

An Examination of Imputation Patterns for Data on the Foreign-Born Population in the American Community Survey, 2006-2010

Thomas Gryn, Edward Trevelyan (U.S. Census Bureau)

Jeff Pongsiri, Joanna Kling, Megan Benetsky, Samaneh Tabrizi (University of Maryland), and Savet Hong (University of California, Berkeley)

**Poster Presentation at the Population Association of America Annual Meetings,
May 3-5, 2012**

The American Community Survey (ACS) utilizes data imputation to increase data quality. Imputation involves replacing missing, incomplete, or inconsistent responses with modified values. This poster analyzes the imputation patterns for the foreign-born population using the 2006 through 2010 1-year ACS, focusing on the variables of citizenship (CIT), year of entry (YOE), year of naturalization (CITW), and place of birth (POB). After presenting the longitudinal imputation trends of these four variables for the years 2006 through 2010, imputation patterns for the citizenship variable in 2010 are examined for various demographic characteristics, such as race/ethnicity and language proficiency.¹

As with any survey, the American Community Survey (ACS) faces the problem of responses that are missing, incomplete, or inconsistent. Through a series of data edits developed by Census Bureau subject matter experts, the ACS utilizes data imputation to maintain data quality. The editing process looks at internally contradictory responses and attempts to resolve them, as well as attempting to estimate missing responses from other

¹ Note that full implementation of the ACS sample, including group quarters enumeration, began in 2006.

information wherever possible. Imputation involves changing responses either through assignment or allocation. “Assignment” occurs when a response value is changed based on other information reported by the respondent. For example, if a respondent indicated that his/her place of birth was in the United States, but also responded that his/her citizenship was not “born in the United States” or was missing, the response for citizenship would be imputed to change his/her citizenship response to “Yes, born in the United States.”²

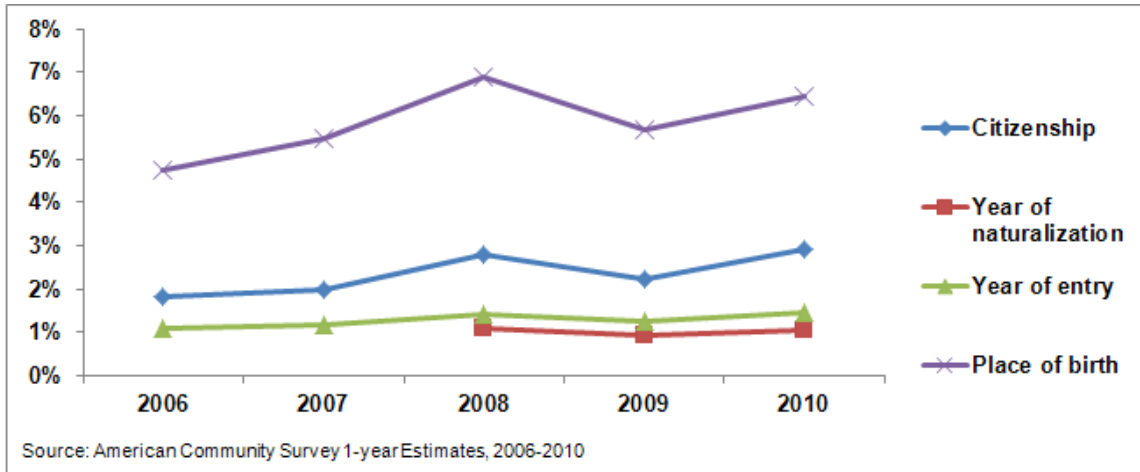
When there is not enough information to assign a value, allocation is used instead. “Allocation” refers to the use of valid values from other respondents who share similar demographic characteristics to the respondent to replace the respondent’s invalid or missing value on the variable.³ For example, a Hispanic respondent age 24 who did not report a value for citizenship may have allocated a valid citizenship response from another 24-year-old Hispanic respondent in the survey.

