

**Assessing the Quality of Data Collected on the Foreign Born:
An Evaluation of the American Community Survey (ACS)**

Pilot and Full Study Findings

August 2003

Steven A. Camarota
Center for Immigration Studies



Center for Immigration Studies

Table of Contents

Table of Contents.....	i
Acknowledgements.....	ii
Executive Summary.....	iii
Introduction.....	1
The American Community Survey.....	3
Evaluation Framework and Key Design Features.....	5
Organization of Report and Presentation of Findings.....	7
Literature Review.....	10
Concerns about Data Collected on the Foreign Born.....	10
Census Literature on the ACS Data Collection Process and Internal Data Quality Checks.....	21
Conclusion.....	29
Focus Groups.....	32
Goals of the Qualitative Analysis.....	32
Methodology.....	34
Findings.....	39
Conclusions.....	63
Internal Validity Checks of the Harris and Fort Bend Counties (Houston, Texas) Sample.....	65
Checking the Internal Validity of the ACS Data: An Overview.....	65
Data and Methods.....	67
Findings.....	68
Conclusion and Implications.....	96
Internal Validity Checks of the C2SS National Sample.....	100
Understanding When Respondents are Captured by the C2SS: A Multinomial Logit Analysis.....	100
Data and Methods.....	102
Results.....	104
Discussion.....	112
Comparing Foreign-Born Estimates from the Harris and Fort Bend Counties (Houston, Texas) Sample to Other Data Sources.....	136
Data Sources.....	136
Findings.....	138
Conclusion.....	151
Comparing Foreign-Born Estimates from the C2SS Sample to Other Data Sources.....	153
Data Sources.....	153
Findings.....	155
Conclusion.....	172
Conclusion and Recommendations.....	173
References.....	179
Appendices.....	182

Acknowledgements

This study, funded by the U.S. Census Bureau under contract 50-YABC-0-66008/0001, subcontract 3701-CIS, would not have been possible without the cooperation and effort of countless Census Bureau employees. The Field Representatives who administer the American Community Survey (ACS) willingly and knowledgeably answered numerous questions and proved to be an excellent source of information on ACS data collection efforts. Kevin Deardorff shared his insights and vision throughout the study, which greatly improved the final product. Angela Brittingham, the study team's project officer and main contact within the Census Bureau, provided substantive and methodological comments while skillfully guiding the project toward completion. In addition, Debbie Klein helped the research team set up site visits to Census Field Offices and Diane Schmidley willingly shared her vast knowledge of the Current Population Survey.

A number of individuals outside of the Census Bureau also contributed to or assisted with this report. The study's advisory panel—Gorge J. Borjas, B. Lindsay Lowell and James G. Gimpel—read and provided pages of comments on both the pilot and full study reports. Mark Kirkorian also read and provided comments on the report and Patrick McHugh was the report's primary editor.

The views expressed here are those of the author and should not be attributed to the Census Bureau or the advisory panel members.

Center for Immigration Studies

Executive Summary

Assessing the Quality of Data Collected on the Foreign Born: An Evaluation of the American Community Survey (ACS)

A Summary of the Pilot Study and Full Evaluation Findings

INTRODUCTION

Since 1996 the U.S. Census Bureau has been piloting a new national data source, called the American Community Survey (ACS), which is intended to replace the decennial census long form in 2010.¹ The ACS contains questions that are nearly identical to those that were on the long form in 1990 and 2000, but unlike the decennial census, is designed to provide information annually rather than by decade.² When fully implemented, the ACS will provide yearly estimates of demographic, housing, social and economic characteristics for the nation, states and other governmental units (i.e., cities and counties) with populations of 65,000 or more.³ Representative data on communities of less than 65,000 will require an initial three or five years to collect (depending on the size of the community), and will then be updated annually. Given that it is implemented each year, the ACS can supply information on the U.S. population in a more timely manner than can the decennial census, allowing policymakers and researchers the opportunity to more quickly identify and respond to social, demographic and economic changes in the population.

This study presents findings from an evaluation of the foreign-born data collected during 1999 in Harris and Fort Bend Counties, Texas, and an evaluation of the foreign-born data from the 2000 ACS national sample, known as the Census 2000 Supplementary Survey (C2SS).⁴ This study was designed to review ACS data collection procedures and to assess the quality of ACS foreign-born data collected in a sampling area with a high concentration of foreign born and in the nationally representative ACS sample.

