STUDY SERIES
(Survey Methodology #2007-16)

BEHAVIOR CODING ANALYSIS REPORT
Evaluating the English and the Spanish Version of the Non-Response Follow-Up (NRFU)

Jennifer Hunter Childs
Ashley Landreth
Patricia Goerman
Dawn Norris
Aref Dajani

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

Report Issued: June 14, 2007

BEHAVIOR CODING ANALYSIS REPORT

Evaluating the English and the Spanish Versions of the Non-Response Follow-Up (NRFU) for the 2006 Census Test

Prepared by: **Jennifer Hunter Childs,**
Ashley Landreth,
Patricia Goerman,
Dawn Norris, and
Aref Dajani
U.S. Census Bureau
Statistical Research Division
Center for Survey Methods Research

Submitted: **April 4, 2007**

TABLE OF CONTENTS

Page

Tables

ACKNOWLEDGMENTS

The authors would like to acknowledge those who lent their efforts to this project. We are grateful for the statistical expertise and assistance of Rolando Rodríguez and the research assistance provided by Diana Simmons and Tina Arbogast. Additionally, we thank Theresa DeMaio for comments on an earlier draft of the paper.

Part I INTRODUCTION

As a part of the decennial census, the U. S. Census Bureau conducts a Non-Response Follow-Up (NRFU) operation in an attempt to gather census data from people who did not respond to the mail-out census form. The NRFU operation involves a face-to-face interview with people who live in non-respondent households.

Essentially, the NRFU contains questions from the self-administered decennial paper form (also known as the "short form"). However, the NRFU question wording sometimes differs slightly from the short form to facilitate interviewer administration. The NRFU is a computer-assisted personal interview (CAPI); it is administered using a handheld computer (HHC). An HHC is an electronic device that collects and stores data much like a CAPI survey would with a laptop, but the HHC is much smaller than a laptop. The HHC contains NRFU questions in both English and Spanish, and interviewers can select the language for the interview and even "toggle" back and forth between languages, if necessary, during an interview.

The interview is relatively short, depending on the number of residents in a household, and it begins at the household level (i.e., it asks questions relevant to the entire household). About mid-way through the interview, the survey switches to person-level questions (i.e., it asks questions about each person who lives in the household). The NRFU interviews range from about 5 to 30 minutes in length and vary by household size.

General portions of the interview include questions about the following:

Household-level topics:

- Usual place of residence
- Roster questions
- Other household members who may live there part time
- Housing status or "tenure" (e.g., own with a mortgage, own without a mortgage, rent, etc.)

Person-level topics:

- Relationship to householder
- Sex
- Age/date of birth
- Hispanic origin
- Race
- Ancestry
- Other places where the person might live

Appendix A provides the question text for both the English and the Spanish-language instruments.

2006 Census Test

During each decade the Census Bureau conducts many small-scale "tests" to ensure that the next decennial census will run smoothly. In 2006, one such test was conducted to evaluate the NRFU instrument. The 2006 Census Test was conducted in two different test sites: Austin, Texas and the Cheyenne River Valley, South Dakota. These test sites were selected for the census test in order to include populations that are typically more difficult to enumerate (i.e., recent immigrant and college student populations and an American Indian reservation).

Staff from the Statistical Research Division (SRD) conducted a behavior coding analysis on data collected during the 2006 Census Test NRFU operation. The behavior coding method was selected for its ability to detect interviewer and respondent behaviors indicative of question administration and response problems. The results of this analysis are reported in this paper.

A similar behavior coding evaluation was conducted during the 2004 Census Test NRFU operation and results from that analysis (reported in Hunter and Landreth, 2005) were used to revise the questionnaire for the 2006 Census Test. Three major changes were made between the 2004 and 2006 tests: 1) the 2004 survey was conducted using a person-based instrument, that is, all the questions were asked for the first person in the household before continuing on and asking the same question for the next person, until data for all household members was reported; in contrast, the 2006 questionnaire was conducted using a topic-based instrument in which the first question was asked for all household members, then the second question was asked for all household members, etc.; 2) in 2004 the Hispanic origin and Race questions were collected using a "traditional" two-part question, while in 2006 this series included a question on Ancestry, making it a three-part question; and 3) there was one Relationship question read to respondents in 2004, which was split into two questions in 2006. Other smaller changes were also made between test years.

In addition to the behavior coding evaluation, Development Associates, under contract with the Census Bureau, conducted an observational study of the NRFU operation. Results from that study (Rappaport et al., 2006) speak to non-verbal aspects of the interview interaction, such as flashcard administration behavior. Results from that study will be cited where appropriate in this report.

The remainder of this report discusses the methods used to carry out the research and presents the results from behavior coding. Within the question-level analysis portion of the results section, we have included suggestions for future testing and recommendations for question wording, where appropriate, that may alleviate question administration and response difficulties identified in this research.

Part II METHOD

A. STUDY DESIGN

Our goal was to capture approximately 120 face-to-face NRFU interviews onto audio-tape at the Austin, TX test site. Collection of the audio-taped interviews was intended to achieve that roughly half the interviews taped were in English and half in Spanish, to assure a sufficient sample for analyzing questions of interest for each version of the survey. The sample was not intended to be a representative sample, but rather a sample of convenience.

The audio-taping of face-to-face interviews required special staff to take charge of the recording equipment and consent procedures. Four trained and specially sworn assistants under contract with the Census Bureau accompanied a total of 22 regular field interviewers on their assignments. The assistants were tasked with gaining respondents' consent to record the interviews, and with operating the recording equipment. One taper typically accompanied an interviewer for the day, taping a few interviews conducted by the same interviewer. The assistants tape-recorded the interviews and also completed observation forms on non-verbal aspects of the interview process. Results from those observations are reported in Rappaport et al. (2006).

Interviews were taped throughout the field period of NRFU, in May and June of 2006. Of the 101 audio tapes collected in the field; only a total of 72 audio tapes were usable.[1] Of these, 18 were in Spanish and 54 were in English. The number of Spanish interviews fell far short of our goal of 60 per language, but this was all that the contractor was able to provide.

Five telephone interviewers from the U.S. Census Bureau's Tucson Telephone Center in Arizona were selected to complete a three-day behavior coding training session in August of 2006. Coders were selected based on their fluency in speaking and reading both English and Spanish, and based on their reliability as interviewers as judged by their supervisors. The training was designed and conducted by some of the authors of this report. After training, two interviewers dropped out of the project for medical reasons and three completed the project. Of those who completed the coding, Spanish was the first language for two of the three. Two of the coders were Arizona natives, while the other was from a Mexican town bordering Arizona. The tenure of the coders as interviewers with the Tucson Telephone Center ranged from a year-and-a-half to eight years. Two supervisors participated in the training and oversaw the coding. Both of the supervisors were native Spanish speakers: one was originally from Mexico and the other was originally from Arizona.

To create coding assignments for each coder, the audio-taped interviews were aggregated by interview language and household size. They were then distributed into coding assignments,

[1] Tapes were deemed unusable when they did not contain respondents' consent, were inaudible, were conducted with non-household members, or contained an interview in a language other than English or Spanish.

such that each coder had an approximately equal number of interviews in each language and for each household size.[2]

Behavior coders applied a prescribed framework of behavior codes to interviewer and respondent behaviors by listening to the audio-tapes and following the interview's progress by reading along with a written version of the questionnaire, called a "Question Guide." The Question Guide contained survey questions formatted similarly to how they appeared on the screen of the hand-held computer with both the English and Spanish text available. By comparing the written document to the interviewers' recitation of the questions, coders made assessments about whether and to what extent the interviewers had read the questions exactly as worded. Coders also made assessments regarding whether or not the respondents' answers to the questions could be easily classified (or "codable") into the existing response categories. It should be noted that the coders based their evaluations solely on the information on the audio-tapes; they did not have access to the data generated for each interview so they did not know how interviewers ultimately recorded respondents' answers into the HHCs.

B. LIMITATIONS

Certain aspects of our research design introduce limitations to this study and necessitate some caution in interpreting and understanding the results. First of all, audio recording restricts observable behavior to verbal communication only. We were not able to capture the nonverbal behavior and communication that occur naturally as part of the face-to-face interviewing process. This artifact of the behavior coding method necessitates caution in interpreting the results, because the data—if taken at face-value—can give a false impression of the interaction. For instance, in face-to-face interviews respondents may nod in response to a yes/no question to indicate agreement. This type of inaudible behavior is undetectable on an audio-tape and thus cannot be adequately captured and represented by the behavior codes. Essentially, the respondent's behavior in this situation would be coded as "inaudible" (which is in contrast to an adequate answer), and therefore the number of adequate and codable responses provided by respondents for a given question may be artificially decreased in the analysis. Furthermore, the inability of the audio-tapes to document respondents' nonverbal behavior may affect the coding of interviewer behavior; an interviewer may offer a paraphrased version of the question after receiving nonverbal feedback, such as an inquisitive look, from a respondent. This may happen so seamlessly that it may sound as if interviewers have altered the question the first time it is administered (i.e., first-level interaction or exchange) without provocation from respondents, causing coders to make negative assessments of an interviewer's success in reading the question exactly as worded.