² With the rationale being that as place of birth is a write-in question and therefore requires more effort and consideration to formulate a response as compared to a check-box question such as citizenship, it is more likely to be the correct response when the two questions give contradictory information.

³ This process is also known as the “sequential hot-deck procedure” (Groves et al., 2004). See Chapter 10 in *Design and Methodology: American Community Survey* (2009) for a more detailed explanation of the imputation procedures used in the ACS.

Imputation Rates for Citizenship, Year of Naturalization, Year of Entry, and Place of Birth in the ACS

Figure 1. Percent Imputed Responses to the Citizenship, Year of Naturalization, Year of Entry, and Place of Birth Questions, United States: 2006-2010



In general, imputation rates for the ACS are relatively low, indicating that responses are infrequently imputed and the quality of the data is relatively high. As shown in Figure 1, all of the four variables of interest – citizenship, year of entry, year of naturalization, or place of birth – showed imputation rates below 7 percent throughout the period 2006-2010, even with modest changes to the questionnaire items in 2008 which included introduction of the year of naturalization section to the citizenship question.^{4, 5, 6,}

7

⁴ Place of birth in 2008 was not significantly different from 7 percent.

⁵ Imputation rates for Figure 1 are calculated for the total population – natives as well as foreign-born.

⁶ See “[Research Note: Changes to the American Community Survey Between 2007 and 2008 and Their Potential Effect on the Estimates of Hispanic Origin Type, Nativity, Race, and Language](http://www.census.gov/population/www/socdemo/hispanic/acs08researchnote.pdf)” at <http://www.census.gov/population/www/socdemo/hispanic/acs08researchnote.pdf> for more information.

⁷ From April to September of 2008, the ACS reduced the number of cases included in the telephone Failed Edit Follow-up (FEFU) operation. In FEFU, households who have returned a mail form but whose data failed a completeness check are called to collect the missing data. This operation is designed to improve the

Note that place of birth comes earlier in the edit process than CIT, CITW, and YOE, and its somewhat higher imputation rate is in part due to this position. For example, respondents with a missing or incomplete value for CIT may later on in the editing process have their CIT value assigned or allocated based on other information provided, but because this occurs later than the edits for POB, imputation of POB cannot take advantage of any later imputation of CIT to further inform POB. Place of birth is also a write-in question, which may cause it to have a wider variation in responses compared to check-box questions such as citizenship (see Figure 2).

Figure 2: Place of Birth, Citizenship, and Year of Entry Questions in the ACS

7 Where was this person born?

In the United States – *Print name of state.*

Outside the United States – *Print name of foreign country, or Puerto Rico, Guam, etc.*

8 Is this person a citizen of the United States?

Yes, born in the United States → *SKIP to 10a*

Yes, born in Puerto Rico, Guam, the U.S. Virgin Islands, or Northern Marianas

Yes, born abroad of U.S. citizen parent or parents

Yes, U.S. citizen by naturalization – *Print year of naturalization*

No, not a U.S. citizen

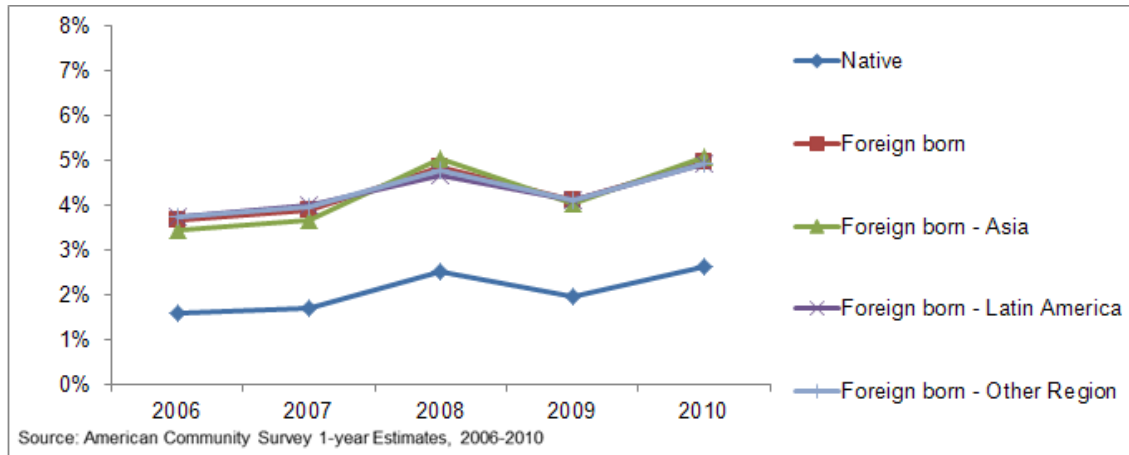
9 When did this person come to live in the United States? Print numbers in boxes.