¹ While the ACS is intended to replace the long form, this change will be dependent upon congressional funding. While full implementation of the ACS was scheduled for 2003, it has been delayed due to cutbacks in survey funding by Congress.

² A number of differences exist between the decennial census and ACS, most notably in how each defines a member of the household. While the decennial census uses a “usual resident” construct to define a member of the household, the ACS employs a “two month” rule. For more information on this and other differences between the ACS and the 2000 Census, see “Changes to the American Community Survey Questions”, www.census.gov/acs/www/Downloads/House.pdf and “Data Collection and Processing”, www.census.gov/acs/www/AdvMeth/CollProc/CollProc.htm.

³ For more information on the methodology of the ACS, see U.S. Census Bureau (2001).

⁴ Throughout the report, we refer to the ACS as either ACS or C2SS where appropriate.

About ACS Data Collection

ACS data collection procedures vary dramatically from other surveys that collect data on the foreign born. The ACS draws a monthly sample from a Master Address File (MAF), which is designed to be an exhaustive list of residential addresses in the United States.⁵ After the sample households are drawn, data collection operates on a series of three-month cycles. During the first month of a cycle, an ACS questionnaire (see Appendix A) is mailed out and data are collected through mail-in self-response. A series of reminders in addition to the questionnaire are mailed to boost response rates, including a pre-notice letter, a post card, and for those households that do not respond to the first questionnaire within three weeks, a replacement questionnaire. These questionnaires are only in English. On the questionnaire, instructions are given in Spanish that direct Spanish-speaking respondents to call a toll-free number to have a phone interview conducted in Spanish.

If a sampled household does not respond to the questionnaire within one month, the Census Bureau attempts to obtain the telephone number of the household using a commercial vendor's list. While it is not always possible to match addresses and telephone numbers, ACS Field Representatives (FRs) spend the second month of the data collection cycle following-up on successfully matched non-responding households using computer-assisted telephone interviewing (CATI). Experienced interviewers collect the data using a sophisticated interview instrument to ensure data accuracy.

If a telephone number is not available for a non-responding household or if CATI interviewers fail to obtain an interview in one month of calls, a one-in-three sub-sample of remaining non-responding households receive a personal visit by a Census FR. These FRs conduct computer-assisted in-person interviewing (CAPI) with an instrument very similar to that used by CATI interviewers. These FRs repeatedly visit non-responding households at various times of the day until an interview is obtained. FRs are very effective in obtaining interviews from these sub-sampled non-respondents.

METHODOLOGY

In light of the complex data collection procedures of the ACS, this project employed a series of different methodological techniques to examine the quality of foreign-born data collected by the ACS, including:

- **A literature review** of studies and governmental reports critiquing the overall scope, quality and utility of foreign-born data from a number of different sources, as well as Census Bureau documents and methodological reports outlining the implementation of the ACS;

⁵ When fully implemented, the survey will draw a monthly sample that will provide an annual sample of three million households.

- **Focus groups** of ACS FRs who administer the survey. Over 50 FRs who conduct in-person and telephone interviews participated in five focus group discussions. Data were obtained from FRs at all three sites where telephone interviews are conducted, and from FRs conducting in-person interviews in Harris and Fort Bend Counties, Texas—an area with a high concentration of foreign born;
- **Internal quality checks** to test the “success rates”⁶ of the modes of ACS data collection, the non-response rates of individual ACS items and the consistency of responses across items in the survey. Methods used to conduct these checks included bi-variate and tri-variate crosstabulations and an unordered multinomial logistic regression analysis; and
- **External Validity or Benchmarking Analyses** where ACS estimates of the size and characteristics of the foreign-born population are compared to other well-established data sources; specifically, the 1990 Census, the 2000 Census and multiple years of the Current Population Survey (CPS).