Furthermore, the act of taping an interview may introduce unknown effects into the interview process. For instance, interviewers may be more vigilant in reading questions exactly as worded and administering the survey in the prescribed manner in circumstances when they know their behavior is being recorded and evaluated. Additionally, the mere presence of a taper may have an effect on interviewers' or respondents' behaviors (e.g., respondents may be less likely to inquire about vague terms or complex questions in the presence of two Census Bureau

[2] Each coder coded approximately 28 audio-taped interviews. The coders' caseloads included duplicates of tapes used for reliability purposes.

employees than they would be in a one-on-one interview). As a result, this might cause reported results to underestimate actual rates of problematic behavior in the field.

A total of 72 taped interviews were collected and of these only 18 were conducted in Spanish. Any analysis of the Spanish interviews should be viewed with extreme caution due to the unusually small sample size.

Finally, these results are limited in their generalizability. Results are not necessarily generalizable to whole household proxy or telephone interviews, as these interviews were all non-proxy, in-person interviews. Additionally, the results of the statistical tests performed for this study are intended to be used for heuristic purposes only. The tests were performed as if the data were collected in a simple random sample, with replacement. However, these data were not collected randomly; and therefore, the results are not generalizable.

C. BEHAVIOR CODING

The behavior coding method is used in survey research to analyze the interactions between interviewers and respondents during the administration of survey questions (Cannell, Fowler, and Marquis, 1968). The method involves the systematic application of codes to behaviors (in this case, verbal behavior) that interviewers and respondents display during the question/answer process, and is often used to identify problematic questions (Oksenberg, Cannell, and Kalton, 1991; Sykes and Morton-Williams, 1987).

Behavior coding is a useful method for gathering information about the quality of the survey instrument and the data it collects. If questions and response options are worded and structured in ways that respondents can easily understand and answer, then our confidence grows regarding the ability of the survey instrument to meet its intended measurement objectives. In an ideal interaction between an interviewer and a respondent, the interviewer asks the question exactly as worded and the respondent immediately provides feedback that is easily classified into one of the existing response categories associated with the question. When the interaction deviates from this ideal, however, we begin to suspect there may be problems with the question and/or response options that may be causing comprehension/response difficulties. The application and analysis of behavior codes for these types of interactions allows researchers to pinpoint where such issues are occurring in the survey instrument (Fowler and Cannell, 1996).

A framework of behavior codes is designed to account for and capture instances of ideal and non-ideal interactions, and to indicate particular types of cognitive issues that can occur (Fowler and Cannell, 1996). Codes assigned to interviewer behavior illustrate whether questions were asked as worded; when they are not, this may indicate that questions are awkwardly worded (Fowler and Cannell, 1996) or overly complex. In addition, skipping questions that should be read might indicate that interviewers judge the information to be redundant or the question to be sensitive. Codes assigned to respondent behavior document when feedback from respondents met the measurement objective of the questions, and when responding to a survey question was more complicated. For instance, when terms are unclear, respondents may ask for clarification (Fowler and Cannell, 1996), or when a question is lengthy or complex, respondents may ask interviewers to reread all or a portion of the question. Additionally, refusals to answer questions

may indicate that respondents perceive a request for information to be too sensitive, whereas "don't know" responses may indicate certain types of information are simply unavailable to the respondent.

Behavior coding can be as complex or as simple as the researcher deems necessary. Coding can be implemented at the first level of interaction only (i.e., when an interviewer first asks the question and the respondent provides feedback before the interviewer speaks again) or several interactional levels may be analyzed. Typically, when research intends to identify problem questions, coding the first level of interaction is sufficient because major problems are often evident either when the question is first read or during the initial response from a respondent (Burgess and Paton, 1993; Esposito, Rothgeb, and Campanelli, 1994; Oksenberg et al., 1991; Smiley and Keeley, 1997).

For this project, we coded the "final response outcome" of the interaction as well as the first level of interaction between interviewers and respondents—also sometimes referred to as the first-level exchange. Coding only the first-level of interaction generates insufficient data for demonstrating whether the interviewer and the respondent were ultimately successful in resolving difficulties with the question-and-answer process before moving on to the next survey item. Therefore, we also coded the final outcome to determine whether an acceptable resolution was reached. We used outcome codes to identify whether some type of acceptable or codable answer was negotiated, or whether undesirable respondent behavior persisted as the interviewer exited the question and continued with the interview.

In addition to entering codes to describe interviewer behavior, respondent behavior and final outcome, coders were instructed to take notes on the entire interaction in some cases. This was the case for specific survey questions that we deemed of special interest and any time that non-ideal interactions occurred throughout the interview. These notes were used for later qualitative analysis by the authors and are described throughout the results section.

The framework of behavior codes used for this project was adapted from Oksenberg, Cannell, and Kalton's (1991) research. The codes, and an explanation of their analytical function, are listed in Appendix B. The behavior codes were designed to capture four main aspects of behavior that occur for each question: 1) question-asking behavior for interviewers; 2) response behavior for respondents during the first-level exchange; 3) interruptions by respondents (i.e., "break-ins"); and 4) final response outcome.

D. INTER-CODER RELIABILITY

To assess reliability for the behavior coding results in general, we must determine whether the coders were sufficiently trained to apply the same codes to the same observable behaviors. The bilingual coders independently coded the same 6 interviews, 3 in English and 3 in Spanish, and agreement statistics were generated with the resulting data. For this project, inter-coder reliability was assessed using the kappa statistic. The kappa statistic provides a conservative measure of agreement among coders in their application of the behavior codes, because it accounts for the possibility of agreement by chance (Fleiss, 1981). According to Fleiss, kappa scores greater than

0.75 indicate an excellent level of agreement across coders, while scores ranging from 0.40 to 0.75 indicate a good to fair level of agreement; scores below 0.40 represent poor agreement.

The average kappa score for the three coders was 0.50—a fair level of agreement. This number represents a composite agreement score across languages and accounts for the following variables: interviewer behavior and respondent behavior (i.e., first response behavior; break-in; and final response outcome).[3] Table 1 displays individual kappa scores aggregated by language. Kappa scores for the English-language interviews ranged from 0.34 to 0.61. Kappa scores for the Spanish-language interviews ranged from 0.39 to 0.64. These scores reflect a poor to good level of agreement.[4] The poor agreement level is due to low kappa scores for interviewer behavior in the English-language interviews (0.34) and the first response behavior in both languages (0.40 and 0.39 for English and Spanish, respectively), although scores for the first response behavior do approach a fair level of agreement (i.e., 0.40). The remaining kappa scores reached acceptable levels.[5]

Table 1. Behavior Coders' Kappa Scores by Language

		Respondent Behaviors		
Interview Language	Interviewer Behavior	First Response Behavior	Break-In	Final Response Outcome
English	0.3401	0.3999	0.6141	0.5175
Spanish	0.4933	0.3860	0.6426	0.4764

The low level of agreement for interviewer (English) and first response behavior (both languages) might normally depress our confidence in the reliability of the data for the entire study. A closer inspection of the results, however, increased our confidence in the analysis process and allowed corrective action to be taken with the larger data set. First, the depressed kappa scores were mainly due to a particular coder whose learning curve was steeper than that of

[3] In previous studies, the kappa scores corresponding to break-ins were low, relative to other interviewer and respondent behaviors. In this study, they are highest, relative to the other behaviors. This may be an artifact of how missing data was dealt with in each study. Research is ongoing and results will be forthcoming.

[4] Kappa scores were calculated for the three researchers as well; however, scores could only be computed from the four English-language interviews that they coded since two of the researchers were not fluent in Spanish. The researchers' average kappa score was 0.84, and their individual scores were as follows: interviewer behavior (0.86); first response behavior (0.75); break-in (0.90); and final response outcome (0.79).

[5] For comparison purposes, percent inter-coder agreement (for the three coders) was also calculated for the behavior variables above for which kappa scores were generated. The average agreement for those four variables for the English- and Spanish-language interviews was 70 percent (individual scores were 54 percent agreement for interviewer behavior; 66 percent for first response behavior; 85 percent for break-ins; and 77 percent for final response outcome) and 72 percent (individual scores were 69 percent agreement for interviewer behavior; 59 percent for first response behavior; 90 percent for break-ins; and 70 percent for final response outcome), respectively.

her cohort. The coding case load was organized such that an English-language interview was coded first, possibly contributing to the low score for interviewer behavior in English. While this coder depressed the agreement score early on, the kappa analysis suggests her coding skills increased over time.

Second, we would ordinarily suggest that caution be exercised when interpreting the behavior coding results for these particular variables, but when analyzing the full dataset we noticed that there were a few ways in which some of the interviewers consistently coded respondents' behavior incorrectly. In order to increase our confidence in the analysis, we were able to recode a large number of cases, based on notes associated with the interaction recorded by the coders (see discussion in the following section, E. Quality Check, for an explanation). By taking this corrective action, we are able to more accurately and reliably reflect the response issues present in these questions.

It is interesting to note the kappa scores for the Spanish-language interviews were different from those produced by the English-language interviews. Some kappa scores were higher, while others were lower. This is in contrast to past findings where Spanish-language kappa scores were consistently lower than the English-language counterparts (see Edwards et al., 2004 and Hunter and Landreth, 2005). Future research should be conducted to understand why and under what conditions Spanish-language behavior coding produces lower reliability scores than does English-language behavior coding.

