2011 American Community Survey discussed in the introduction of this paper represent the resident population. However, the logistic regression analysis presented in this study is limited to the household population and excludes group quarters.

[3] Note that respondents are directed to designate the person living or staying in the household in whose name the house or apartment is owned, being bought, or rented (i.e., the householder) as "person 1." If there is no such person, any adult living or staying in the household can be designated person 1.

telephone interview (CATI); and 3) face-to-face, through a computer assisted personal interview (CAPI).[4] In all three modes, information about each household member may be passed to the primary respondent in one of two ways: directly from household member to primary respondent; and indirectly from one household member through another and then to the primary respondent (Figure 1).[5] For telephone and in-person interviews, a primary respondent would then provide the information directly to a census interviewer.

This information path is altered notably for households that respond using a mail-back questionnaire. While it is possible (and even likely) that a single primary respondent completes the form, it is also possible that others in the household would access the questionnaire directly and provide their own information. Shortening the path between respondent and survey instrument would likely improve response rates and the accuracy of the information provided. Not only do mail-back questionnaires increase the likelihood that multiple household members have access to the survey instrument, they also provide more time for them to do so. Information collected through CATI and CAPI usually is completed during a delimited period, such as a single phone call or interview, and with a single respondent.[6] By comparison, a household may hold on to a questionnaire for several days, improving the likelihood that multiple respondents would participate in its completion.

[4] The U.S. Census Bureau added an additional mode – data collection through the Internet – as an option to respond to the American Community Survey in 2013.

[5] Here, we are talking about characteristics of an individual that a primary respondent would not know through obvious visual cues. It is possible that the primary respondent would supply some information, such as sex or race, without consulting the individual or other household members. Information about other characteristics, such as year of naturalization, would have to be obtained by the primary respondent directly from the individual or through interaction with other household members.

[6] Sometimes during a CATI/CAPI interview, multiple people respond. Unfortunately, the data are not available to determine how often this happens.

A second factor that influences how information moves from household members to the primary respondent is social distance. In this report, social distance refers to how closely related a household member is to the householder. Closely related individuals, such as spouses or parents and children, continually share information on a wide variety of topics that is both current and accurate. This would improve the householder's ability to respond to the survey (Figure 2). However, as the social distance between the primary respondent and a household member increases – for example, to an extended family member, a distant relative, or a nonrelative – the flow of information would likely decrease in variety, frequency, and accuracy, decreasing the householder's ability to respond from memory and thereby increasing nonresponse rates.

The flow of information to the primary respondent is augmented by the strength of ties among the remaining household members. For example, a wife may provide information about her live-in parents to her husband – the householder – which is then reported accurately. In this sense, the composition and strength of the social ties within the household can help or hinder the flow of information. This would be especially important for telephone or face-to-face interviews, as information flows to the interviewer or enumerator most often through a single respondent and during a single interview session. For households using mail-back questionnaires, however, the number of options through which information can be transmitted increases notably. Information about household members can flow: 1) through the primary respondent, possibly augmented by information exchanged by the remaining household members, then to the questionnaire; or 2) directly from household members to the questionnaire. Greater access to the questionnaire by all household members, in addition to more time with the instrument, could result in lower item nonresponse and more accurate data.

According to this model, households responding through CATI and CAPI should demonstrate higher rates of item nonresponse than households responding by mail. Additionally, householders should exhibit the lowest rates of item nonresponse, since they are likely to be primary respondents, knowing and providing their own information. In contrast, nonrelatives should have the highest rate of item nonresponse, with comparatively lower – and declining – rates for the householder's other relatives, extended family, and immediate family, in that order.

The hypothesized association between survey mode and item nonresponse rates can be seen in year of naturalization reporting. According to the 2011 American Community Survey, 17 percent of all naturalized citizens did not have a reported year of naturalization. When divided by survey mode, 13 percent of mail, 18 percent of CATI, and 22 percent of CAPI respondents did not have a reported year of naturalization (Figure 3). The association also can be seen when the population is divided by broad relationship categories. Among householders who reported they were naturalized citizens, 13 percent did not report a year of naturalization (see Figure 4). Compared with their relatives, 16 percent of immediate family members, 27 percent of extended family members, and 40 percent of other relatives did not have a reported year of naturalization. Nonrelatives, at 39 percent, were not significantly different from other relatives.

While survey environment and social distance from the primary respondent may influence item nonresponse, the survey methodology literature has clearly demonstrated that other characteristics, such as age, sex, and socioeconomic status, also play a role. The potential influence of these variables is especially important when considering mode since the

characteristics of households responding by different modes are known to vary. In the 2011 American Community Survey, sampled housing units received a request to complete a mail-back questionnaire.[7] Those that did not respond by mail were eligible to be contacted by phone (if a phone number was available) and asked to complete a CATI interview. A sample of the remaining housing units that did not or could not respond by mail or phone were interviewed in person (U.S. Census Bureau 2009). Differences in the item nonresponse rate by mode seen among naturalized citizens who have a reported year of naturalization may not be due to mode *per se*, but may simply reflect the behavior of those households that ultimately places them into either mail or CATI. In other words, households less willing or able to cooperate may have specific characteristics that set them apart from more willing households.

Note that our conceptual framework alters how item nonresponse is defined and operationalized. In general, when item nonresponse is discussed in the survey methodology literature, it refers to the behavior of an individual respondent who is assumed to be directly interviewed. However, the American Community Survey is a household survey, with information about all household members passed to the Census Bureau most commonly by a primary respondent. In this sense, item nonresponse does not simply reflect individual behavior or characteristics. Rather, it stems from the interaction among all members of a household, their collective characteristics, and their willingness and ability to pass accurate and timely information among themselves and, ultimately, to the Census Bureau. Unfortunately, we do not know definitively who reported information for a household member. Thus, in this paper, we forego more conventional wording

[7] There are some housing units that, while in sample, are not mailed a questionnaire but are instead interviewed in person only. For example, interviews in remote Alaska are conducted using personal visit procedures only, as are areas with a large number of nonmailable addresses (U.S. Census Bureau 2009).

(i.e., naturalized citizens were more likely *to report* a year of naturalization) for more precise terminology (i.e., naturalized citizens were more likely *to have a reported* year of naturalization) that better reflects the reporting behavior of households.

4. Data and Methods

This paper uses logistic regression analysis and data from the 2011 American Community Survey to determine the characteristics associated with those naturalized foreign born who have – and do not have – a reported year of naturalization. Based on our conceptual framework, we posit two primary explanatory variables: 1) survey environment and 2) social distance from the primary respondent. We also control for several individual and household characteristics previously reported in the literature to be associated with item nonresponse behavior.

4.1. Data

This study uses 2011 American Community Survey (ACS) data derived from internal Census Bureau files.[8] The ACS is a nationally representative survey that uses a series of monthly samples to produce annually updated data. In 2011, more than 3 million housing unit addresses were selected to be in sample.[9] Data for the 2011 ACS were collected continuously throughout the year using three successive modes: mail, computer-assisted telephone interviews (CATI), and

[8] Data were analyzed using SAS proc surveylogistic in order to use the replicate weights to estimate variances. For additional information on replicate weights, see Chapter 12 in U.S. Census Bureau (2009).
[9] The American Community Survey sample was expanded in mid-2011. The 2011 sample was 3.3 million addresses, while the 2012 sample was about 3.5 million.

computer-assisted personal interviews (CAPI).[10] Because of the sampling of non-respondents to mail and telephone modes of data collection in addition to overall nonresponse, the final number of interviewed housing units was approximately 2 million in 2011.[11]

The ACS provides data users and the public with a wealth of information on the characteristics of the foreign-born population in the United States. In addition to the other social, demographic, economic, household, and geographic data available, migration-related items include place of birth, U.S. citizenship status, year of naturalization, residence one year ago, and language spoken at home. For additional information on the design and methodology of the ACS, see U.S. Census Bureau (2009). For a discussion of the benefits and limitations of using the ACS to analyze immigration and the foreign-born population, see Grieco and Rytina (2011).

4.2. Level of Analysis

Because citizenship status is an individual-level characteristic, the likelihood that foreign-born citizens' year of naturalization was reported is modeled at the level of the individual. The model includes two primary explanatory variables – survey environment and social distance from the primary respondent. The model also includes additional individual and household variables as well as controls for nonresponse. Table 2 presents descriptive statistics of the explanatory

[10] The ACS collects data continuously throughout the year and includes 12 monthly independent samples. Data collection for each sample lasts for 3 months, with mail returns accepted during the entire period, CATI operations beginning about five weeks after the first mail package is sent out, and in-person interviews occurring in the third month. After mail and CATI operations have been completed, a CAPI subsample is selected from the remaining housing units in sample. Results are weighted so the published data represent the whole population. For more information, see U.S. Census Bureau (2009).