Year

Source: ACS-1(2009)KFI

final quality of mail-returned questionnaires. As expected, the reduced number of FEFU cases resulted in an increase in missing data among mail respondents and thus an increase in item allocation rates in 2008. For more information about FEFU, please refer to Chapter 7, page 7-5 of the ACS Design and Methodology report at http://www.census.gov/acs/www/methodology/methodology_main/

Figure 3. Percent Imputed Responses to the Citizenship Question by Nativity and Region of Birth, United States: 2006-2010

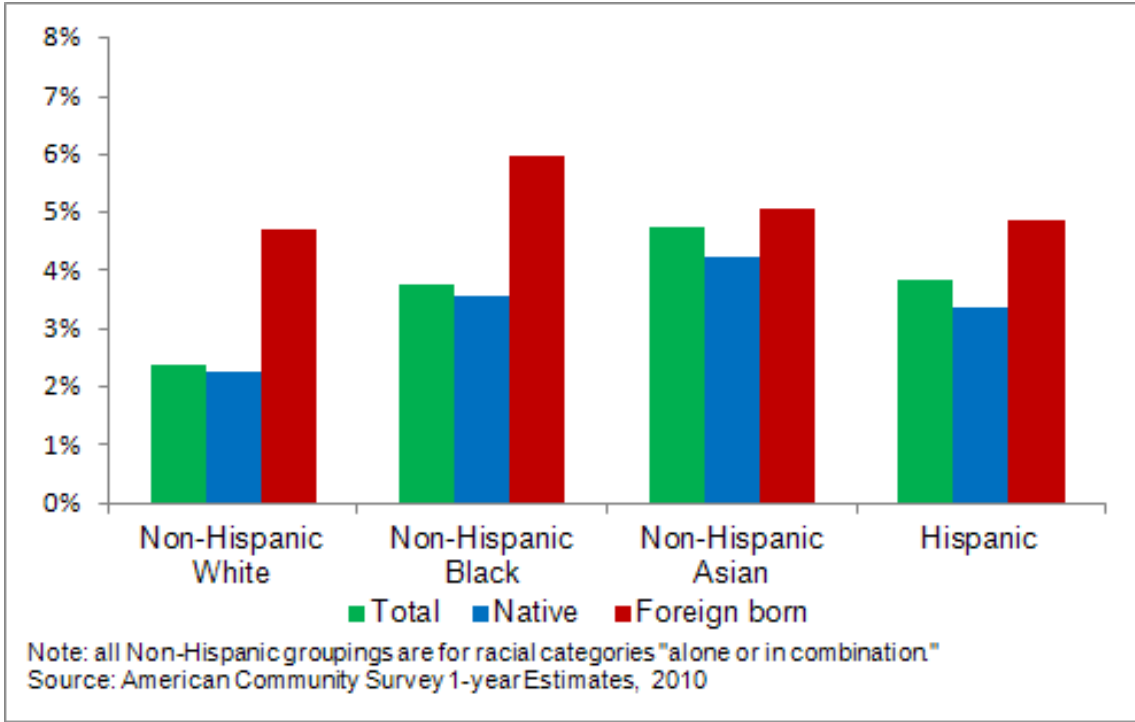


When response rates to the citizenship question are categorized by nativity and region of birth of the foreign-born population, the native born exhibited lower imputation rates on responses to the citizenship question than the foreign born, with rates for the native population between 1 and 3 percent from 2006 through 2010. There was little variation in imputation rates for the foreign born by region of birth in 2009 and 2010.