Given the relatively low kappa scores uncovered, we will implement various enhancements in future studies. First, we will increase the number of coders as well as the number of interviews coded for reliability purposes. We originally attempted to train a total of five coders, but two dropped out of the study. Of the remaining coders, two were novices and only one had participated in similar studies in the past. The coder that had the lowest reliability scores demonstrated her relatively long learning curve with her first Kappa interview, an English-interview, which was coded inconsistently with the other coders and brought down the English kappa scores substantially over those that would have been calculated were that interview or that coder not included in the calculations. As language of interview was a key factor in this study, it would have been preferable to have more interviews in each language. In the future, if there are strata of comparison across interviews that comprise key factors, we will need more interviews in each stratum of comparison.

Finally, there were inconsistencies in the rates of missing data between interviewers for the Kappa cases. This had a downward effect on the kappa statistics. Apparently the rules for how to code skipped questions were not applied consistently by the coders. In future studies, we will not permit coders to leave any responses blank. If a question was skipped, then there will be an appropriate code to be entered across all fields for that question. With a larger cadre of coders, more interviews to be coded, and a better process for missing data, we expect to report substantially higher reliability statistics for future research studies.

E. QUALITY CHECK

During an initial review of the dataset by the researchers, it was evident that the coders experienced some difficulty in applying the codes; many times they coded the same type of situation in different ways. This was apparent through a review of the notes that accompanied all less-than-ideal interactions. The researchers recoded variables to more accurately reflect interactions when the notes were both clear and justified the assignment of an alternative code.

For example, "verification" occurs when an interviewer verifies information that the respondent already provided, rather than asking a question exactly as worded (see Appendix B for a more detailed explanation of each of the codes). So an interviewer might ask, "You said you were Hispanic, right?" Verifications were classified as either correct or incorrect depending upon whether the coder had actually heard the respondent providing this information at an earlier time in the interview. In some cases the researchers noticed that a coder had classified a question such as "You're Hispanic, right?" as a major change to the question wording instead of as a verification. When the notes clearly indicated that this was the case, the variable was recoded.

Coders also had difficulty determining when a response met the measurement objective of the question. For instance, when a question was asked with a major change to question wording, yet produced an adequate answer, coders sometimes miscoded the acceptable response as an inadequate answer. This may have been because the coder thought that the respondent was answering a different question than intended, so technically the response was invalid. An example that easily demonstrates this phenomenon occurred frequently for the Race question. Rather than reading the question as worded, an interviewer might ask only "Are you white?" The respondent might say "yes" in response. The interviewer's behavior should have been coded as a major change to question wording and the respondent's behavior should have been coded as an adequate answer. Although the respondent did not hear the whole question or all of the response options, a response of "white" does meet the measurement objective of the question and is easily codable as one of the response categories in the survey instrument.

The recoded data produced a more conservative picture of the errors for both interviewer and respondent behavior. Overall, through recoding, the percentage of questions classified as "read exactly as worded" increased, while major change percentages decreased. The number of both positive and negative verifications also increased. For respondent behavior, adequate answer percentages generally increased while inadequate answer percentages decreased.

F. LOGISTIC REGRESSION USING INTERVIEW LANGUAGE AND REPEATED QUESTION ADMINISTRATION

Similar to the 2004 analysis (see Hunter and Landreth, 2005), we hypothesized that two variables would affect interviewer and respondent behavior. The first was interview language (Spanish or English) and the second was "repeated question administration" (i.e., whether a question was being asked for the first time or being repeated for additional members of a household).

We examined the results by language to see if there were problems concentrated in one language version more than the other. Behavior coding is a useful method both for evaluating the

translation of a survey instrument and for identifying questions and concepts that may differ in interpretation across languages or cultures.

The "repeated question administration" variable reflects the fact that the NRFU instrument asks the same questions about multiple household members. We were interested in whether we would find more undesirable interviewer behavior when interviewers were required to repeat survey questions for people in the household after "Person 1" (i.e., Person 2, Person 3, and so on). Once an interviewer has asked the question one time and must repeat it for later household members, we hypothesized that he or she might shorten or rephrase the question to speed up the interview or to avoid sounding redundant. Additionally, we looked at what effect, if any, this behavior has on respondent behavior.

We ran a logistic regression examining the effect of interview language (i.e., English and Spanish) and repeated administration (i.e., the first administration of the question versus all other administrations of the same question) on "good" behavior by the interviewer and the respondent for each question.[6] We looked at the effects of language and repeated administration on interviewer behavior, respondent first-level exchange behavior, and respondent outcome behavior.

[6] Good Interviewer Behavior was defined as exact wording/slight change (Code E/S) and correct verification (Code V+). The correctly skipped category (Code CS) was omitted from the analysis because it is neither "good" nor "bad" behavior. The only behavior that was considered good respondent behavior was an adequate answer (Code AA). Inaudible respondent and outcome behaviors (Code I/O) were removed from this analysis because they are neither "good" nor "bad" behaviors.

Part III RESULTS

A. INTRODUCTION

To present the results of the behavior coding, first we created tables containing rates of behavior types by question (Tables A1 and A2 in Appendix A) and describe how to interpret them. We present general findings, looking at overall assessments of the performance of these questions as well as general results from the logistic regression analysis. Finally, we present detailed question-by-question findings and recommendations for each substantive question in the NRFU instrument.

Behavior Coding Results Tables

The aggregate results of the behavior coding for interviewer and respondent behaviors are contained in Tables 2a and 2b. Table 2a contains interviewer behavior parsed across seven types of behaviors per question. The information in this table accounts for approximately 100 percent of interviewers' behavior (taking into account rounding error), and includes:

Exact or slight changes to question wording (E/S);
Major change (MC);
Correct verification (V+);
Incorrect verification (V-);
Inaudible or other (I/O);
Incorrect skip (IS); and
Correct skip (CS).

In addition, Table 2a presents respondent behavior at the first-level exchange parsed across seven types of behavior per question and includes the following variables:

Adequate answer (AA);
Inadequate answer (IA);
Qualified or Uncertain answer (QA);
Clarification or Request to re-read question (CL);
Don't know (DK);
Refused (R); and
Inaudible or other (I/O).

The percent of respondent interruptions (i.e., "break-ins") to the initial question administration is also provided in Table 2a. These calculations were based on the total number of first-level exchange respondent behaviors for each question. Break-ins are calculated separately from the seven respondent behaviors mentioned above because when a respondent breaks-in, he or she does so by saying something that can be coded (e.g., the respondent could break in with an answer that may be codable or uncodable or they may interrupt for clarification).

Table 2b contains percentages for the respondents' final response outcome and contains the same types of behavior included for the first-level respondent behavior, excluding requests for clarification and/or re-reading of the question (because we expect that no interaction would end on a request for clarification or re-read).

These tables represent 72 households (containing a little over 200 people in total) interviewed for the NRFU survey. Questions 1 through 6 are household-level questions and are asked only once of the respondent. The remaining questions are person-level questions. Data for these questions are gathered from the respondent for every member of the household with the exception of Questions 7 and 8 (Relationship), which are collected about Person 2 and higher. This accounts for the large increase in the number of observations for Questions 7 and above.

Interpreting Behavior Coding Results

In analyzing behavior coding data, the standard practice for identifying flawed survey questions is to flag questions for which non-ideal interviewer and respondent behaviors exceed 15 percent for any behavior type (e.g., major change or inadequate answer). Though this is an arbitrary cut-off point, this level of non-ideal behavior suggests that a question has a "high level" of a problem that merits some attention (Oksenberg, et. al, 1991; Fowler, 1992).

B. GENERAL OBSERVATIONS

According to Oksenberg, et al.'s standard for interpreting behavior coding data, the interviewer behavior results in this study are quite striking; all but two questions (i.e., Relationship 1 and Age) exceeded the 15 percent threshold for major changes made by interviewers to the question wording (see Table 2a). Interviewers altered wording to the extent that question meaning could have been interpreted differently than intended between 8 percent and 64 percent of the time, depending on the question of interest. In fact, on average, ideal question-asking behavior across all of the questions that we analyzed was only 40 percent (aggregate data not shown). This demonstrates that for most of the NRFU instrument, the interviewers had a difficult time achieving standardized question reading at satisfactory levels.

In terms of respondent behavior, the results indicate that respondents were able to provide adequate answers to most questions; however, there was evidence of problems with some questions. Six questions generated a high level of initial responses that failed to meet the measurement objective of the questions. The percent of inadequate answers received on the first exchange was at unacceptable levels in the questions on Usual Residence, Ownership, Relationship (both 1 and 2), Hispanic Origin, and Race. When looking at the final response outcome (Table 2b) these issues decrease below the problematic rate for Usual Residence, Ownership, and Hispanic Origin, but an unacceptable level of inadequate answers persists for Relationship 1 and 2, as well as Race. The persistent problem in the Relationship and Race questions, as evidenced by the high percent of inadequate final outcome responses, indicates that the concepts may be poorly communicated and/or respondents may not be able to respond within the parameters of the response set provided. In either case, these results indicate that a significant number of these respondents were unable to negotiate the Relationship and Race questions to the satisfaction of the questions' measurement goals, based on the answers provided to the

interviewers.[7] Particular issues related to each question are discussed in greater detail in the question-level presentation of results.