[11] Most of the differences between the initial and final sample sizes is the result of sampling for non-respondents. The weighted response rate for the ACS has consistently been 97 percent or greater across all modes of data collection.

variables considered in this analysis, first for the naturalized citizen population and then separately for those who had and did not have their year of naturalization reported. As Table 2 shows, the two groups differ, often considerably, in their characteristics.

However, while the level of analysis is at the individual level, the ACS is a household-level survey. This is important because, as discussed earlier, the information collected from a household about an individual member may be provided by that individual or – more likely – by a primary respondent, usually the householder. Unfortunately, it is unknown how many interviews were completed by a primary respondent alone and how many involved individual self-reporting by other household members.

4.3. Universe and Sample Size

Only those foreign born who were reported as naturalized citizens were included in the analysis; individuals with an imputed response of naturalized citizen were excluded. The sample also was restricted to naturalized foreign born who were aged 18 years and older. The analysis only includes residents of the United States; residents of Puerto Rico were excluded.

The American Community Survey collects data from the resident population living in both households and group quarters (such as prisons, dormitories, hospitals, and nursing homes). However, because the model used in this analysis focuses on the behavior of individuals within households, we excluded the population living in group quarters facilities.

After all universe restrictions were applied to the data, the resulting sample size is 225,350 naturalized citizens, including 200,869 with a reported year of naturalization and 24,481 without one. After weighting, this represents 16.4 million naturalized citizens, including 14.3 million with a reported year of naturalization and 2.1 million without one. This suggests a total item nonresponse rate of about 13 percent. This is lower than the imputation rate of 17 percent for all 18.1 million naturalized foreign born, but still considerably higher than the overall person characteristics imputation rate of 5.8 percent.

4.4. Dependent Variable

In this report, naturalized citizens are divided into one of two categories based on whether or not a year of naturalization was reported. This creates a dichotomous dependent variable comparing those with a reported a year of naturalization and those without it. For this analysis, we are modeling the probability of a year of naturalization being reported, either directly by a respondent or indirectly by a primary respondent.

4.5. Primary Explanatory Variables

4.5.1. Survey Environment

A single survey environment variable is included in the analysis to examine the influence of data collection mode (i.e., mail, CATI, and CAPI) on response patterns. Mode is a three-category dummy variable with *mail* as the reference (i.e., omitted) category. We hypothesize that

naturalized citizens in households using mail-back questionnaires will demonstrate lower nonresponse rates to the year of naturalization question because of a longer interview period and greater access to the survey instrument. Conversely, because of the restricted information paths between household members and the primary respondent, limited direct access of the household members to the interviewer or enumerator, and a shorter interview period, households that respond through CATI and CAPI will have a higher nonresponse rate.

The descriptive statistics presented in Table 2 support this hypothesis. About 8 percent of naturalized citizens living in households that responded by mail did not have a reported year of naturalization, compared with more than 18 percent of those in households responding by CAPI and more than 15 percent of households responding by CATI.[12]

4.5.2. Social Distance from the Primary Respondent

A single variable – relationship to householder – is included in the analysis to measure the social distance between household members and the primary respondent. For all households participating in the American Community Survey, one member is designated as person 1. In most cases, person 1 is the householder (defined as the person, or one of the people, in whose name the home is owned, being bought, or rented). The householder usually acts as the primary respondent for the household by completing the questionnaire or answering questions asked by telephone interviewers and field representatives (Hill et al. 2008). As the primary respondent, the householder provides his or her own information in addition to the information about other

[12] All comparative statements in this article have undergone statistical testing and comparisons are significant at the 90-percent confidence level unless noted otherwise.

household members. We hypothesize that the closer the household member is to the householder, the more likely a year of naturalization will be reported for that household member.

Relationship to householder is a five-category dummy variable with *householder* as the reference group. The remaining categories group household members together depending on social distance from the householder. Immediate family members include the householder's spouse and children, including adopted and step children. Extended family members include the householder's father and mother, brothers and sisters, and grandchildren, and parents-, sons- and daughters-in-law. Other relatives include more distant kin, such as cousins. Nonrelatives include roomers and boarders, housemates and roommates, unmarried partners, foster children, and other nonrelatives.

In our study universe we estimate that about 10 percent of householders did not have a reported year of naturalization, compared with 12 percent of immediate family members, 20 percent of extended family members, 29 percent of other relatives, and 27 percent of nonrelatives (Table 2).

4.6. Control Variables

4.6.1. Household Characteristics

The statistical models used in the analysis incorporate several variables representing household characteristics: region of residence, number of adults, homeownership, and poverty status. They

also include three variables reflecting the householder's ability to answer a survey: level of education, English speaking ability, and nativity.[13]

Region of residence is a four-category dummy variable with *West* as the reference category. This variable was included in the analysis to control for possible regional variations in households' willingness to complete surveys. Of those living in the West, 12 percent did not have a year of naturalization provided, compared with about 15 percent in the Northeast, 12 percent in the South, and 11 percent in the Midwest.

Number of adults is a continuous variable controlling for the number of household members aged 18 and older. In a study of coverage error, Fein and West (1988) found a positive association between household size and risk of omission of individual household members. We hypothesize that the likelihood of item nonresponse increases as the number of adults in the household increases. Additional adults may indicate socially complex households, such as those that include more distant kin or nonrelatives, which may increase the chance that primary respondents can't or forget to provide the information requested by a survey on other household members. While 11 percent of naturalized citizens living with either one or two other adults did not have a reported year of naturalization, 14 percent living with three other adults, 15 percent living with four other adults, and 16 percent living with five or more other adults did not have a reported year of naturalization.[14]

---

[13] The American Community Survey questionnaire and the CATI/CAPI instrument are available in two languages – English and Spanish. In addition, the U.S. Census Bureau has staff available to complete telephone or in-person interviews in other languages. However, the majority of all questionnaires and interviews are completed in English.

[14] The American Community Survey mail-back questionnaire is designed to accommodate detailed answers for households with five or fewer people and basic information, including name, sex, and age, for an additional seven people. Cases requiring additional detailed information, such as persons 6 through 12 in households with more than five people, are eligible for the telephone failed-edit follow-up (FEFU) operation. A higher nonresponse rate for the

Two variables were included in the analysis to account for socio-economic status of the household: homeownership and poverty status. Research has found that socio-economic status is negatively associated with item nonresponse (Bell 1984, Ferber 1966, Schuman and Presser 1996). Homeownership is a dummy variable with *rented* as the reference category. We hypothesize that the naturalized foreign born who live in households that are owned rather than rented are more likely to provide a year of naturalization. Almost 12 percent of naturalized citizens who lived in owned households did not have a reported year of naturalization compared with 15 percent in rented households. Household poverty status is a dummy variable with *in poverty* as the reference group. We hypothesize that the naturalized foreign born who are members of households that are not in poverty are more likely to report a year of naturalization. About 15 percent of those in poverty did not have a reported year of naturalization compared with 12 percent not in poverty.

Three additional variables were included in the model to control for the ability of the household to complete a survey questionnaire or answer an interviewer's questions. These include the householder's level of education, language spoken at home (i.e., either English or another language), and nativity status. The characteristics of the householder are used to assess this ability as the householder most commonly acts as the primary respondent for the household.

year of naturalization item could be a reflection of these missing cases. However, we do not believe this is the case for two reasons. First, the proportion of the total naturalized foreign born represented by persons 6 through 12 is small (about 2 percent). Second, by restricting our universe to those 18 years and older, we reduced this proportion further: the proportion of the naturalized foreign born aged 0 to 17 represented by persons 6 through 12 (10 percent) was higher than that for those aged 18 and over (1.7 percent).

Householder's level of education is a two-category dummy variable with *less than a high school degree* as the reference category. Previous research has found educational attainment to be negatively associated with item nonresponse (Craig and McCann 1978, Dillman et al. 2002, Dodge et al. 1993, Messer, Edwards, and Dillman 2012, Francis and Busch 1975). We hypothesize that householders who are more educated are more likely to understand and respond appropriately to survey questions, thus reducing item nonresponse rates for other household members. Fourteen percent of naturalized citizens living in households where the householder had less than a high school degree did not have a year of naturalization reported, compared with 12 percent of those in households where the householder had a high school degree or higher education.