2010 Imputation by Nativity and Demographic Characteristics

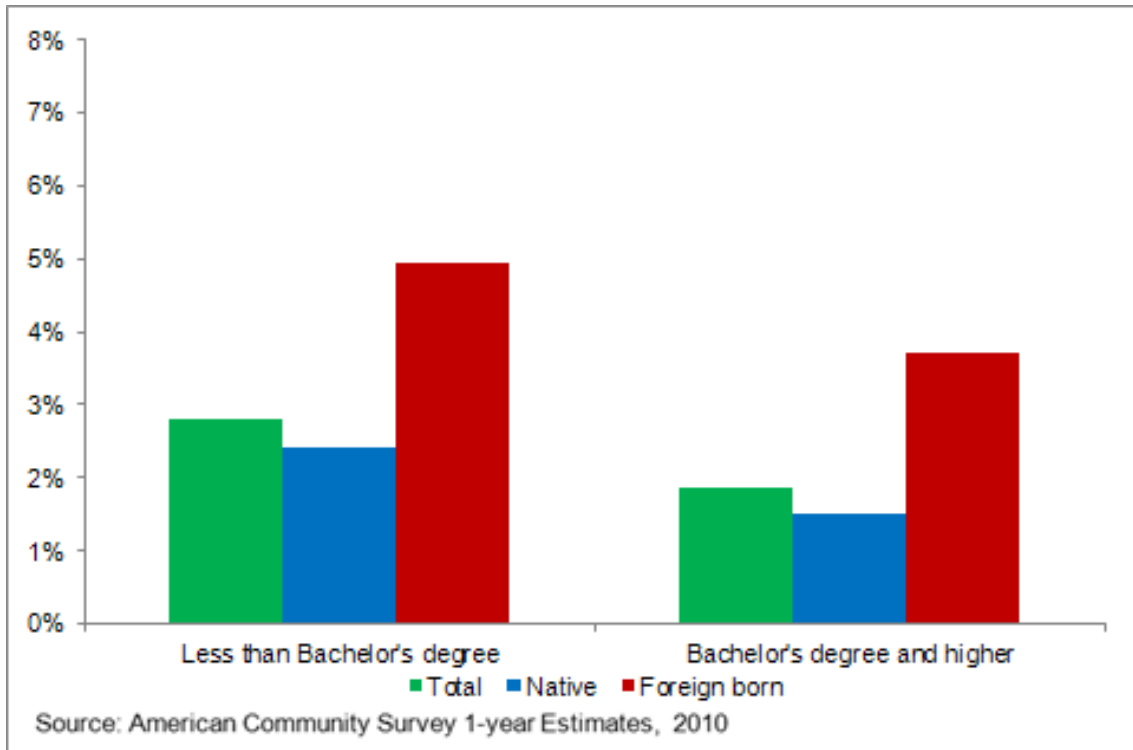
While Figures 1 and 3 displayed imputation patterns year-by-year for 2006 through 2010, Figures 4 through 7 illustrate differences for the ACS survey year 2010 in imputation by nativity and selected basic demographic characteristics, such as race/ethnicity and poverty status.