There were a few additional questions for which unacceptable or non-ideal respondent behaviors approached the 15 percent threshold. There were higher than expected rates of Qualified or Uncertain Answers given for the questions on Undercount and Age. Additionally, there was a higher than expected rate of "Don't Know" responses for the question on Age. These issues are also discussed in greater detail in the question-level presentation.

Logistic Regression Results

The findings from the logistic regression analysis are included in Tables 3 and 4 (for Repeated Question Administration and Language, respectively) at the end of this report.[8]

Repeated Question Administrations[9]

The trend for every one of the person-level questions we analyzed was that it was asked as intended (i.e., exactly as worded, with slight changes or correctly verified) more often the first time it was administered than for repeated administrations. This means that it was asked appropriately more often for Person 1 than for Persons 2 and higher in the household (see Table 3). Collapsing across questions, this difference in behavior was significant. For the first administration, questions were asked as intended 50 percent of the time; for subsequent administrations, correct question-asking behavior decreased to 32 percent (see Table 3).

The logistic regression analysis showed significant negative effects of repeated question administrations on good interviewer behavior for the questions on Date of birth, Confirming age as of Census Day, and Hispanic origin (see Table 3). Question-level analyses of these differences will be presented in the sections pertaining to each question.

There were no significant effects of repeated administration on respondent behavior or outcome behavior. This means that, despite the fact that interviewers often shortened the question for later people in the household, there was not a significantly increased likelihood that the respondent would provide an inadequate response to the question for later people in the household.

Between 2004 and 2006, the structure of the person-level questions was changed from a person-based format, where the entire series of questions was asked about each person before moving onto the next person, to a topic-based format, where the individual question was asked about each person before moving onto the next question.

Although the 2006 instrument used the topic-based format, the interviewer was required to read the same complete question for each household member. So, for example, the interviewer was

[7] Though it was not in the scope of this study, it would be interesting to compare the ways in which interviewers field coded/recorded these responses into the HHC with the data captured on audiotape for this study.
[8] The first model we ran included an interaction term between question administration and language. It was not hypothesized to be significant, and it was shown not to be (nor was it shown to be a significant predictor of effects at the question level). Therefore, it was omitted from the model for the analysis presented here.
[9] Results reported for repeated question administration are holding the effect of language of the interview constant.

required to read: “Are you of Hispanic, Latino or Spanish origin?” followed by “Is Mary of Hispanic, Latino or Spanish origin?” followed by, “Is John of Hispanic, Latino or Spanish origin?” A more conversational implementation of the topic-based approach would employ using a shortened version of the question after initially asking it in full. So the interviewer would read, “Are you of Hispanic, Latino or Spanish origin?” followed by “How about Mary?” followed by “How about John?” Our results indicate that despite the requirement to read the question in exactly the same way for each household member, many interviewers took it upon themselves to shorten questions for later household members to the “How about NAME?” format. This accounted for many of the interviewer behaviors classified as a major change (MC) to question wording.

This finding indicates that our problems with non-standardized question administration were not solved by switching to the topic-based approach alone. However, the topic-based approach in combination with a simplification of questions on later household members to “How about NAME?” would likely improve standardized question administration rates. Programming this type of wording into the instrument from the outset would most likely encourage interviewers to follow the script, therefore deviating less from the exact question wording. In this way, we might be able to control the ways that interviewers shorten question so that we can maintain a standardized approach.

Despite a similarly low level of ideal interviewer behavior with that found using the 2004 instrument, the topic-based approach makes more sense from a cognitive processing perspective (i.e., respondents do not have to reprocess the same question anew for each household member). If shortened “how about” questions had been used in 2006, we believe the good interviewer behavior rates would have shown significant increases for many questions. We believe the poor interviewer results are an artifact of a non-conversational question-administration script.

Interview Language[10]

There were significant effects of language on overall interviewer, respondent, and outcome behaviors. Questions in English were more often administered correctly than were those in Spanish. This trend was evident for each of the 16 survey questions (see Table 4). In English interviews, questions were asked in a “good” way 46 percent of the time, while they were asked in a “good” way only 31 percent of the time in Spanish interviews. For respondent behavior, English questions yielded a rate of good response behavior 82 percent of the time, while Spanish questions yielded a rate of good response behavior only 69 percent of the time. Outcome behavior was similar with a good response outcome in 89 percent of English interactions but in only 79 percent of Spanish interactions.

Logistic regression analysis revealed no significant effects of language on good interviewer behavior for any specific question (though this is likely due to lack of power with our small Spanish sample). We did find significant effects of language on good respondent and outcome behavior for the questions on Relationship1 (the first of two questions about relationship), Hispanic Origin, and Race. Analyses for these questions are presented in the question-level section.

[10] Results reported for language are holding the effects of repeated question administration constant.

The fact that Spanish cases exhibited poorer interviewer and respondent behavior and outcomes may be explained by a number of factors. First of all, the fact that the Spanish instrument is a translation (and not an instrument developed in Spanish) may cause it to sound less natural or conversational than the English version in some instances. In addition, not all of the terms and questions in the Spanish instrument have been properly pretested to be sure that respondents will comprehend them as intended. This may lead interviewers to contextualize or alter question wording in places where they have found that questions do not tend to "work" well with respondents.

There are different norms of politeness across cultures and it may not always seem appropriate to interviewers to launch into the scripted interview without making some small talk or framing questions in some way. Similarly, due to cultural conversational norms or difficulty with the translation, Spanish-speaking respondents might feel at times that a discussion is warranted, rather than just giving a brief response. It is not surprising that this is particularly the case on the Hispanic Origin and Race questions, which that have been shown to be particularly difficult for both English- and Spanish-speaking Hispanic respondents to answer in cognitive testing (Caspar, et al., 2006, Caspar, et al., forthcoming, Goerman, forthcoming). In addition, the relationship question may be difficult for immigrant respondents, who are more likely to live in complex living arrangements with people other than just nuclear family members.

Another issue is that the Spanish-speaking interviewers employed for the Census test were not tested or certified as to their Spanish-language proficiency levels. In listening to some of the tapes, we observed that some Spanish-speaking interviewers had difficulty reading Spanish aloud and had problems with Spanish pronunciation and grammar. To complicate the situation even more, immigrants of limited English proficiency often have lower educational levels than the average population in the U.S., and this may contribute to the need for greater discussion in answering the questions in Spanish.

Finally, we want to note here that there were many typographical errors, translation mistakes and fill errors in the Spanish version of the 2006 NRFU instrument. These errors are enumerated in Appendix C. These problems may be due to a lack of thorough review/proofreading of the Spanish translation (possibly during the programming phase). In most cases they will cause interviewers not to be able to read questions exactly as worded. In some cases there is a strong likelihood that English and Spanish speakers will interpret and answer the questions differently because of these errors, which will most certainly impact data quality. Because we see these things as errors, we will not expand on them in the text, but it is important that they be corrected and not exist in future instruments.

General Recommendations

There are some findings that lend themselves to general cross-cutting recommendations. These recommendations are outlined here. They will be mentioned again in the question-by-question analysis where they apply.

- Several previous studies have demonstrated the need to use a series of shorter questions to convey residence rules instead of the current method of a long question with a flashcard visual aid (Childs et al, 2006; Childs et al, 2007; Hunter, 2005; Hunter and

Landreth, 2005; Rappaport et al., 2006; U.S. Census Bureau, 2006). Findings from the current study support that recommendation. Childs et al. (2007) presents cognitive testing of a questionnaire that fits this recommendation.

- Generally, we found that interviewers often omit the date (April 1, 2006) from questions. We recommend careful review of the 2008 NRFU specification to ensure that the date is not repeated extensively as it was in the 2006 instrument. Once the date has been established for the interview, it should not be repeated in full over and over. Consider also replacing the year with "this year" after it has been mentioned once. Repeating the year can cause respondents to misunderstand that the interviewer is indeed talking about THIS year.
- As has been recommended in past research (Hunter, 2005; Hunter and Landreth, 2005), we recommend automatically filling "house," "apartment," or "mobile home" based on interviewer observation early in the interview to avoid this awkward fill. Additionally, we recommend including (esta/este) in the Spanish fill for completeness.
- Take full advantage of the topic-based format by scripting shortened versions of the questions for later persons in the household (e.g., "How about John?) when appropriate. By scripting the shortened version when appropriate (and by always having the full text as optional text on the screen), the interviewers should learn when it is, and when it is not appropriate to shorten the question. This is key for the relationship question which cannot be shortened or it loses its meaning (see the question-by-question analysis on Relationship for further discussion).

C. QUESTION-LEVEL ANALYSIS AND RECOMMENDATIONS

In this section, we provide results from a question-by-question analysis of each question a respondent should hear in a non-proxy personal visit NRFU interview. When appropriate, we provide recommendations for implementation of changes for 2008 or for future testing.

Live Here Question

This question is aimed at determining whether or not the respondent or anyone in his/her household lived at the address in question on April 1, 2006. This question determines the path that the interview will follow in the instrument. It is intended to be read as:

> **Question 1. Live Here**
>
> **Did you or anyone in your household live at [ADDRESS] on April 1, 2006?**
>
> **¿Vivía usted u otro miembro de su hogar en [ADDRESS] el 1 de abril de 2006?**

This question suffered from major change 40 percent of the time (see Table 2a). Analysis of behavior coding notes revealed that these major changes consisted of omission of the date in the question (either the whole date, or just the year). Additionally, interviewers sometimes replaced the address fill with "here." These two issues typically did not co-occur. In a small number of Spanish cases and in one English case, the interviewer read the question as worded and then

added an explanation of why it was being asked. The same types of changes occurred in both English and Spanish cases.