Householder's language spoken at home is a two-category dummy variable with *speaks a language other than English at home* as the reference category. We hypothesize that householders who use English at home are more likely to understand and respond appropriately to the survey questions, thus lowering the likelihood of item nonresponse for other household members. However, the imputation rates suggest a different association: 12 percent of naturalized citizens living in households with householders who spoke a language other than English at home did not have a year of naturalization reported, compared with 15 percent of those living in households with householders who spoke only English at home.

Householder's nativity is a two-category dummy variable with *foreign born* as the reference category. We hypothesize that native householders, because of their increased exposure to and knowledge of the federal government and federal surveys – especially when compared with the

foreign born – are more likely to be willing to participate. As with language spoken at home, however, the imputation rates suggest a different association: 12 percent of naturalized citizens living in households with foreign-born householders did not have a year of naturalization reported compared with 17 percent in households with native-born householders.

4.6.2. Individual Characteristics of Naturalized Citizen Household Members

The statistical models used in this analysis incorporate several variables representing both demographic and social characteristics of foreign-born citizens: age, sex, place of birth, period of entry, migration in the last year, and relationship to the householder. These variables are included to control for the influence these characteristics may have on response behavior if the foreign-born citizen is 1) the householder (i.e., the likely primary respondent) or 2) a household member interacting with the householder. As with year of naturalization, these characteristics may have been reported directly by the naturalized citizen respondent or indirectly by a primary respondent.

Past research has shown that older respondents tend to have higher item nonresponse rates in many surveys (Dillman et al. 2002, Messer, Edwards, and Dillman 2012) and that questionnaires completed by females contain more nonresponse than those completed by males (Ferber 1966, Francis and Busch 1975, Francis and Robbins 1995). In this analysis, age is a four-category dummy variable with *60 and older* as the reference category. Sex is a binary dummy variable with *female* as the reference category. We expect that both younger and male naturalized citizens

– when they respond for themselves – will be more likely to report the year they naturalized than older and female naturalized respondents.

A place of birth variable is included in the model to control for possible differences in reporting year of naturalization among broad cultural groups. The literature suggests that differences in response patterns among race and Hispanic origin groups may exist which, in part, reflect broad cultural differences (Groves and Couper 1998, Johnson et al. 2002, Johnson et al. 1997, Rowland and Forthofer 1993). As the focus of this research is on the foreign born, we use place of birth rather than race or Hispanic origin to control for possible cultural differences. Place of birth is a five-category dummy variable with *Europe* as the reference group. Only 10 percent of those born in Europe did not have a year of naturalization reported, compared with 11 percent from Asia and other regions and 15 percent from Latin America.

The models used also incorporate two variables representing the migration process that may influence response rates: 1) period of entry into the United States and 2) moved last year (either domestically or internationally). Period of entry is a four-category dummy variable with *1996 and before* as the reference category. We hypothesize a negative association between time in the United States and reporting year of naturalization because of recall error. That is, as more time passes since citizenship was obtained, the more likely a respondent will forget and not report a year of naturalization. However, the imputation rates suggest a different pattern: year of naturalization was not reported for more than 45 percent of naturalized citizens who arrived between 2007 and 2011 compared with between 10 to 13 percent for the remaining groups.

While notable, the number of naturalized foreign born who were reported to have arrived between 2007 and 2011 is small, representing less than 1 percent of the total sample.[15]

Moved in the last year is a binary dummy variable with *yes* as the reference group. This variable attempts to account for the effects of recent migration – either domestic or international – on the likelihood of having a year of naturalization reported. Persons recently moved into a household may have a greater social distance from the primary respondent making it less likely that the respondent would know what year they became a citizen. Therefore, we hypothesize that naturalized citizens who did not move in the last year will be more likely to provide a response to the year of naturalization question. As shown in Table 2, of the naturalized citizens who moved in the previous year, 15 percent did not have a year of naturalization reported, compared to 12 percent of those who did not move in the previous year.

4.6.3. Nonresponse Controls

In this analysis, we include several dummy variables to control for patterns of nonresponse. These include relationship to the householder, householder's level of education, householder's language spoken at home, householder's nativity status, homeownership, poverty status, age, sex, place of birth, year of entry, and moved in the last year. For each dummy, the reference category is *reported*, indicating an acceptable response was given (i.e., not imputed) for that

[15] In some circumstances, it is possible for legal permanent residents to be eligible to naturalize before completing 5 years of continuous residency in the United States. The primary exception to the 5-year residency requirement is spouses of U.S. citizens, who have a 3-year residency requirement for naturalization. See the "Citizenship Through Naturalization" page on the US Citizenship and Immigration Services website at www.uscis.gov/us-citizenship/citizenship-through-naturalization.

item. Nonresponse controls were not included in the model for three variables: mode, region, and number of adults aged 18 and over. As mode and region are survey operational variables, they do not have missing data. Number of adults is simply a count of persons 18 and over in the household; unfortunately, there is no way to account for adults in the household missed by the survey.[16]

Some researchers address the issue of item nonresponse by discarding incomplete cases from their sample (Dillman et al. 2002). We included nonresponse controls to maximize sample size. By controlling for imputation, we are able to retain those cases with some missing information, rather than excluding them from the analysis and losing the data that were provided. In addition, we can control for the correlation of nonresponse between the dependent and each explanatory variable. It is possible that naturalized citizens who don't have a reported year of naturalization also may be very likely not to have an answer for a related variable, such as year of entry or place of birth, while other information, such as age and sex, is provided. For example, 70 percent of naturalized citizens who did not have a reported year of entry also did not have a reported year of naturalization. In addition, 35 percent of those who did not have a reported country of birth also did not have a reported year of naturalization. However, similar rates of nonresponse for those who did not have a reported age and sex were notably lower (27 percent and 15 percent, respectively).

[16] The correlations among all of the explanatory variables included in the models were analyzed and were below an acceptable level (below 0.4).

5. Results

Table 3 shows the results of the logistic regression analyses. A total of nine models are shown: model 1 and four others (A models) without the nonresponse controls and another four (B models) with the nonresponse control variables.

5.1. Primary Explanatory Variables

Both primary explanatory variables – mode and relationship to the householder – are significant in all of the models. As hypothesized, mode has a significant influence on the likelihood that foreign-born citizens have a year of naturalization reported. As shown in Model 5-A, when compared to households that responded using a mail-back questionnaire, naturalized citizens who lived in households interviewed through either CATI or CAPI were significantly less likely to have a reported year of naturalization. Specifically, naturalized citizens living in a CATI household were 91 percent less likely and those living in a CAPI household were 142 percent less likely to have a reported year of naturalization than those responding by mail-back questionnaire.[17] Our second primary explanatory variable – relationship to householder – also behaves as expected. When compared to householders, immediate family members were 22 percent, extended family members 95 percent, other relatives 177 percent and non-relatives were 158 percent less likely than the householder to have a reported year of naturalization.

[17] For ease of interpretability, some of the odds ratios discussed in the text are the inverse of the odds ratios presented in Table 3. For example, the inverse of the odds ratio .523 is 1.912, interpreted as respondents in CATI households being 91 percent less likely than respondents in mail-back households to report a year of naturalization.

When the nonresponse control variables are added to the final model, both mode and relationship to the householder remain significant. However, the size of the negative coefficients are reduced, suggesting that the increased likelihood of not having a year of naturalization reported is tempered by the addition of the control variables. Model 5-B shows that, when compared to naturalized citizens living in households that responded by mail-back questionnaires, CATI households were 62 percent and CAPI households were 71 percent less likely to have reported a year of naturalization for their naturalized citizen members. The odds ratios of the social distance variable were also reduced. Immediate family members were 19 percent, extended family members 83 percent, other relatives 110 percent, and nonrelatives 124 percent less likely than the householder to have a year of naturalization reported.[18]

5.2. Household Characteristics

All of the household-level controls are significant and most behaved as predicted. As shown in Model 5-A, as hypothesized, the householder's educational attainment does positively influence year of naturalization reporting. Naturalized citizens in households with householders who had a high school degree or higher education were 6 percent more likely to have a year of naturalization reported than those living with householders who had less than a high school degree. Also, the number of adults showed the expected results, with each additional adult in the household decreasing the likelihood that foreign-born citizens had a reported year of naturalization. Both homeownership and poverty status influence the likelihood of reporting year

[18] In model 5-B, the coefficients for other relatives and nonrelatives are not statistically different from one another.

of naturalization in the expected direction as well. Naturalized foreign born who live in households that are owned were 7 percent more likely to have a reported year of naturalization, while those living in households not in poverty were 19 percent more likely.