KEY FINDINGS

Overall, the findings from both the pilot and full evaluations of the ACS indicate that the survey is a highly valuable source of information on the foreign born. In almost all cases, ACS estimates of the size and characteristics of the foreign born are comparable to those derived from the 2000 Census and the CPS. Moreover, C2SS data have certain distinct advantages over both the 2000 Census and CPS. While 2000 Census data are more precise than the C2SS (especially at smaller geographic levels), the C2SS data, even though collected in the same year, were made publicly available months, and in some cases years before the 2000 Census data. In addition, estimates of the size and characteristics of the foreign-born population produced by the ACS and C2SS are significantly more precise than estimates derived from CPS data.

However, the evaluation also revealed sources of non-sampling error in the data collection process, which may affect data quality.⁷ In addition, other issues were uncovered and distinct differences in the quality of the data were found across the modes of ACS data collection. Finally, in some cases, C2SS estimates of the foreign-born population were statistically different from those generated by the 2000 Census; however, the size of these differences was quite modest.

Findings from the Literature Review

Over the last 25 years, numerous reports and journal articles have been written expressing concern over the poor quality or limited utility of data on the foreign born. Five issues have been emphasized as areas of particular concern:

⁶ By “success rates,” we simply mean how successful (in terms of the percentage of respondents) each mode is in capturing data on the foreign born.

⁷ It is important to note that many of the sources of non-sampling error discovered through this evaluation are not unique to the ACS and are found in the data collection efforts of other surveys as well.

- Under-representation or under-coverage of the foreign born by survey sampling;
- Over-representation of citizenship among foreign-born respondents;
- Difficulty recording the correct year of entry into the United States;
- Issues with capturing ancestry; and,
- Biases in capturing language characteristics of the foreign born.

Findings from the Focus Groups

A number of important findings related to the administration of the ACS emerged from the focus groups of Census Field Representatives (FRs). These include:

- **Language barriers are a significant problem.** FRs repeatedly indicated that problems resulting from language barriers represent a significant impediment to collecting high-quality data on the foreign born. The extent to which language barriers posed a problem varied across the modes of data collection and the telephone call centers. Both formal and informal methods are used by Census FRs to overcome language barriers with varying degrees of effectiveness;
- **Census FRs represent a source of non-sampling error.** While Census FRs were found to be knowledgeable and dedicated to the collection of high-quality data on the foreign born, some were unaware of specific coding instructions related to foreign-born questions such as year of entry. Some FRs also did not know whether to code certain categories of the foreign born, such as Lawful Permanent Residents, as citizens or non-citizens. Finally, there was evidence that FRs committed administrative errors or confused question wordings because the ACS survey instrument differed from other surveys on which they also worked;
- **FRs provided anecdotal evidence that coverage error and selection bias may exist.** FRs suggested that particular groups of foreign-born respondents are more reluctant than others to participate in the survey. FRs noted that willingness to participate varied by country of origin and citizenship. Foreign-born respondents from countries where there is a distrust of the government were less likely to participate in the survey than other foreign-born respondents. In addition, FRs believed that foreign-born respondents living in the country illegally were more reluctant than other foreign-born respondents to participate in the survey. FRs believed that in some cases, foreign-born respondents failed to make known all the residents usually living in the household;
- **FRs believed that respondents often incorrectly respond to specific questions on the survey.** FRs noted that they thought response errors to specific questions—in particular those regarding citizenship status and year of entry—occur often among foreign-born respondents. FRs felt that these errors occur because of recall bias, a fear of the implications of certain responses or a desire to answer questions in a socially desirable way;
- **There were also other sources of non-sampling error.** The FRs mentioned a number of other potential sources of non-sampling error less frequently. One such point was that the

length of the ACS survey caused some respondent burden. While this burden is no greater than the Census long form it will replace, the length of the survey did cause some FRs to move quickly through the instrument when interviewing respondents. In addition, FRs thought the survey was poorly marketed. The fact that few respondents had ever heard of the survey made it more difficult for the FRs to build rapport and gain the trust of the respondents.