Figure 4. Percent Imputed Responses to the Citizenship Question by Nativity and Race and Ethnicity, United States: 2010



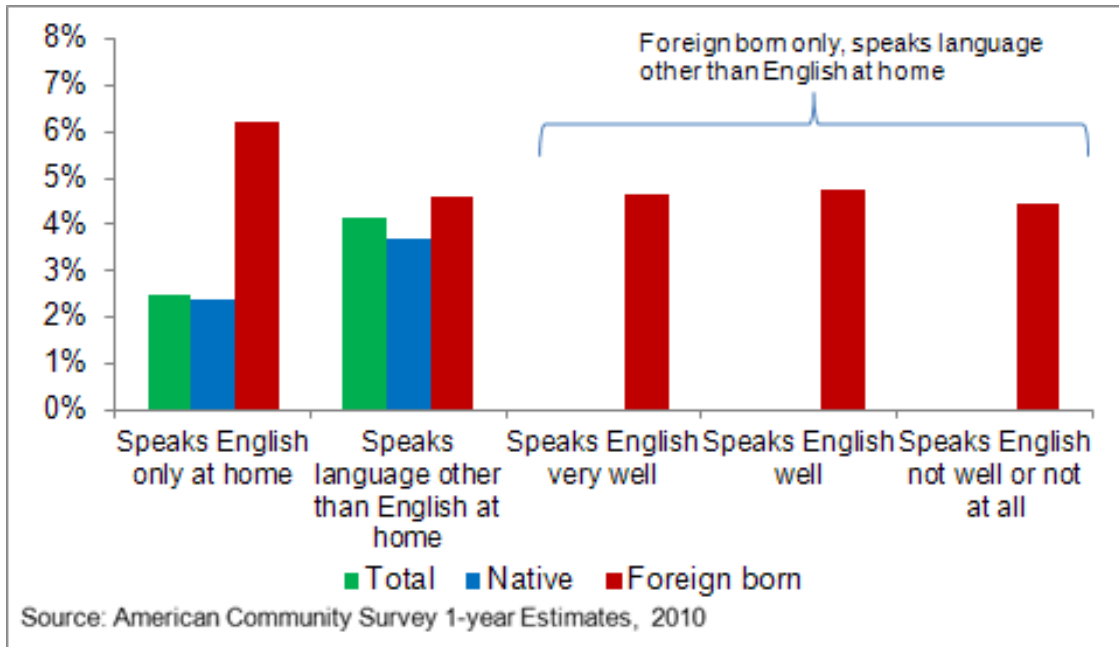
When imputation is analyzed by nativity and race/Hispanic origin, imputation rates for foreign-born non-Hispanic Whites and Blacks were about 2.5 percent higher than for natives, while rates for foreign-born non-Hispanic Asians were about 1 percent higher than for natives. For Hispanics, foreign born had approximately 1.5 percent higher imputation rates than natives.

Figure 5. Percent Imputed Responses to the Citizenship Question by Nativity and Educational Achievement For Those Age 25 and Older, United States: 2010