However, two variables – householder's language spoken at home and householder's nativity status – demonstrated results opposite of what we hypothesized. Naturalized citizens living in households where the householder spoke only English at home and where the householder was native born were less – not more – likely to have a reported year of naturalization. Perhaps households with foreign-born householders who speak a language other than English at home have a greater orientation towards immigrant culture and a stronger awareness of the naturalization process. If so, household members would likely have greater knowledge of the details associated with each other's citizenship status, which could boost the response rate associated with year of naturalization reporting. We recommend more research on this topic.

The addition of the nonresponse controls to the final model does alter the likelihoods associated with householder's education and number of adults in the household. In Model 5-B, householder's level of education no longer significantly influences the likelihood a respondent in the household will report a year of naturalization for members who are naturalized citizens. In addition, the coefficient for the number of adults in the household is now positive, suggesting that more adults increases the likelihood naturalized citizens will have a year of naturalization reported. This is interesting, as the number of adults does not have a companion nonresponse control. These results seem to suggest that – all other things being equal – the number of adults in a household *alone* does not increase nonresponse rate but may actually work to reduce it.

5.3. Individual Characteristics of Naturalized Citizen Household Members

Like the household variables, all of the individual-level variables are significant. As shown in Table 3 (Model 5-A), age follows the hypothesized pattern, with naturalized citizens in younger age groups more likely to have a reported year of naturalization than those in older age groups.[19] The variable moved in the last year also demonstrated the expected effect: naturalized foreign born who did not move in the last year were more likely to have a reported year of naturalization. However, males were less likely than ~~females which~~females, which is opposite the results generally found in the nonresponse literature. In addition, the period of entry variable did not exhibit the expected pattern: naturalized citizens who entered between 2007 and 2011 were much less likely to have a reported year of naturalization than those who entered in 1996 or earlier.

When the nonresponse controls are included in the final model, however, the likelihoods associated with age, sex, and especially period of entry change. In Model 5-B, only naturalized citizens 45 to 59 are now significantly more likely to have a reported year of naturalization than those 60 and older, with the remaining age groups no longer significant. While this generally supports the negative association between age and year of naturalization reporting, it appears to only matter in the older age groups. Also in Model 5-B, the sign of the sex coefficient flips, with males now more likely than females to have a year of naturalization reported, which was the association originally hypothesized. Additionally, the period of entry dummy coefficients exhibit a more plausible pattern. Once the nonresponse controls were added to the model, the

[19] The coefficient for the 30-44 age group did not differ statistically from the coefficient for the 45-49 age group.

hypothesized association emerged, with naturalized citizens who entered between 1997 and 2001 being 54 percent more likely and those entering between 2002 and 2006 72 percent more likely to have a year of naturalization reported than those who arrived in 1996 or before. The coefficient for the 2007 to 2011 entry period is no longer significant.

6. Discussion and Conclusion

The results of the logistic regression analysis demonstrate that mode has a significant and substantive effect on whether or not naturalized citizens have a reported year of naturalization. Naturalized foreign born living in households who respond to the American Community Survey by mail-back questionnaires are more likely than those who respond through CATI or CAPI to report the years their members obtained citizenship. This association remains significant even after the inclusion of the nonresponse controls, although the difference between the mail and CATI/CAPI likelihoods is somewhat diminished.

It could be argued that differences by mode simply reflect household response behavior. That is, households willing to participate in the ACS would likely complete the mail-back questionnaire. Those less able or willing to complete a paper questionnaire may be willing to be interviewed by phone. Those less able or willing to be surveyed by CATI may respond to face-to-face interviews. Those who refuse to participate through any mode may simply fall out of the sample. In other words, it is likely that households that are difficult to survey – for whatever reason – are more likely to be surveyed by CATI or CAPI rather than mail-back questionnaires.

But the different modes used are more than just alternative-yet-equal methods to collect data as they represent three different ways the U.S. Census Bureau interacts with households and their members. In addition, they present households with fundamentally different ways to pass the information requested from each household member to the survey instrument. In the mail mode, household members provide their information either indirectly through a primary respondent or directly through the questionnaire, with access to the instrument lasting as long as the household takes to return the form. In CATI and CAPI, household members are most likely to provide their information only through a primary respondent, with access to the interviewer limited to one or a few short interview sessions. This would suggest that foreign-born citizens in households that respond by mail would be more likely than those that respond by CATI or CAPI to have a reported year of naturalization. The results of the logistic regression analysis support this view, albeit indirectly.

Additional support for this model comes from the significant results associated with the "social distance from primary respondent" variable. As hypothesized, the farther naturalized citizen household members are from the householder in social distance, the less likely they are to have a reported year they obtained citizenship. This likely reflects, at least in part, the increasingly weak ties as relationships move from those between immediate family members to those between more distant kin or between nonrelatives. Put another way, the householder knows less and less about others in the household as relationships become more distant. This would, in turn, diminish the likelihood that timely and accurate information would be reported on the survey instrument. The results of the logistic regression analysis show that – even with the household, individual, and nonresponse controls – naturalized citizens who were immediate family members, followed by

extended family members, and other relatives and nonrelatives, were less likely than the householder to have a reported year of naturalization.

Initially, the results of this research seem to contradict the more conventional wisdom about item nonresponse by mode. According to the survey methodology literature, item nonresponse is a larger problem for mail surveys, especially when compared with telephone or face-to-face interviewing (de Leeuw 1992, de Leeuw, Hox, and Huisman 2003), and other self-administered questionnaires (Tourangeau, Rips, and Rasinski 2000). However, in our paper, the logistic regression analysis suggests that naturalized citizens living in households that respond to the ACS using mail-back questionnaires exhibit *lower* rates of item nonresponse than those living in households responding by either CATI or CAPI. While initially inconsistent, we argue that the results described in the literature as well as our research are both correct and likely reflect differences in how item nonresponse is conceptualized.

In general, the survey research literature views item nonresponse as respondent behavior significantly influenced by individual characteristics (e.g., age, sex, socioeconomic status, etc.). When a questionnaire is self-administered, the respondent has direct access to the survey instrument and, as suggested by Tourangeau, Rips, and Rasinski (2000), control over which questions are read, in what order, and whether or not an answer is recorded. By comparison, during phone or in-person interviews, whether an individual answers a question or not is mitigated by the presence of an interviewer who reads the script, clarifies confusing concepts, and encourages respondents to provide complete answers. Thus, phone and in-person interviews

that directly survey individuals could, in some instances, have a lower overall item nonresponse rate than self-administered questionnaires, as has been suggested by the literature.

However, when considering data collection within the household context, the relationship between respondent and survey instrument changes. Most households have more than one member so the majority of people who are part of household surveys 1) are not directly interviewed and 2) do not have direct access to the survey instrument or interaction with the interviewer. As discussed earlier, most information about a household and its members included on a questionnaire or passed to an interviewer is through a primary respondent. Of the three modes used in the ACS, self-administered questionnaires – which can remain with the household for an extended period – offer the greatest likelihood that more than one household member would access the instrument and/or provide information directly to the primary respondent to place on the form. More people with more time providing more accurate and timely information could reduce the rate of item nonresponse for all individuals in the household.

By comparison, phone and in-person interviews are usually completed with a single respondent and during a delimited period of time. While it is possible that other household members are present during phone and in-person interviews, their participation is likely limited. In this sense, a primary respondent with limited response time providing information that is pulled predominantly from memory would likely result in a higher item nonresponse rate for all members of the household. The view that item nonresponse is more than an individual-level characteristic and is significantly influenced by the wider social context in which response behavior occurs is supported by the results of the logistic analysis presented in this study.

Placing the response behavior of individuals within a household survey environment can also help explain differential item nonresponse in two ways. First, it is likely that a primary respondent can accurately provide, in addition to his/her own information, information about the more obvious characteristics of other household members, such as sex and race. It is also likely that the primary respondent can approximate or guess, with reasonable accuracy, certain other characteristics, such as age, nativity status, and country of birth. In this way, nonresponse for such variables would be comparatively lower than the overall rate, a supposition supported by the pattern of imputation for these items in the ACS. Second, knowledge about more obvious characteristics would require less interaction and familiarity with other household members than those characteristics that may not be generally known or discussed (such as year of naturalization), may be considered sensitive (such as citizenship status), or both (such as personal income). Thus, information likely not discussed between socially-distant household members would be more likely to not be provided by the primary respondent, resulting in an overall higher nonresponse rate for such items for some relationship groups, such as distant relatives or nonrelatives. The negative association between social distance from the primary respondent and the likelihood of having a reported year of naturalization demonstrated in this research supports this conclusion.

References

Baker, B. (2012). Estimates of the Naturalized Citizen Population: Comparisons Between the ACS and Administrative Records. Poster presented at the Annual Meeting of the Population Association of America, San Francisco, CA. May 2012.