Findings from the Internal Validity Checks

- **Data from foreign-born respondents are collected later in the ACS data collection process compared to natives.** In both the Harris and Fort Bend Counties sample and in the national sample, foreign-born respondents were nearly twice as likely to have their data collected during the in-person phase of data collection compared to natives. There was also large variation among foreign-born respondents in that some groups (e.g., Mexicans, El Salvadorans and Guatemalans) were more likely to be captured during the in-person phase of data collection than others (e.g., Canadians).
- **Multivariate modeling revealed a number of reasons why the foreign born are captured later in the data collection process.** Results from an unordered multinomial logit model, which predicted the mode in which an individual would respond to the ACS, revealed a number of important predictors. Low income, low educational level, race (e.g., being black, American Indian, or Hispanic), and poor English skills were the largest predictors of data being collected during the in-person phase of data collection. Therefore, one explanation for the foreign born being interviewed later is that they are more likely to possess at least one of these characteristics than natives. Moreover, even when controlling for these and other variables, being foreign born increases the likelihood of responding later in the data collection process. This means that there are other characteristics associated with being foreign-born not identified in the model such as legal status or cultural differences, that increase the likelihood that this population will not respond until the in-person phase.
- **The weights of foreign-born respondents interviewed during the in-person phase of data collection are significantly larger than those returning the mail-in questionnaire or answering by phone.** The weights of foreign-born respondents are not large relative to other surveys. However, the average weight of a foreign-born respondent with data collected during the C2SS in-person phase of data collection is nearly three times as large as foreign-born respondents interviewed in the mail-in and telephone modes of data collection. While this is less of an issue at the national level, the large weights of foreign-born respondents interviewed during the in-person phase may affect the reliability of estimates at the sub-national level and estimates of sub-populations of the foreign born with smaller sample sizes.
- **With the exception of a few items, the foreign-born non-response rates to ACS items were modest.** With the exception of questions associated with income and one or two other questions, the non-response rates of the foreign born to specific ACS questions were low. It is important to note that high non-response rates to income questions are not unique to the ACS or the foreign born. Indeed, other national surveys also have high non-

response rates to these items. In addition, the native non-response rates to the ACS income questions were similar to those of the foreign born.

- **Data collected from in-person interviews have lower item non-response rates than data collected using the other modes.** Interestingly, the non-response rates of foreign-born respondents were substantially and consistently lower among those who responded to the survey through in-person interviews compared to other modes. This finding appears to create an interesting trade-off between larger standard errors caused by sub-sampling non-respondents for the in-person interview and better data quality obtained through this mode. That is, while the weights of foreign-born respondents interviewed during the in-person phase are three times as large as other foreign-born respondents, the data for these respondents require significantly less editing and imputation.

Findings from the Benchmarking Analyses

- **Sub-national (Harris and Fort Bend Counties) ACS estimates of the characteristics of the foreign-born population benchmarked well against the CPS and 1990 Census.** ACS estimates of the proportion and number of the foreign born living in Harris and Fort Bend Counties, as well as their distribution by country and region of origin, year of entry, age, sex, race, marital status, education and employment status were all similar to estimates generated using a three-year combined sample of the CPS (1998-2000) and 1990 Census data. ACS estimates, however, had dramatically smaller confidence intervals around the estimates compared to the three-year sample of the CPS. The increased precision of the ACS estimates points to a distinct strength of the survey in examining the foreign-born population at the sub-national level.
- **At the national level, the ACS estimates of the foreign born benchmarked reasonably well against the 2000 Census and CPS, with some exceptions.** With only a few exceptions, C2SS estimates of the proportion and number of foreign born living in United States, as well as their distribution across a number of different characteristics were similar to estimates generated using 2000 CPS and 2000 Census data.

The benchmarking of the C2SS did, however, reveal some differences with the 2000 Census. Most notably, the percentage of the total foreign-born population nationally in the 2000 Census falls outside of the confidence interval of this estimate for the C2SS. While the difference is not large, it is statistically significant. The smaller C2SS estimate suggests that the 2000 Census more completely captured the foreign-born population than the C2SS, though the difference is relatively modest. It is important to keep in mind, however, that the C2SS was administered at only a fraction of the cost of the decennial census long form, which includes over four million foreign-born cases compared to 130,000 in the C2SS. Moreover, the C2SS was only collected in a little over one-third of the nation's counties. Despite these limitations, the C2SS still produced estimates that were very similar to the decennial census, which should be taken as an indication of the survey's value.

RECOMMENDATIONS

Below, a number of recommendations based on the findings of this evaluation are provided for improving the quality of foreign-born data collected by the ACS. These recommendations are organized into matrix table E.1. by the perceived cost of the recommendation and the difficulty of its implementation.