Respondents usually gave adequate answers in both of these situations. However, we cannot determine if their answers would have been the same had they heard the question read exactly as scripted. For example, if the date was omitted, despite the fact that the respondent answered with a codable "yes" or "no," he or she could have been answering regarding any point in time, not necessarily April 1, 2006.

Respondents occasionally provided inadequate answers by giving detail on exactly when they lived there, or who lived there, instead of answering "yes" or "no" (e.g., "I've lived here for 7 years;" "Only me"). This occurred in small numbers of both Spanish and English interviews.

In a few cases, we inadvertently gathered and coded proxy interview tapes. In those cases, the question asked if the address was "vacant or occupied," leading the respondent to answer "occupied" or "people are living here" instead of "yes/no."[11] Therefore, the major change and inadequate answer rates recorded by our coders were somewhat inflated for this question.

Recommendations for Live Here

Interviewer training should emphasize the importance of reading the question exactly as worded. Interviewers might be more likely to read this question as worded if we were to establish the reference date early in the interview, and instruct them to read only an abbreviated version (without the year) at this point and possibly during the remainder of the interview.

Usual Residence Question

The usual residence question is aimed at determining whether the housing unit is the usual residence of the residents or whether it is a seasonal unit. It is intended to be read as:

Question 2. Usual Residence

Is this [house/apartment/mobile home] a vacation home, seasonal residence, held for occasional use, or does someone in this household usually live here?

[] Vacation home, seasonal residence, held for occasional use
[] Usual Residence

¿Es [esta/este] [casa/apartamento/casa móvil] una vivienda para vacaciones, una vivienda [estacional/de temporada] o una vivienda que se mantiene para uso ocasional, o vive aquí usualmente algún miembro de este hogar?

[] Una vivienda para vacaciones, una vivienda estacional/de temporada, o una vivienda que se mantiene para uso ocasional
[] Residencia habitual

[11] These cases were not removed from analysis due to our already small sample size and because questions after this one had identical question wording.

Changes recommended from the 2004 Census Test (Hunter and Landreth, 2005) included changing "usual residence of someone in your household" in the question text to "someone in this household usually lives here" and putting the most commonly selected option - "someone usually lives here" - at the end of the question to take advantage of recency effects in aurally administered questions.

Interviewer Behavior

This question suffered from major change 42 percent of the time (see Table 2a). Interviewers often asked whether the abode was a "usual residence" instead of asking whether someone "usually live[s] here." This type of change occurred in both English and Spanish interviews. This is most likely due to the fact that "Usual residence" was a response option shown to the interviewer on the handheld computer screen (but not meant to be read to the respondent). This phrase was listed in the instrument even though it should have been changed to match the question text. We consider this an instrument error, and one that likely affected question-administration behavior negatively.

When interviewers read the question correctly, they would at times emphasize or repeat the "usually live here" option at the end of the actual question, thus creating a leading question (or in some cases aiding the respondent in choosing the appropriate answer). Neither of these behaviors appeared to produce inadequate respondent behavior. However, we do not know if respondents would have answered differently had the interviewer read the question correctly.

Interviewers also occasionally left out one or more response options or an important word/concept in the question (e.g., "someone usually lives here," "held for occasional use," "household"). At times these behaviors initially produced inadequate answers, but further interviewer/respondent negotiation resulted in adequate answers for most cases in the final response outcome. Again, these are codable answers but they might not be the same answers the respondents would have given had the question been administered correctly.

Respondent Behavior

Respondents provided inadequate responses 21 percent of the time at the first response (see Table 2a) and the answer remained inadequate in the final outcome 13 percent of the time (see Table 2b). Inadequate respondent behavior consisted of two types. First of all, respondents often focused on the word "residence" and used it (and sometimes alone) to answer the question (giving responses such as "for now, a residence" or "residential"), even though this term is associated with a response option (i.e., "seasonal residence") that did not reflect their intended response. Secondly, respondents often provided details on *how long* they had lived in the residence or *who* lived there instead of answering the question (e.g., "We live here.", "I live here since 1975", "My husband since 1990," or "Just me"). This occurred even when the interviewer read the question as worded.

In Spanish interviews, 4 respondents answered by saying only that their home was rented, perhaps being uncertain as to whether a rented home should be counted as a "usual residence" or simply being uncertain as to the intent of the question. This also occurred in one English-language case.

Response difficulties may have arisen because the first few response options set up an expectation that the question seeks to determine *housing type* ("vacation home, seasonal residence"), but the next two categories describe—with less precision—other kinds of living/housing arrangements that emphasize amounts of time respondents spend at a particular dwelling ("held for occasional use, or does someone in this household usually live here?") and do not describe housing *types*. This inconsistency in the formation of the response options could leave some respondents searching for the "correct" term to provide the interviewer (e.g., "residence"), or worse yet, misunderstanding the question intent altogether (e.g., respondents providing their tenure at a dwelling).

Spanish-Specific Problems

The way the question is currently programmed in the instrument, the interviewer sees a great deal of text that he/she does not have to read in the Spanish version: "(esta /este) [casa/apartamento/casa móvil]" and has to choose one term out of both sets of brackets him/herself. If the instrument were changed to display only one of the [house/apartment/mobile home] terms based on interviewer observation, the translation for "this" could be a part of that fill so that the interviewer would see and read either the masculine or feminine form of "this" depending on whether the instrument displays the fill for a house, apartment or mobile home.

Two different translations for the term "seasonal residence" are listed in the Spanish translation. The translation reads: "una vivienda estacional/de temporada" meaning literally "a seasonal/seasonal residence." The translator may have supplied two terms because "estacional" is a literal translation for "seasonal" but is not a common term in Spanish. In fact, cognitive testing of the Spanish contained in the 2005 Test version of the (paper) Census Bilingual Questionnaire found that large numbers of Spanish-speaking respondents did not recognize or understand the phrase "residencia estacional" as intended (Caspar, et. al., 2006). The biggest concern was that some respondents interpreted it to mean "stationary" or permanent residence, while others expressed confusion and said that they had no idea what it meant. Some respondents were able to guess at the meaning but they reported that this was not a term that they had heard before. The confusion cut across educational levels and nationalities.

If one of these two terms must be used we recommend "de temporada." This is a more common term in Spanish and would probably convey closer to the intended meaning of "seasonal," although further testing is probably in order. Another option would be to test the term "temporary" or "temporal" in Spanish to see if this conveys closer to the intended meaning of "seasonal residence."

As the question is written in the 2006 NRFU instrument, an enumerator cannot read it exactly as worded and sound natural. He or she would have to pick one term or add "or" between the two terms. This type of situation does not encourage enumerators to read as worded.

Recommendations for Usual Residence

- Automatically fill house, apartment, or mobile home based on interviewer observation early in the interview to avoid this awkward fill. In Spanish, make this fill include (esta/este).

- Make the response options match the question text. By changing the response options, the tendency for the interviewer to change "usually live here" to "usual residence" may be reduced. This, in turn, may reduce the tendency of respondents to focus on the word "residence" in their answers and may increase the likelihood of an appropriate answer from the outset.
- We have two alternatives to suggest for future testing to improve comprehension of this question:
    1. Since the acceptable responses to this question fall into only two categories (i.e., home is used occasionally or someone usually lives there), the question could be simplified. Confusion over question intent could be avoided by simplifying the response options to avoid mixed metrics in the response options. A question like this could be tested for future use:

       Is this [H/A/MH] used occasionally (as a vacation or seasonal home), or does someone in this household usually live here?
       [] Used occasionally (vacation/seasonal home)
       [] Usual residence

       ¿Se utiliza [esta/este] [casa/apartamento/casa móvil] solo ocasionalmente (como vivienda para las vacaciones o de temporada), o vive aquí usualmente algún miembro de este hogar?
       [] Ocasionalmente (vivienda para las vacaciones/de temporada)
       [] Residencia habitual

    2. Alternatively, we could test putting the most likely response option ("someone usually lives here") first. By shortening the question, respondents should be able to remember each of the remaining two options without any difficulty. Additionally, we could test omitting the phrase "held for occasional use" from the question because respondents in cognitive testing did not understand this term (Hunter, 2005). This option could remain as a "blind" response category in case the respondent offers something to that effect.

       Does someone in this household usually live at this <house/apartment/mobile home> or is it a vacation or seasonal residence?
       [] Someone usually lives here
       [] Vacation or Seasonal residence
       [] Held for occasional use

       Usualmente, ¿vive alguien de este hogar en [esta/este] [casa/apartamento/casa móvil], o es [esta/este] [casa/apartamento/casa móvil] una vivienda para las vacaciones o de temporada?