Bell, R. (1984). Item Nonresponse in Telephone Surveys: An Analysis of Who Fails to Report Income. Social Science Quarterly, 65, 207-215.

Bishop, G., Tuchfarber, A., and Oldendick, R. (1986). Opinions on Fictitious Issues: The Pressure to Answer Survey Questions. Public Opinion Quarterly, 50, 240-250.

Craig, C. and McCann, J. (1978). Item Nonresponse in Mail Surveys: Extent and Correlates. Journal of Marketing Research, 15, 285-289.

de la Puente, M. (1993). Why Are People Missed or Erroneously Enumerated in the Census: A Summary of Findings From Ethnographic Research. In Proceedings of the 1993 Research Conference on Undercounted Ethnic Populations, May 5-7. 29-66.

de Leeuw, E. (1992). Data Quality in Mail, Telephone, and Face-to-Face Surveys. Amsterdam: TT-Publications.

de Leeuw, E., Hox, J., and Huisman, M. (2003). Prevention and Treatment of Item Nonresponse. Journal of Official Statistics, 19, 153-176.

Dillman, D., Eltinge, J., Groves, R., and Little, R. (2002). Survey Nonresponse in Design, Data Collection, and Analysis. In Survey Nonresponse, R. Groves, D. Dillman, J. Eltienge and R. Little (eds). New York: Wiley, 3-26.

Dodge, J., Clark, N., Janz, N., Liang, J., and Schork, M. (1993). Nonparticipation of Older Adults in a Heart Disease Self-Management Project: Factors Influencing Involvement. Research on Ageing, 15, 220-237.

Fein, D. and West, K. (1988). Towards a Theory of Coverage Error: An Exploratory Assessment of Data from the 1986 Los Angeles Test Census. In Proceedings of the Fourth Annual Research Conference, March 20-23. 540-562.

Ferber, R. (1966). Item Nonresponse in a Consumer Survey. Public Opinion Quarterly, 30, 399-415.

Francis, J. and L. Busch. (1975). What We Now Know About 'I Don't Knows.' Public Opinion Quarterly, 39, 207-218.

Francis, L. and M. Robbins. (1995). Survey Response Rate as a Function of Age: Are Female Clergy Different? Psychological Reports, 77, 499-506.

Goyder, J. (1982). Further Evidence of Factors Affecting Response Rates to Mailed Questionnaires. American Sociological Review, 47, 550-553.

Goyder, J. (1987). The Silent Minority. Boulder, CO: Westview Press.

Grieco, E. and Rytina, N. (2011). U.S. Data Sources on the Foreign Born and Immigration. International Migration Review, 45, 1001-1016.

Groves, R. and Couper, M. (1998). Nonresponse in Household Surveys. New York: Wiley.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., and Tourangeau, R. (2004) Survey Methodology. New York: Wiley.

Harris, P., Bhaskar, R., Shook-Finucane, C. and Ericson, L. (2007) Evaluation Report Covering Place of Birth, U.S. Citizenship Status, and Year of Arrival: Final Report. Washington, D.C.: U.S. Census Bureau. (2006 American Community Survey Content Test Report P.1)

Hill, J., Lestina, F., Machowski, J., Rothhaus, C., and Roye, K. (2008). Study of Respondents Who List Themselves as Person 1. Washington, D.C: U.S. Census Bureau. (DSSD 2008 Memorandum Series # G-09)

Johnson, T., O'Rourke, D., Chavez, N., Sudman, S., Warnecke, R., Lacey, L., and Horm, J. (1997). Social Cognition and Response to Survey Questions Among Culturally Diverse Populations. In Survey Measurement and Process Quality, L. Lyberg, P. Biemer, and M. Collins (eds). New York: Wiley, 87-113.

Johnson, T., O'Rourke, D., Burris, J., and Owens, L. (2002). Culture and Survey Nonresponse. In Survey Nonresponse, R. Groves, D. Dillman, J. Eltienge, and R. Little (eds). New York: Wiley, 55-69.

Kaldenberg, D., Koenig, H., and Becker, B. (1994). Mail Survey Response Rate Patterns in a Population of Elderly: Does Response Deteriorate with Age? Public Opinion Quarterly, 58, 68-76.

McDaniel, S., Madden, C., and Verille, P. (1987). Do Topic Differences Affect Survey Nonresponse? Journal of the Market Research Society, 29, 55-66.

Messer, B., Edwards, M., and Dillman, D. (2012). Determinants of Item Nonresponse to Web and Mail Respondents in Three Addressed-Based Mixed Mode Surveys of the General Public. Pullman: Washington State University. (Social and Economic Sciences Research Center Technical Report 12-001)

Messmer, D. and Seymour, D. (1982). The Effects of Branching on Item Nonresponse. Public Opinion Quarterly, 46, 270-277.

Rowland, M. and Forthofer, R. (1993). Investigation of Nonresponse Bias: Hispanic Health and Nutrition Examination Survey. Vital and Health Statistics, 119, 1-75.

Schwede, L. (2003). Complex Households and Relationships in the Decennial Studies and in Ethnographic Studies of Six Race/Ethnic Groups. Washington, D.C.: U.S. Census Bureau. (Census 2000 Testing, Experimentation and Evaluation Program Ethnographic Research Report)

Schwede, L., Blumberg, R., and Chan, A. (eds). (2006). Complex Ethnic Households in America. Lanham, MD: Rowman and Littlefield Publishing.

Schwede, L. and Terry, R. (2013). Comparative Ethnographic Studies of Enumeration Methods and Coverage Across Race and Ethnic Groups. Washington, D.C.: U.S. Census Bureau. (2010 Census Program for Evaluation and Experiments)

Shoemaker, P., Eichholz, M., and Skewes, E. (2002). Item Nonresponse Distinguishing Between Don't Know and Refuse. International Journal of Public Opinion Research, 14, 193-201.

Schuman, H. and Presser, S. (1996). Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. New York: Academic Press.

Tourangeau, R., Rips, L., and Rasinksi, K. (2000). The Psychology of Survey Response. Cambridge: Cambridge University Press.

U.S. Census Bureau. (2009). Design and Methodology: American Community Survey. Washington, D.C.: U.S. Government Printing Office. Available at: www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology.pdf

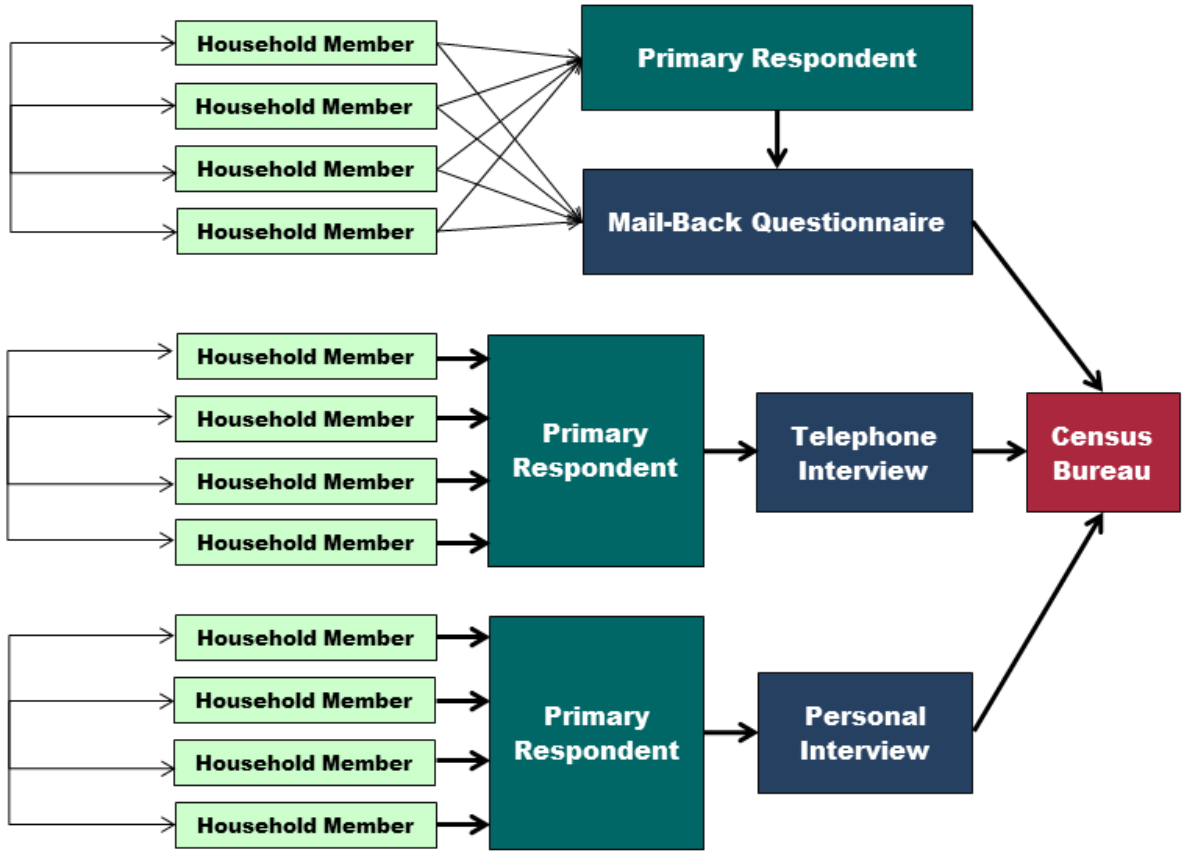
Van Hook, J. and Bachmeier, J. (2013) How Well Does the American Community Survey Count Naturalized Citizens? Demographic Research, 29, 1-32.