Relatively Easy and Inexpensive to Implement

- **Change the telephone number for Spanish interviews so that it rings at the Tucson Call Center rather than the Jeffersonville Center. In addition, have a Spanish interviewer receive the call or be sure that the voicemail message is in Spanish.** Through this evaluation, it was found that Spanish-speaking respondents who call the toll-free number on the ACS form to receive assistance with the survey in Spanish call the Jeffersonville Call Center where he or she receives a recording in English. It is recommended that the message be translated into Spanish and the calls be routed to the Tucson Call Center where there are a number of FRs who are certified to administer the survey in Spanish.⁸
- **Enhance FRs training to include a module that deals with issues related to obtaining data on the foreign born.** The focus groups revealed that Census FRs were a potential source of non-sampling error, which can be minimized by providing a modest level of training specifically related to gathering data on the foreign born. For example, FRs should have a basic understanding of the different immigration statuses a foreign-born respondent may hold, and understand that foreign-born non-citizens can be in the country on non-immigrant visas, as lawful permanent residents, and can also reside in the country illegally.

The second area of training should be directed at how the year of entry question is asked. While it is stated in the ACS manual, FRs clearly did not know that they should be asking foreign-born respondents for the most recent time they came to the United States to live.

Finally, according to the focus groups, the FRs who were most successful in obtaining interviews with foreign-born respondents were those who were the most empathetic to the respondent. Therefore, it may be useful to provide sensitivity training to FRs making them aware of the reasons why foreign-born respondents may be reluctant to cooperate with the ACS (e.g., distrust of government, legal status, etc.). The Census Bureau may also want to devise distinct strategies for FRs to deal with foreign-born respondents who are fearful of responding to the survey.

⁸ This recommendation was reported to the Census Bureau in the interim report of this evaluation and has been implemented.

- **Conduct more regular feedback sessions with ACS survey interviewers.** FRs are willing to provide information and should be considered a valuable resource in the Census Bureau's efforts to improve the ACS. As evidenced by this study, much can be learned from feedback sessions with the FRs. Allowing FRs to discuss their interactions with respondents can help to identify problems and potential remedies for sources of non-sampling error. With regard to the foreign born, Spanish-speaking FRs might provide useful suggestions about how to better word the Spanish translation of the ACS so as to reduce confusion.

Perhaps equally important is the impact that obtaining input from FRs can have on morale. When brought together to discuss issues associated with the ACS, FRs feel empowered. If the Census Bureau were to conduct regular meetings with its FRs, perhaps twice a year, with the specific intent of listening to the observations and suggestions they have, it may have a significant beneficial impact on the ACS data collection process.

Relatively Easy but Relatively Expensive to Implement

- **Do more to promote the ACS to the general public, and specifically to the foreign born.** FRs were very clear that awareness of the ACS was virtually non-existent among the public at large, and members of Congress. Data from the focus groups provide some evidence that this lack of public knowledge hurts response rates. Given that there is both strong anecdotal and systematic evidence that the 2000 Census benefited significantly from a well-coordinated advertising and community outreach effort, we recommend that in the years to come, similar efforts, though on a smaller scale, be employed to ensure adequate participation rates for the ACS. While costly, the Census Bureau has extensive experience with the promotion of the decennial census and could use this experience to develop an effective advertising campaign.

Relatively Difficult but Relatively Inexpensive to Implement

- **Allow FRs to specialize in ACS data collection.** Census FRs may work on as many as three different surveys each week. Given that non-sampling error occurred when FRs confused the wording of ACS questions that were similar to those in other surveys, the Census Bureau should consider allowing FRs to specialize in ACS data collection. Given the size of the ACS data collection effort, it should be possible to allow most FRs who work on the ACS to work exclusively on the survey. This may serve to minimize errors made by FRs when collecting data using the ACS.
- **Include a series of toll-free numbers on the ACS questionnaire for respondents speaking languages other than English or Spanish.** There is no single greater impediment to collecting quality data on the foreign born than the problems associated with language barriers, especially among non-Spanish speakers. The evaluation provides compelling evidence to suggest that language barriers cause foreign-born respondents to have their data collected in later stages of the ACS data collection process, if at all. In

addition, there is some evidence that language-barriers also affect the quality of the data that are collected on foreign-born respondents.