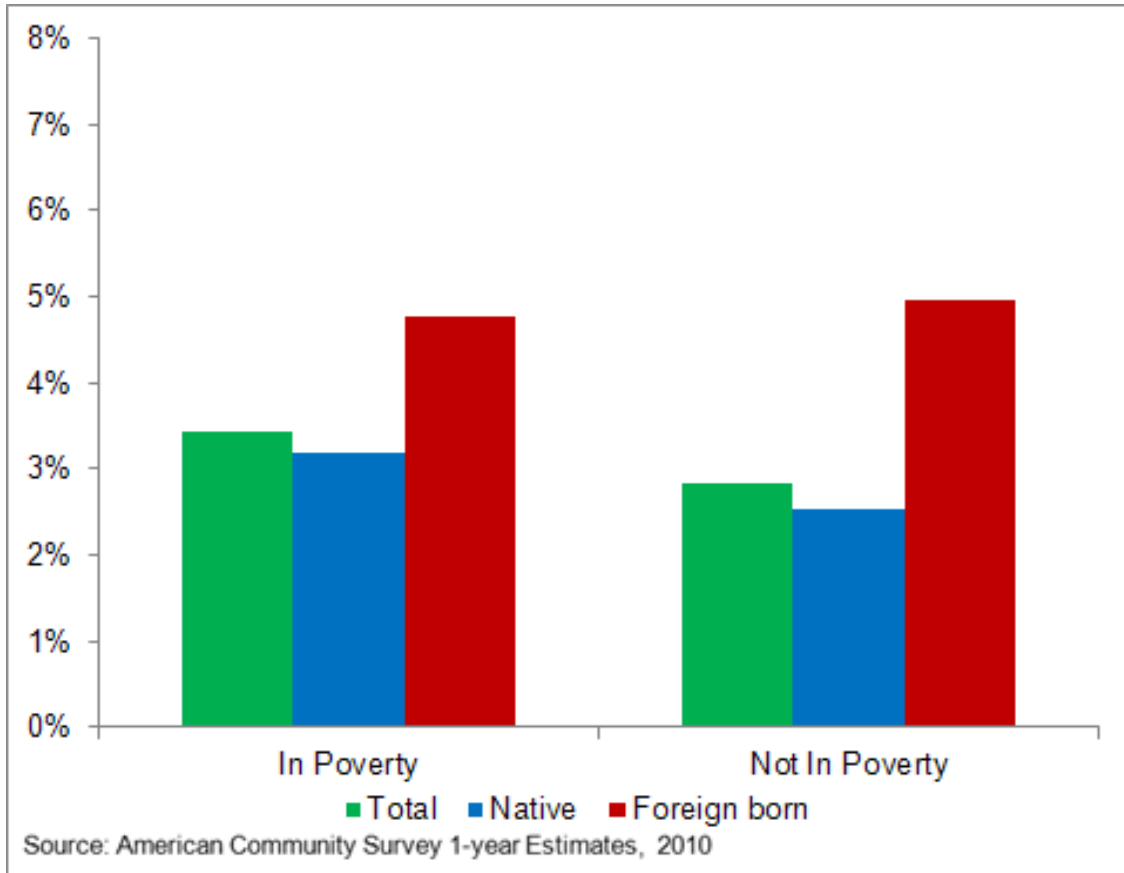
When imputation is analyzed by nativity and educational attainment, in general, individuals with a bachelor's degree or higher were less likely to have their citizenship responses imputed than those with less than a bachelor's degree. This was found for both the native and foreign-born populations.

Figure 6. Percent Imputed Responses to the Citizenship Question by Nativity and English-Language Ability, United States: 2010



Among those who spoke English only, the foreign-born had higher imputation rates than natives. However, among the foreign born who spoke a language other than English at home, imputation rates were all about 4.6%.

Figure 7. Percent Imputed Responses to the Citizenship Question by Nativity and Poverty Status, United States: 2010



When imputation is analyzed by nativity and poverty status, natives in poverty had higher imputation rates than natives who were not in poverty. However, no difference in imputation rates were found for the foreign born by poverty status.

Conclusion

In summary, the quality of responses to the ACS in 2006 to 2010 was relatively high, as indicated by the relatively low rate of imputation of responses.. In general, the native-born population had lower imputation rates on responses to the citizenship

question than the foreign born. There was also little variation by region of birth among the foreign born in imputation for the citizenship question. Imputation was generally higher for the foreign born than the native born regardless of race, Hispanic origin, educational attainment, English-speaking ability, and poverty status. Imputation rates were all about 4.6% for the foreign born regardless of their English-speaking ability. Natives born in poverty had higher imputation rates than the native born not in poverty. The foreign born showed no difference in imputation rates by poverty status. Potential extensions of this research may include exploring causal relationships behind the findings, for example whether correlations remain strong once the influence of other factors such as socioeconomic status is controlled for statistically.

References

Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau, 2004. Survey Methodology. Hoboken, NJ: Wiley Series in Survey Methodology.

U.S. Census Bureau, 2009. "*Design and Methodology: American Community Survey*. (http://www.census.gov/acs/www/methodology/methodology_main/)