       [] Alguien del hogar vive aquí usualmente

[] Es una vivienda para vacaciones o una vivienda de temporada
[] Es una vivienda que se mantiene para uso ocasional

Household (POP) Count

The question used to gather the Household Count aims at conveying the residence rules through a two-bullet summary of the main concept of usual residence. It also includes showing the respondent a residence rules flashcard, which delineates more detailed instructions about who to include and exclude. There is no way of telling from the audio-tapes whether or not the flashcard was administered, but we do have observation summaries that report this information. Rappaport et al. (2006) report that about 25 percent of respondents in this study were shown the flashcard. It is interesting to note that Rappaport and colleagues report the flashcard was shown to the respondent in 28 percent of English cases, but in only 17 percent of Spanish cases.

This question is scripted as follows:

Question 3. POP Count

The census must count every person living in the United States on April 1, 2006.

- **We want to count people where they usually live and sleep.**
- **For people with more than one place to live, this is the place where they sleep most of the time.**

How many people were living or staying in this [house/apartment/mobile home] on April 1, 2006?

El censo tiene que contar a todas las personas que viven en los Estados Unidos el 1 de abril del 2006.

- **Queremos contar a las personas en el lugar donde viven y duermen habitualmente.**
- **Para las personas que tienen más de un lugar para vivir, éste sería el lugar donde duermen la mayor parte del tiempo.**

¿Cuántas personas estaban viviendo o quedándose en [esta/este] [casa/apartamento/casa móvil] el 1 de abril de 2006?

Interviewer Behavior

The POP count question was administered exactly as scripted in only 52 percent of cases (see Table 2a). In 45 percent of cases the question was reworded, which may have led to it being interpreted differently by respondents. In about a third of the major change cases, the interviewer omitted one or more of the four sentences contained in this question. This provides a strong indication that the question is too long for interviewers to read as worded consistently. In about half of the major change cases, the interviewer omitted part or all of the date in one or both places where it appears in the question text. It maybe too burdensome to ask interviewers to read the reference date twice in the same question. The fact that some interviewers are omitting the reference date in both instances where it occurs in this question may indicate that interviewers feel that the reference date has already been established prior to this point in the interview.

In about a third of the major change cases, the interviewer omitted the word "staying" from the question itself. This changes the meaning of the question by potentially making it more restrictive–implying that we are only interested in the people who "officially" live in the household, and not the people staying there temporarily. Because we know we miss people who have no connection to a single housing unit, reading this question this way could lead to an undercount. Again, this may be an indication that, more generally, the question is too complicated and that interviewers are looking for any possible way to shorten it.

Sometimes, though less frequently, the interviewer incorrectly filled the house/apartment/mobile home fill by saying "here." This gives an indication that the fill that appears on the screen is either causing confusion or is perceived as too long or burdensome. This indicates that having the interviewer pick a fill on the fly may be causing them undue burden.

The types of changes made by interviewers were similar across Spanish and English-language cases. In a few Spanish-language cases, interviewers read the entire question as worded but then immediately rephrased it in a simpler, shorter way before the respondent answered. In some cases this may have been due to a pause or confused look on the part of the respondent (which would not have been recorded in our data).

Respondent Behavior

Despite poor interviewer behavior, respondents did provide adequate (codable) responses in 94 percent of cases. The fact that respondents gave adequate answers to this question more often than to any other question in the survey should not be interpreted to mean that the data quality is best for this question. Rather, while we know that respondents provided codable answers to the question, we also know that they were often hearing different versions of the question across cases. We have no way of knowing whether a respondent's answer accurately reflects his or her living situation.

Recommendations for POP Count

- Use a series of shorter questions to convey residence rules. This would be easy to achieve in an interviewer-administered instrument. See Childs et al. (2007) for an example of how this might be accomplished.
- Review the screens at which the date appears and avoid repeating the reference date to the extent possible.
- Automatically fill house, apartment, or mobile home based on interviewer observation early in the interview to avoid this awkward fill. Include (esta/este) in the Spanish fill.

Gathering the Roster (Names)

The question used to gather the household roster has two purposes. The first is to have the respondent list one of the owners or renters of the household as Person 1. The second is to gather the names of all of the people living or staying in the housing unit. Having the respondent choose Person 1 correctly is crucial because household relationships are later recorded through each

person's relationship to Person 1. In census data analysis, family and non-family households are classified through responses to the relationship question. For this reason, it is critical that Person 1 be one of the owners or renters of the household (as opposed to one of the children or a boarder). The question was scripted as follows:

> **Question 4. Names**
>
> **What is the name of each person who lived or stayed at this [house/apartment/mobile home] on April 1, 2006? Start with the name of one person who owned or rented this [house/apartment/mobile home].**
>
> **¿Cómo se llama cada una de las personas que vivía o se quedaba en esta [casa/apartamento/casa móvil] el 1 de abril de 2006? Comience por el nombre de la persona que era [el/la] [propietario/propietaria] o quien alquilaba esta [casa/apartamento/casa móvil].**

Interviewer Behavior

This question was administered with a major change in 64 percent of cases (see Table 2a). Of the major change cases (N=46), a third of the time behavior coders reported that the interviewer dropped the last part of the question, which asks the respondent to start with the person who owns or rents the home (data not reported in table form). In another 22 percent of the major change cases, the interviewer changed the instruction and asked the respondent to start with someone other than the owner (e.g., the respondent himself or herself in many cases). In these instances, the person listed as Person 1 may or may not have been the owner or renter of the household and this could later cause problems for determining relationships within the household.

In over half of the major change cases, the interviewer omitted the month, day and year. In an additional 26 percent of major change cases, the interviewer left out just the year. Put together, in 80 percent of all major change cases, the interviewer either completely omitted or modified the reference date in this question. This indicates a problem with perceived redundancy with this question immediately following question 3 (which contains the date "April 1, 2006" twice).

As with Question 3, the house/apartment/mobile home fill caused problems for the interviewers in a third of the major change cases. Again, similarly to Question 3, in 37 percent of major change cases, the interviewer left out either "live" or "stay" or both words in the administration of this question. Having such a long and complex question seems to be leading interviewers to omit key terms that are meant to convey the concept of usual residence. The changes made by interviewers did not differ a great deal between English and Spanish cases.

Respondent Behavior

Again, respondents seemed to answer this question relatively easily, despite the sometimes differing question wording they had heard. In 93 percent of cases, an adequate answer was received on the first exchange (see Table 2a). Like in other questions, however, we do not know if respondents' answers would have been different if interviewers had administered the question exactly as worded instead of rephrasing it.

Spanish-Specific Problems

There is a grammatical problem in the way the term "owner" and the preceding article are written in the instrument. The term is translated with brackets containing both a masculine and feminine form of the word "owner." While appropriate for a written questionnaire, this is not a construction that can be read aloud as worded. This will require interviewers to make a modification in order to read the question.

The Census Bureau Mode Consistency Guidelines state that interviewer-administered instruments should contain questions written the way they would be read aloud by an interviewer. This will sometimes differ from the way something may be written or abbreviated on a paper instrument (U.S. Census Bureau Mode Consistency Guidelines, 2006).

The way the term "owner" is currently written, interviewers will have a number of choices as to how to handle it. To avoid wordiness they might choose only one term. They might read only the masculine form of the word, which—when the sex of the individual is unknown—would be appropriate for use in referring to either a man or a woman in Spanish. Interviewers might read only the feminine form, which would not be appropriate as this form assumes a female owner and could only be used to refer to a woman. Finally, they might choose to read both the masculine and feminine forms, but in order to do this they would have to make a change to the question wording. They would have to say "el propietario o la propietaria." This phrase would give extra emphasis to the fact that an owner could be either a man or woman.

Rather than leaving this to the discretion of each interviewer, the Census Bureau should make a decision on this and incorporate it into the instrument. We recommend using just the masculine term (el propietario) to refer to either a man or a woman, since this would make the question less wordy and complex. If there is a concern that this will cause people to leave women owners off of the form, cognitive testing of the two options may be in order. Since using the masculine to refer to either a man or woman is standard in Spanish, as long as the sex of the individual is unknown, cognitive testing for this purpose does not seem necessary.

Recommendations for Names

- In non-proxy interviews, it is much more natural to list the respondent first, even if he or she may not be the householder. Having the respondent list him/herself first eliminates the risk of leaving himself/herself off the form. Additionally, if we gather the respondent's name first, it eliminates the need to ask or verify with whom the interviewer is speaking (this is an item in the 2006 NRFU instrument that was not tested in this study). A separate question could be added to determine the owner or renter to be used to reckon relationships for the relationship question. See Childs et al. (2007) for an example of how to do this.
- Avoid repeating the date so frequently throughout the question series.
- Automatically fill house, apartment, or mobile home based on interviewer observation early in the interview to avoid this awkward fill.
- If the question stays as it is, use just the masculine term (el propietario) to refer to either a male or female owner rather than the current wording of "propietario/propietaria."