Table 1.
Imputation Rates for Selected Variables: 2008 to 2011

Rate	2008	2009	2010	2011
Overall person characteristic	6.2	4.9	5.8	5.8
Year of naturalization	18.2	14.7	16.6	17.0
Year of entry	9.9	8.7	9.8	9.9
Year last married	8.3	6.8	7.7	7.7
Year householder moved into unit	1.9	2.4	3.4	3.2
Citizenship				
Total	2.5	2.0	2.7	2.7
Native	2.2	1.8	2.4	2.4
Foreign born	4.5	3.8	4.6	4.5

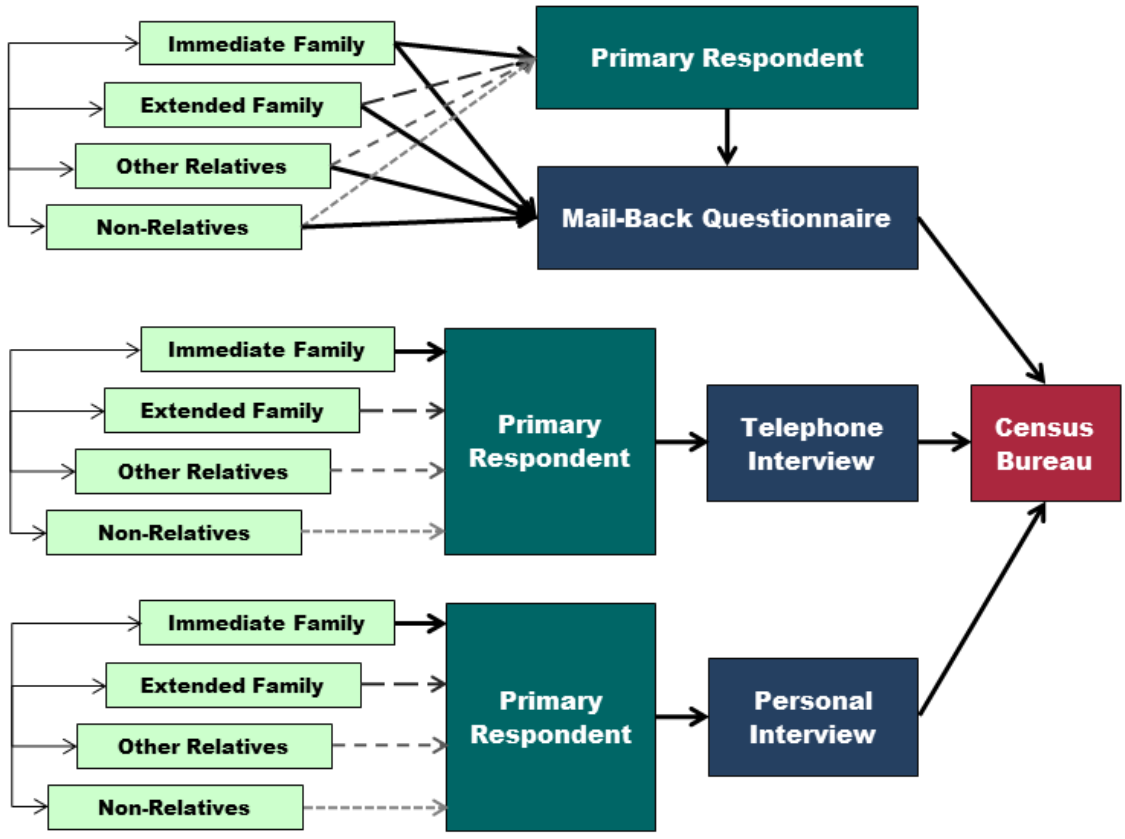
Source: U.S. Census Bureau, 2008 to 2011 American Community Survey.

Figure 1.



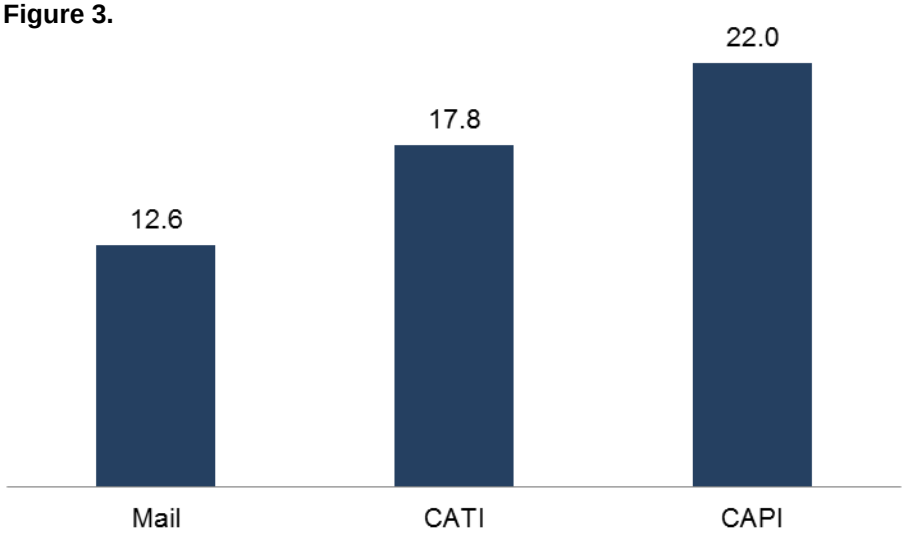
Flow of Information From Household Members to the U.S. Census Bureau

Figure 2.



Flow of Information From Household Members to the U.S. Census Bureau, Showing Social Distance From the Primary Respondent

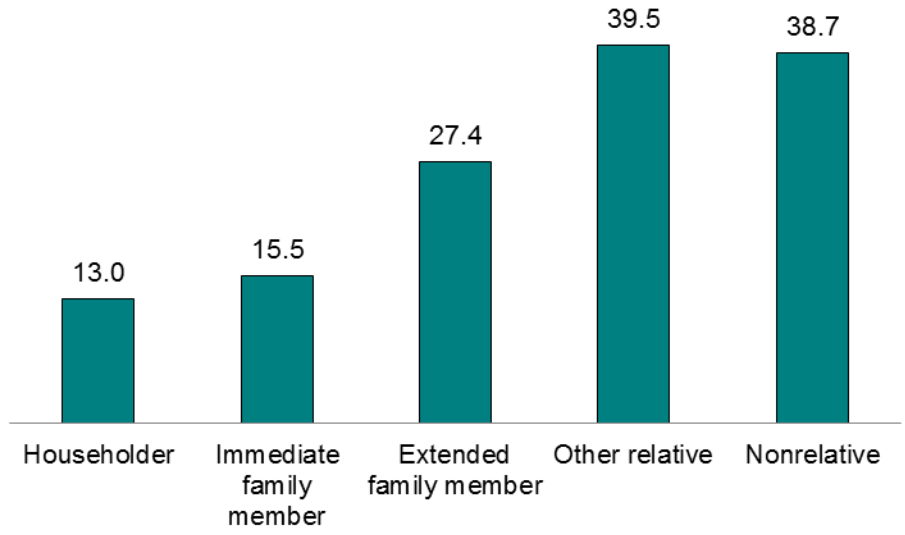
Figure 3.



Item Nonresponse Rate for Naturalized Citizens for Year of Naturalization by Mode: 2011

Note: CATI = Computer Assisted Telephone Interview. CAPI = Computer Assisted Phone Interview.
Source: U.S. Census Bureau, 2011 American Community Survey.