Given that the ACS questionnaire is printed only in English, it is unlikely that the paper survey will be returned by households in which no one speaks English. It is also clear that CATI interviewers have serious difficulties communicating with respondents who do not speak either English or Spanish, and frequently fail to obtain data from these respondents. Consequently, foreign-born respondents with poor or no English skills are interviewed disproportionately during the in-person phase of data collection. This may be of concern given that, as discussed, only a one-in-three sample of non-respondents from the mail-in and CATI phases of ACS data collection receive an in-person visit.

One way to help with this issue is to include a series of 800 numbers on the ACS survey for respondents who speak different languages. These languages should at least include Chinese (both Cantonese and Mandarin), Russian, Vietnamese and Korean. The 800 numbers should ring at a call center where Census FRs speaking those languages can administer the interview. If FRs who speak certain languages are not available, the Census Bureau can hire and train personnel to work part-time to respond only to these calls. The Census Bureau can monitor the call volume of the 800 numbers and hire staff accordingly.

Relatively Difficult and Expensive to Implement

- **Create mail-out questionnaires in multiple languages.** In addition to the recommendation above, the Census Bureau may, over time, wish to incorporate other foreign languages into the ACS data collection process. Creating mail-out questionnaires that incorporate the languages above may increase coverage of the foreign born, and may not be as difficult to implement as one would think given that most ACS questions were already translated for the 2000 Census.
- **Cognitively test the ACS questions on foreign-born respondents.** The focus group discussions with FRs suggest that some foreign-born respondents either have difficulty understanding ACS questions or interpret their meaning differently than natives. For example, the FRs noted that foreign-born respondents who speak English well sometimes feel self-conscious about their ability and respond that they speak English “not well.” Foreign-born respondents educated outside the United States also have difficulty answering the educational attainment questions using the U.S. metric offered on the ACS. Examples like these highlight the need for systematic cognitive testing of the ACS questions to better understand how the foreign born comprehend and interpret the questions, and how comprehension and interpretation may vary by different characteristics of the foreign born.
- **Study the ACS reinterview data to test the reliability of responses given by foreign-born respondents.** Census FRs believe that response errors occur among foreign-born respondents, especially for the citizenship and year of entry questions. Currently, the ACS reinterviews approximately four percent of the households that complete the survey

as a means to measure data reliability (i.e., achieving consistent results over multiple trials of the survey). One way to empirically test the beliefs of the FRs would be to analyze the initial and reinterview data of foreign-born respondents to see how they compare. It is important that analyses done comparing the initial ACS data with the reinterview data include a special focus on the foreign born in order to determine the reliability of data from these respondents.

CONCLUSION

The American Community Survey is poised to become the most valuable yearly data source in the country for information on the foreign born. The findings from both the pilot and full evaluations of the ACS indicate that the survey will provide data on the foreign-born population that will be of similar quality to decennial census data, but will be provided on a much more timely basis. While this evaluation points out ways in which the ACS can be improved, the issues presented are found in all large data collection efforts. Indeed, if one were to conduct the same investigation of the Census 2000 data collection procedures, there is no doubt that similar issues would arise.

Figure E.1. Matrix of Recommendations for the Improvement of ACS Data Collection by Ease of Implementation and Expense

		Cost to Census Bureau	
		Relatively Inexpensive	Relatively Expensive
Administrative Burden	Relatively Easy	<ul style="list-style-type: none"> • Enhance FR training to include a module that deals with issues related to obtaining data on the foreign born. • Conduct regular feedback sessions with ACS survey interviewers. 	<ul style="list-style-type: none"> • Do more to promote the ACS to the general public.
	Relatively Difficult	<ul style="list-style-type: none"> • Allow FRs to specialize in ACS data collection. • Include a series of toll-free numbers on the ACS questionnaire for respondents speaking languages other than Spanish. 	<ul style="list-style-type: none"> • Create mail-out questionnaires in multiple languages. • Cognitively test the ACS questions on foreign-born respondents. • Study the ACS reinterview data to test the reliability of responses given by the foreign born.