Undercount Question

The Undercount question follows the roster section of the survey and was designed to produce a more robust list of those staying at the household on the reference date. It intends to stimulate additional reporting for people who may not be considered "typical" or "regular" household members:

> **Question 5. Undercount**
>
> **We do not want to miss any people who might have been staying here on April 1, 2006. Were there any additional people staying here that you did not include, for example:**
> **Children, such as newborn babies or foster children?**
> **Relatives, such as adult children, cousins, or in-laws?**
> **Nonrelatives, such as roommates or live-in baby sitters?**
> **People staying here temporarily?**
>
> **No queremos dejar de contar a ninguna persona que pudiera haber estado quedándose aquí el 1 de abril de 2006. ¿Hubo alguna persona adicional que se quedaba aquí a la cual usted no incluyó, por ejemplo:**
> **Niños, tales como recién nacidos o hijos de crianza?**
> **Parientes, tales como jóvenes adultos, primos o parientes políticos?**
> **Personas no parientes, tales como compañeros[as] de cuarto o niñeras que viven en el hogar?**
> **Personas que se quedaban aquí temporalmente?**

Interviewer Behavior

In the 72 interviewed households, interviewers read the Undercount question exactly as worded or with minor changes only 35 percent of the time (see Table 2a). A substantial amount of undesirable interviewer behavior was evident; this question was part of the 3-way tie for the highest major change rate in the survey, with major changes in an astounding rate of 64 percent of cases.

An examination of the notes about major-change behavior indicates that most of the problems were due to omitting some (or parts) of the response options, rather than to omitting or changing parts of the question stem itself. In 85 percent of the major change cases all, some, or parts of response options were omitted (data not shown). In only 5 cases did interviewers omit all response options. However, in many more cases, interviewers omitted some of the response options or parts of the response options. This indicates that the response options are too long and complex to be administered in full by interviewers. Supporting this is the fact that this question had the highest rate of break-in by respondents at 9 percent. Though this is not an alarmingly high rate, it does suggest that the response categories may be too lengthy for a face-to-face interview.

As it is currently arranged, administration and response problems happen when the interviewer is prompting respondents to respond to categories that occur fairly infrequently. This causes a lot of work for interviewers and respondents. It would be less problematic for both parties if the response task was altered such that respondents could passively listen to the list and only stop the interviewer when a category is read that might pertain to them.

A less significant problem with this question was omission of all or part of the date in the question, similar to behavior seen in previous questions. In about a quarter of all major change cases, the date was omitted in part or whole.

Respondent Behavior

Despite the poor interviewer behavior, respondents were able to give an adequate answer in 80 percent of cases (see Table 2a). Though, as mentioned with previous questions, it is impossible to tell how respondents would have answered had they heard the entire stimulus as scripted.

Interestingly, this question had the second highest rate of qualified answers, at 10 percent of respondent behaviors. Of our 72 cases, 6 respondents mentioned a person (or persons) as possible household members and the interviewer did not add those persons or pursue discussion to determine whether or not those persons should have been listed. Since this question is aimed at gathering possible omissions, the fact that interviewers disregarded possible omissions is an alarming problem.

Recommendations for Undercount

- This question is a good candidate for a question series that combines the residence rule presentation with both coverage questions in a series of shorter questions. See Childs et al. (2007) for an example of how to do this.
- Interviewer training should emphasize enumerating, and not disregarding people that respondents mention here.
- Avoid repeating the reference date in every question.
- If the current method is kept, we recommend testing shortened response categories and not forcing an answer until all categories have been read. For example, the question could say:

  Please listen as I read a list of people that are sometimes missed in the census. These people include:

  Newborn babies, foster children;
  Adult children, cousins, in-laws;
  Roommates, live-in baby sitters; or
  People staying here temporarily.

  Were any of these kinds of people staying here on April 1, 2006?

Ownership Question

This question collects information about property ownership and various methods for securing the property (i.e., whether household members own the property free and clear or by way of a mortgage, and if not, whether they pay rent or occupy the property without payment of rent):

Question 6. Ownership

Is this [house / apartment / mobile home]…
Owned by you or someone in this household with a mortgage or loan?
Owned by you or someone in this household free and clear?
Rented for cash rent?
Occupied without payment of cash rent?

¿Es [esta/este] [casa/apartamento/casa móvil]
Propiedad suya o de alguien en este hogar con una hipoteca o préstamo?
Propiedad suya o de alguien en este hogar libre y sin deudas?
Alquilada por pago de dinero en efectivo?
Ocupada sin pago de dinero en efectivo?

Interviewer Behavior

Table 5 illustrates that coders' notes showed that interviewers routinely omitted various response options and often simply asked (or verified), "Is this house owned by you or someone in this household with a mortgage?" or "Is this house rented for cash rent?" (of the 48 cases with major change interviewer behavior, this occurred 33 percent and 25 percent of the time, respectively, across languages). This type of interviewer behavior often meant the "free and clear" owner option was not read to respondents and sometimes produced persistent inadequate responses (i.e., the interviewer never determined whether the owner had a mortgage or owned the property free and clear. This issue is discussed later in the Respondent Behavior section). Other frequent, but less common, major changes to question wording were, "Do you rent or own?" (17%) and omitting the last response option (i.e., "occupied without payment of cash rent"—15%). Interviewers may have noted cues from their physical surroundings and read only those portions of the question which applied to renters or owners—based on whether a respondent appeared to live in a house or apartment.

In the Spanish-language interviews, there was some evidence that interviewers were changing the third response option from "alquilada" to "rentada" (31%, see Table 5 above). Alquilada is the term commonly found in the dictionary for "rented." Rentada is an Anglicism that many Spanish-speaking immigrants in the U.S. use (commonly referred to as "Spanglish"). Further research is needed to see whether the term "rentada" is more understandable than the term "alquilada" to Spanish-speaking respondents in the U.S. Presumably, usage of this term would vary by educational level, English proficiency, proximity to the border and other factors. The fact that interviewers are using the term "rentada" instead of "alquilada" may also say something about their Spanish language proficiency.[12]

[12] Currently there is no formal proficiency test to certify NRFU field interviewers' Spanish proficiency level.

Table 5. Types of Major Changes to Question Wording for the Ownership Question

Interviewer Behavior, Major changes, N= 48	English N	English %	Spanish N	Spanish %	Total N	Total %
1. Asked or verified if rented (only)	9	28	3	19	12	25
2. Asked or verified if owned (only)	11	34	5	31	16	33
3. Asked "Rent or own?"	5	16	3	19	8	17
4. Read most, but not all, phrases (e.g., omitted "by you or someone in this household," read only "free and clear" for 2nd category)	2	6	6	38	8	17
5. Omitted last response option only	5	16	2	13	7	15
6. Reworded 3rd option to "rentada"	0	0	5	31	5	10
7. Other	0	0	3	19	3	6
Total	32	100	16	169	48	123

Note: In some cases interviewer behavior was coded under multiple categories and the N represents the actual number of problems identified, not cases.

Respondent Behavior

Respondents were able to provide adequate answers a good deal of the time (70%), but exceeded the 15 percent threshold in giving answers that did not initially seem to meet the question's objective (inadequate answers = 19%; see Table 2a). Coders' notes reveal that in about half the cases when respondents gave inadequate responses, they provided related but insufficient information for field coding purposes. For instance, respondents reported "I'm buying it," or in the case of a renter, "it's deducted from my check." In a few cases respondents seemed confused about the question intent; they reported who actually owned or rented the unit/home when it was someone other than the respondent (e.g., "my mother owns it," or "the owners own it"). Most delivered acceptable responses after further negotiation with interviewers, but in a few cases interviewers neglected to obtain enough information to properly differentiate between a respondent who owned the property with a mortgage or one who owned it free and clear.

Recommendations for Ownership

- A simplified version of the question could help eliminate the need to rely on interviewers to follow-up consistently and correctly when homeowners do not provide enough information to field code a response (i.e., they neglect to tell the interviewer whether they have a mortgage or own the property free and clear). This could be accomplished by dividing the existing questions in two, first asking whether the unit is owned, rented or occupied without payment and then whether it is owned with a mortgage or not. Further evaluation would be needed to ensure the proposed question functions as intended. Such a question could look like this:

Is this house owned by you or someone in the household, is it rented or is it occupied without payment of rent?
Owned **–** GO TO FOLLOW-UP QUESTION
Rented
Occupied without payment of rent

[Follow-Up Question**] Is it owned with a mortgage or owned free and clear?**

¿Es esta casa propiedad suya o de alguien en este hogar, está alquilada, o está ocupada sin pago de alquiler?
Propiedad — GO TO FOLLOW-UP QUESTION
Alquilada
Ocupada sin pago de alquiler

- The last two options present the terms rented and occupied (alquilada and ocupada) only in feminine form. If the respondent lives in an apartment, those adjectives would need to be masculine. This issue could be resolved by having a house/apartment/mobile home fill throughout the instrument. The terms "rented" and "occupied" would also need to be fills in Spanish to account for masculine or feminine forms of the adjectives. They would read [Alquilado/Alquilada] and [Ocupado/Ocupada].

Relationship Questions (1 and 2)

These questions collect information to define how each household member is related to "Person 1," or the householder who owns or rents the home. These questions first attempt to determine whether or not the pair of household members is related. Then, respondents are asked to review a flashcard and select a category that denotes the relationship for those two particular people. The wording for both questions appeared as follows:

Question 7. Relationship1

[Are you/is NAME] related to [NAME]?
Yes – Go to 8a
No – Go to 8b

¿Está [NAME/usted] relacionado(a) con [NAME/usted]?
Sí, es pariente – Go to 8a
No, no es pariente – Go to 8b

Question 8. Relationship2

a. Which one of these categories best describes how [you are/NAME is] related to [NAME]?