Figure 4.



Item Nonresponse Rate for Naturalized Citizens for Year of Naturalization by Relationship to Householder: 2011

Source: U.S. Census Bureau, 2011 American Community Survey.

Table 2.

Descriptive Statistics for Variables Used in the Model for Year of Naturalization Reporting: 2011

(Numbers in thousands)

Variable	Total naturalized		Reported year of naturalization: Yes		No			
	Number	Percent	Number	Percent	Number	Percent	Percent of total	Margin of error
Total sample size	16,398	100.0	14,331	100.0	2,067	100.0	12.6	0.2
Survey environment								
Survey mode								
Mail (reference)	8,356	51.0	7,700	53.7	656	31.7	7.8	0.2
Computer assisted telephone interview (CATI)	1,769	10.8	1,512	10.6	257	12.4	14.5	0.4
Computer assisted personal interview (CAPI)	6,273	38.3	5,119	35.7	1,154	55.8	18.4	0.4
Social distance from primary respondent								
Relationship to householder								
Householder (reference)	8,380	51.1	7,519	52.5	861	41.7	10.3	0.2
Immediate family member	5,825	35.5	5,124	35.8	700	33.9	12.0	0.3
Extended family member	1,297	7.9	1,038	7.2	258	12.5	19.9	0.6
Other relative	309	1.9	219	1.5	90	4.4	29.2	1.9
Nonrelative	587	3.6	431	3.0	156	7.6	26.6	1.3
Household characteristics								
Region of residence								
Northeast	4,005	24.4	3,418	23.9	587	28.4	14.7	0.4
Midwest	1,809	11.0	1,608	11.2	201	9.7	11.1	0.4
South	4,710	28.7	4,143	28.9	567	27.4	12.0	0.4
West (reference)	5,874	35.8	5,162	36.0	712	34.5	12.1	0.3
Number of adults aged 18 and over								
1	2,009	12.2	1,789	12.5	220	10.6	11.0	0.4
2	7,714	47.0	6,842	47.7	872	42.2	11.3	0.2
3	3,475	21.2	2,990	20.9	485	23.5	14.0	0.4
4	2,004	12.2	1,708	11.9	296	14.3	14.8	0.5
5 or more	1,196	7.3	1,003	7.0	193	9.3	16.1	0.9
Homeownership								
Owned	11,622	70.9	10,263	71.6	1,359	65.8	11.7	0.2
Rented (reference)	4,776	29.1	4,068	28.4	708	34.2	14.8	0.4
Poverty status								
In poverty (reference)	1,741	10.6	1,473	10.3	268	13.0	15.4	0.6
Not in poverty	14,656	89.4	12,858	89.7	1,798	87.0	12.3	0.2
Householder's level of education								
Less than a high school degree (reference)	3,085	18.8	2,654	18.5	431	20.8	14.0	0.4
High school degree or higher	13,313	81.2	11,677	81.5	1,636	79.2	12.3	0.2
Householder's language spoken at home								
Speaks only English at home	3,873	23.6	3,294	23.0	579	28.0	14.9	0.4
Speaks a language other than English at home	12,524	76.4	11,037	77.0	1,487	72.0	11.9	0.2
Householder's nativity status								
Native	2,211	13.5	1,828	12.8	384	18.6	17.4	0.5
Foreign born (reference)	14,186	86.5	12,504	87.2	1,683	81.4	11.9	0.2
Individual characteristics of naturalized citizens								
Age								
18-29	1,549	9.4	1,349	9.4	200	9.7	12.9	0.7
30-44	4,494	27.4	3,917	27.3	578	28.0	12.9	0.4
45-59	5,427	33.1	4,751	33.1	676	32.7	12.5	0.3
60 and older (reference)	4,928	30.1	4,315	30.1	613	29.6	12.4	0.3
Sex								
Male	7,482	45.6	6,548	45.7	933	45.2	12.5	0.2
Female (reference)	8,916	54.4	7,783	54.3	1,133	54.8	12.7	0.2
Place of birth								
Asia	6,065	37.0	5,400	37.7	664	32.1	11.0	0.3
Europe (reference)	2,790	17.0	2,509	17.5	281	13.6	10.1	0.3
Latin America	6,455	39.4	5,457	38.1	998	48.3	15.5	0.4
Other	1,089	6.6	965	6.7	124	6.0	11.4	0.6

Source: U.S. Census Bureau, 2011 American Community Survey. For more information, see *www.census.gov/acs/www/.*

Table 2.

Descriptive Statistics for Variables Used in the Model for Year of Naturalization Reporting: 2011

(Numbers in thousands)

(continued)

Variable	Total naturalized		Reported year of naturalization: Yes		No			
	Number	Percent	Number	Percent	Number	Percent	Percent of total	Margin of error
Period of entry								
2007-2011	156	1.0	85	0.6	71	3.4	45.4	2.8
2002-2006	813	5.0	709	4.9	105	5.1	12.9	0.9
1997-2001	1,905	11.6	1,705	11.9	201	9.7	10.5	0.5
1996 and before (reference)	13,523	82.5	11,833	82.6	1,691	81.8	12.5	0.2
Moved in the last year								
Yes (reference)	1,536	9.4	1,306	9.1	230	11.1	15.0	0.8
No	14,862	90.6	13,026	90.9	1,837	88.9	12.4	0.2
Nonresponse control variables								
Relationship to householder								
Reported (reference)	16,201	98.8	14,182	99.0	2,019	97.7	12.5	0.2
Imputed	197	1.2	149	1.0	48	2.3	24.2	2.5
Homeownership								
Reported (reference)	15,712	95.8	13,735	95.8	1,977	95.7	12.6	0.2
Imputed	685	4.2	596	4.2	89	4.3	13.0	0.8
Poverty status								
Reported (reference)	11,159	68.1	10,312	72.0	847	41.0	7.6	0.1
Imputed	5,239	31.9	4,019	28.0	1,219	59.0	23.3	0.4
Householder's level of education								
Reported (reference)	15,954	97.3	14,066	98.1	1,888	91.3	11.8	0.2
Imputed	444	2.7	265	1.9	179	8.7	40.3	1.7
Householder's language spoken at home								
Reported (reference)	16,237	99.0	14,211	99.2	2,026	98.0	12.5	0.2
Imputed	161	1.0	120	0.8	40	2.0	25.1	2.4
Householder's nativity status								
Reported (reference)	16,285	99.3	14,242	99.4	2,043	98.9	12.5	0.2
Imputed	113	0.7	89	0.6	24	1.1	21.0	2.3
Age								
Reported (reference)	16,041	97.8	14,069	98.2	1,972	95.4	12.3	0.2
Imputed	357	2.2	262	1.8	95	4.6	26.5	1.4
Sex								
Reported (reference)	16,328	99.6	14,272	99.6	2,056	99.5	12.6	0.2
Imputed	70	0.4	60	0.4	10	0.5	14.7	2.2
Place of birth								
Reported (reference)	16,196	98.8	14,199	99.1	1,996	96.6	12.3	0.2
Imputed	202	1.2	132	0.9	70	3.4	34.6	1.9
Year of entry								
Reported (reference)	15,375	93.8	14,026	97.9	1,349	65.3	8.8	0.1
Imputed	1,023	6.2	305	2.1	718	34.7	70.2	0.8
Moved in the last year								
Reported (reference)	16,011	97.6	14,015	97.8	1,996	96.6	12.5	0.2
Imputed	387	2.4	316	2.2	71	3.4	18.3	1.0

Source: U.S. Census Bureau, 2011 American Community Survey. For more information, see *www.census.gov/acs/www/.*

Table 3.
Logistic Regression Analysis of Foreign-Born Citizens Reporting Year of Naturalization: 2011

	Model 1			Model 2-A			Model 2-B		
Variable	b	S.E.	Odds ratio	b	S.E.	Odds ratio	b	S.E.	Odds ratio
Survey mode									
CATI	-0.691 ***	0.013	0.501						
CAPI	-0.973 ***	0.011	0.378						
Relationship to householder									
Immediate family				-0.177 ***	0.010	0.838	-0.177 ***	0.010	0.838
Extended family				-0.776 ***	0.013	0.460	-0.775 ***	0.013	0.461
Other relative				-1.280 ***	0.028	0.278	-1.269 ***	0.033	0.281
Nonrelative				-1.154 ***	0.020	0.316	-1.152 ***	0.021	0.316
Household characteristics									
Region of residence									
Northeast									
Midwest									
South									
Number of adults aged 18 and over									
Homeownership									
Poverty status									
Householder's level of education									
Householder's language spoken at home									
Householder's nativity status									
Individual characteristics of naturalized citizens									
Age									
18-29									
30-44									
45-59									
Sex									
Place of birth									
Asia									
Latin America									
Other									
Period of entry									
2007-2011									
2002-2006									
1997-2001									
Moved in the last year									
Nonresponse controls									
Relationship							-0.037	0.051	0.964
Homeownership									
Poverty status									
Householder's level of education									
Householder's language spoken at home									
Householder's nativity status									
Age									
Sex									
Place of birth									
Year of entry									
Moved in the last year									
Intercept	2.463 ***	0.007		2.167 ***	0.007		2.167 ***	0.007	
N (weighted, in thousands)	16,398			16,398			16,398		
-2 log likelihood (in thousands)	12,051			12,179			12,179		
Pseudo R^2	0.030			0.020			0.020		
Somer's D	0.216			0.119			0.121		
Chi-square (Wald statistic)	8,972 ***			7,676 ***			7,745		
Degrees of freedom	2			4			5		

*p <.05; **p <.01; ***p <.001
Source: U.S. Census Bureau, 2011 American Community Survey. For more information, see *www.census.gov/acs/www/.*

Table 3.

Logistic Regression Analysis of Foreign-Born Citizens Reporting Year of Naturalization: 2011

(continued)

	Model 3-A			Model 3-B			Model 4-A		
Variable	b	S.E.	Odds ratio	b	S.E.	Odds ratio	b	S.E.	Odds ratio
Survey mode									
CATI									
CAPI									
Relationship to householder									
Immediate family									
Extended family									
Other relative									
Nonrelative									
Household characteristics									
Region of residence									
Northeast	-0.227 ***	0.011	0.797	-0.147 ***	0.012	0.863			
Midwest	0.054 **	0.016	1.056	0.075 ***	0.016	1.078			
South	-0.039 **	0.014	0.961	-0.012	0.015	0.988			
Number of adults aged 18 and over	-0.154 ***	0.004	0.858	-0.040 ***	0.004	0.961			
Homeownership	0.303 ***	0.013	1.353	0.267 ***	0.013	1.306			
Poverty status	0.256 ***	0.016	1.291	0.302 ***	0.015	1.352			
Householder's level of education	0.144 ***	0.013	1.155	0.064 ***	0.012	1.066			
Householder's language spoken at home	-0.234 ***	0.013	0.792	-0.208 ***	0.013	0.812			
Householder's nativity status	-0.405 ***	0.015	0.667	-0.517 ***	0.016	0.596			
Individual characteristics of naturalized citizens									
Age									
18-29							0.107 ***	0.023	1.113
30-44							0.033 *	0.013	1.033
45-59							0.047 ***	0.012	1.048
Sex							0.020 *	0.008	1.020
Place of birth									
Asia							-0.087 ***	0.014	0.917
Latin America							-0.482 ***	0.014	0.618
Other							-0.122 ***	0.021	0.885
Period of entry									
2007-2011							-1.761 ***	0.036	0.172
2002-2006							-0.072 **	0.024	0.931
1997-2001							0.158 ***	0.020	1.171
Moved in the last year							0.211 ***	0.020	1.235
Nonresponse controls									
Relationship									
Homeownership				0.092 ***	0.021	1.096			
Poverty status				-1.243 ***	0.010	0.288			
Householder's level of education				-1.213 ***	0.024	0.297			
Householder's language spoken at home				-0.042	0.049	0.959			
Householder's nativity status				0.111 *	0.051	1.118			
Age									
Sex									
Place of birth									
Year of entry									
Moved in the last year									
Intercept	1.983 ***	0.020		2.294 ***	0.021		1.964 ***	0.025	
N (weighted, in thousands)	16,398			16,398			16,398		
-2 log likelihood (in thousands)	12,234			11,417			12,228		
Pseudo R^2	0.015			0.081			0.016		
Somer's D	0.158			0.371			0.143		
Chi-square (Wald statistic)	4,601 ***			27,126 ***			6,020 ***		
Degrees of freedom	9			14			11		

*p <.05; **p <.01; ***p <.001

Source: U.S. Census Bureau, 2011 American Community Survey. For more information, see *www.census.gov/acs/www/*.

Table 3.

Logistic Regression Analysis of Foreign-Born Citizens Reporting Year of Naturalization: 2011

(continued)

Variable	Model 4-B b		S.E.	Odds ratio	Model 5-A b		S.E.	Odds ratio	Model 5-B b		S.E.	Odds ratio
Survey mode												
CATI					-0.649	***	0.013	0.523	-0.480	***	0.014	0.619
CAPI					-0.882	***	0.011	0.414	-0.538	***	0.013	0.584
Relationship to householder												
Immediate family					-0.200	***	0.011	0.819	-0.173	***	0.012	0.841
Extended family					-0.668	***	0.017	0.513	-0.604	***	0.020	0.547
Other relative					-1.020	***	0.029	0.361	-0.740	***	0.042	0.477
Nonrelative					-0.946	***	0.021	0.388	-0.806	***	0.024	0.447
Household characteristics												
Region of residence												
Northeast					-0.216	***	0.011	0.806	-0.200	***	0.012	0.819
Midwest					-0.035	*	0.016	0.966	-0.054	**	0.017	0.947
South					0.021		0.014	1.021	-0.019		0.015	0.982
Number of adults aged 18 and over					-0.055	***	0.005	0.946	0.043	***	0.005	1.044
Homeownership					0.073	***	0.013	1.075	0.066	***	0.014	1.068
Poverty status					0.172	***	0.016	1.187	0.189	***	0.017	1.208
Householder's level of education					0.058	***	0.013	1.060	-0.016		0.013	0.984
Householder's language spoken at home					-0.300	***	0.013	0.741	-0.255	***	0.017	0.775
Householder's nativity status					-0.158	***	0.016	0.854	-0.122	***	0.021	0.885
Individual characteristics of naturalized citizens												
Age												
18-29	-0.072	**	0.024	0.930	0.317	***	0.023	1.373	0.042		0.024	1.043
30-44	0.004		0.015	1.004	0.068	***	0.014	1.070	-0.022		0.016	0.978
45-59	0.072	***	0.013	1.075	0.044	***	0.012	1.045	0.043	**	0.014	1.044
Sex	0.078	***	0.009	1.081	-0.026	**	0.008	0.974	0.021	*	0.009	1.021
Place of birth												
Asia	-0.122	***	0.014	0.885	-0.118	***	0.014	0.889	-0.208	***	0.015	0.812
Latin America	-0.302	***	0.014	0.739	-0.284	***	0.015	0.753	-0.196	***	0.016	0.822
Other	-0.131	***	0.023	0.877	-0.072	**	0.022	0.930	-0.117	***	0.025	0.889
Period of entry												
2007-2011	-0.207	***	0.051	0.813	-1.485	***	0.039	0.227	-0.059		0.049	0.942
2002-2006	0.472	***	0.026	1.603	0.049	*	0.025	1.050	0.544	***	0.026	1.722
1997-2001	0.406	***	0.022	1.501	0.201	***	0.021	1.222	0.433	***	0.021	1.541
Moved in the last year	0.149	***	0.022	1.161	0.087	***	0.020	1.091	0.076	***	0.022	1.079
Nonresponse controls												
Relationship									0.078		0.059	1.081
Homeownership									-0.085	***	0.024	0.918
Poverty status									-0.903	***	0.011	0.405
Householder's level of education									-0.779	***	0.027	0.459
Householder's language spoken at home									-0.074		0.051	0.929
Householder's nativity status									0.003		0.048	1.003
Age	-0.549	***	0.025	0.578					-0.305	***	0.024	0.737
Sex	-0.045		0.061	0.956					-0.115		0.064	0.892
Place of birth	-0.918	***	0.031	0.399					-0.864	***	0.033	0.422
Year of entry	-3.154	***	0.014	0.043					-2.715	***	0.016	0.066
Moved in the last year	-0.160	***	0.029	0.852					-0.108	***	0.038	0.898
Intercept	2.302	***	0.027		2.717	***	0.034		3.032	***	0.033	
N (weighted, in thousands)	16,398				16,398				16,398			
-2 log likelihood (in thousands)	10,293				11,665				9,719			
Pseudo R^2	0.171				0.061				0.218			
Somer's D	0.373				0.307				0.521			
Chi-square (Wald statistic)	76,873	***			30,201	***			132,009	***		
Degrees of freedom	16				26				37			

*p <.05; **p <.01; ***p <.001

Source: U.S. Census Bureau, 2011 American Community Survey. For more information, see *www.census.gov/acs/www/*.