Husband or wife
Biological son or daughter
Adopted son or daughter
Stepson or stepdaughter
Brother or sister
Father or mother
Grandchild
Parent-in-law
Son-in-law or daughter-in-law
Other relative

b. Which one of these categories best describes [your/NAME's] relationship to [NAME]?

Roomer, boarder
Housemate, roommate
Unmarried partner
Foster child or foster adult
Other nonrelative

a. ¿Cuál de estas categorías es la que mejor describe cómo está [usted /NAME] relacionado(a) con [usted/NAME]?

Esposo o esposa
Hijo o hija biológico(a)
Hijo o hija adoptivo(a)
Hijastro o hijastra
Hermano o hermana
Padre o madre
Nieto(a)
Suegro(a)
Yerno o nuera
Otro parentesco

b. ¿Cuál de estas categorías es la que mejor describe la relación [de NAME] con [usted/NAME]?

Inquilino o pupilo
Compañero de casa o cuarto
Compañero(a) no casado(a)
Hijo(a) de crianza o adulto bajo custodia
Otro no pariente

Interviewers were instructed to use a flashcard that accompanied the Relationship2 question; it displays all of the associated response categories. This Relationship flashcard was presented to

respondents in 28 percent of the interviews across both languages; it was used 25 and 33 percent of the time during interviews conducted in English and Spanish, respectively (Rappaport et al., 2006).

Interviewer Behavior

Interviewers read the Relationship1 question as worded (or with slight modifications) 58 percent of the time, but delivered the Relationship2 question far less accurately (reading it as worded only 12% of the time; see Table 2a). Interviewers made major changes to the question wording slightly less frequently for Relationship1 than Relationship2 (14% compared to 22%, respectively; see Table 2a). According to coders' notes, when interviewers were not reading the first question as worded, they were most often attempting to verify the relationship between the reference person and another household member (e.g., "John is your son?"—data not shown). These verifications were usually appropriate interviewer behavior since there was evidence on the tape that the respondent had mentioned the relationship earlier in the interview. However, in a small number of cases, the verifications did not appear to be based on prior information supplied by the respondent and may have been a guess on the part of the interviewer.

Table 6 shows that in the second question (Relationship2/Question 8), coders notes indicated that when interviewers made a major change to the wording, they often abbreviated and paraphrased by merely asking, "And John?" (42%). This behavior was common when the interviewer had established a "question routine" after asking for the relationship between the household members and the reference person (e.g., "And how is Mary related to Sally?", "And John?", "And Ben?"), though coders' notes indicate that paraphrasing did sometimes occur the first time interviewers posed the question as well. This is problematic for this question because direction is not implied when an interviewer asks "And John?" – does it mean how John is related to Mary or to Sally? Would the answer be John is Mary's son, or Mary is John's mother? This may have lead to increased problems with the relationship data observed in the 2004 Census Test (Love and Byrne, 2005). By revising the instrument to script where it is appropriate to shorten questions, we may be able to decrease the likelihood that interviewers shorten questions when it is detrimental to understanding.

Another poor interviewer behavior, though observed less frequently, was observed when the interviewer read nothing at all and simply showed the response options to the respondent (13%, Table 6). Additionally, some interviewers read a few response options that seemed to match information given by the respondent in the Relationship1 question. In these cases, interviewers read a list of categories that were either related (16% in Table 6, e.g., biological son, adopted son) or unrelated in nature (10% in Table 6, e.g., roomer/boarder, housemate/roommate, other nonrelative).

There were very few Spanish cases for which coder notes were available, but Spanish-speaking interviewers appear to have shortened the question by reading only the "related" options, by simplifying the question and asking, for example, "Is he a relative or non-relative?" or by asking about all residents in one question, saying something like "Are they all your sons?"

Table 6. Types of Major Changes to Question Wording for the Relationship2 Question

Interviewer Behavior, Major changes, N= 31	English N	English %	Spanish N	Spanish %	Total N	Total %
1. Rephrased/shortened question (e.g., "And John is?" or "And John?" or "And how're…?")	13	54	0	0	13	42
2. Omitted question, read "related" options only	4	17	1	14	5	16
3. Omitted question, read "unrelated" options only	3	13	0	0	3	10
4. Omitted question, showed list to respondent	4	17	0	0	4	13
5. Asked "Relative or nonrelative?"	0	0	2	29	2	6
6. Verified by reading one response option	0	0	2	29	2	6
7. Asked about all residents at the same time	0	0	1	14	1	3
8. Other	0	0	1	14	1	3
Total	24	101	7	100	31	99

Note: Due to rounding, all total percentages do not equal exactly 100.

The most frequently occurring undesirable interviewer behavior for the Relationship2 question, however, was due to interviewers erroneously omitting the entire question (also called "inappropriate skip" or "IS") from the interview, which occurred 52 percent of the time (see Table 2a). This inappropriate skipping behavior occurred far less frequently in the Relationship1 question (12%—see Table 2a). Evidence in the coders' notes suggests interviewers found either Question 7 or 8 redundant and omitted one or the other. When respondents provided more than a "yes/no" answer in Question 7 (e.g., "she's my daughter"), interviewers tended to skip Question 8; they likely entered a response for Question 8 in the instrument and moved on. There is also evidence in coders' notes that interviewers skipped Question 7 and asked Question 8 instead (e.g., when no previous information was obtained, an interviewer might ask, "And NAME is related to NAME, how?" and when this same question was posed for later household members interviewers merely asked, "And NAME?").

Respondent Behavior

In order to analyze how each of these questions performed, the coders were trained only to accept a "yes/no" response as an adequate answer in Relationship1 because Relationship2 was designed to capture the actual relationship designation. This strict approach to coding acceptable answers for the first question lowered the percentage of adequate answers to 53 percent (see Table 2a). So, at least half the time respondents were complying with the "yes/no" question format. But, when respondents did not comply, they were prematurely offering information that often satisfied (or would begin to satisfy) the next question (Relationship2). For example, when responses to the question "Are you related to John?" were *not* given in a "yes/no" format, coders' notes (in 42 out of 47 cases, in English) indicate that respondents said things like, "she's my daughter" or "I'm the wife." At this point, interviewers might feel as if they have sufficient information to code responses for Question 8 (Relationship2) without asking anything further. It

is easy to see, then, why the incidence of skipped questions is so high for the follow-up question (63 percent, combining correct and incorrect skips; see Table 2a). Respondents often reported the exact nature of the relationship in response to the Relationship1 question.

In looking at final outcome rates to both Relationship questions, the low rate of adequate answers—56 and 68 percent for questions 1 and 2 respectively in Table 2b—may suggest more serious data quality issues. Looking at responses across both relationship questions, many respondents (41 reported children in 72 total households, data not reported in table form) reported "son" or "daughter" without specifying initially "biological," "step," or "adopted." In these 41 instances, there were only 10 cases where the interviewer actually probed about type of child (and in most of those cases, the interviewer simply asked "biological?"). When a "son" or "daughter" response was given, and the interviewer did not probe for more detail, it would have been impossible to properly identify which response should have been coded.

There was a significant difference in respondent behavior (both initial response and final outcome) by language for the first Relationship question (though the difference was not significant for the second question, the trend is similar). Fewer adequate answers were provided by respondents in the Spanish-language interviews, compared to the English-language interviews. For English cases, Questions 7 and 8 rendered adequate responses at the first level of exchange at 76 to 72 percent, respectively (see Table 4). In contrast, for Spanish cases, adequate responses were given around 29 and 25 percent, respectively. In the "final response," adequate responses increased to 33 percent for the Spanish cases in Question 7, but there remains a sizable difference compared to the final outcome in the same question for the English cases at 78 percent.

In both cases, such high rates of inadequate answers are unacceptable, but the issue in the Spanish-language cases is more dramatic. Coders' notes in these cases were sparse (only 16 cases) and do not reveal the reason for the difference. Both English- and Spanish- speaking respondents provided inadequate responses such as "son" or "daughter"—neglecting to report whether these children were biological, adopted, or foster children. In a few cases, responses classified as "unrelated" were impossible to field code without additional probing on the part of the interviewer. For example, respondents often said things like "He's my boyfriend," "one lives in the back and one lives in the front," "she's my niece," or they may have simply said "cousin." These are not acceptable response categories by themselves and interviewers did not always follow this type of response with further probes. The interviewers may have gone ahead and chosen categories such as "other relative" or "other nonrelative" without consulting the respondent.

The greater difficulty observed in Spanish-language interviews is probably due to a number of factors. It may be an indication that the translation of this question does not sound natural, so respondents are confused after hearing it read aloud. The question is quite long and refers to the list of categories in a rather complex way. A more natural and conversational question would read simply "Looking at this list, how are [you/ NAME] related to [you/NAME]?" Translation: "Utilzando esta lista, ¿Cuál es la relación entre [usted/NAME] y [usted/NAME]?" Additional testing of any new translation is recommended, as it is difficult to indicate the direction of the relationship desired. Additional testing could help determine whether respondents can report the

relationship in the direction we desire after hearing the new question wording. This new wording should also be tested in English.

Other Considerations

It should be noted that some of the Spanish terms in this question have been revised through the cognitive testing of the Decennial Bilingual PAPI questionnaire (Caspar, et.al forthcoming). For example, "unmarried partner" is now translated as "pareja no casada." In addition, "Foster child and adult" have been dropped in both the English and the Spanish versions of the question. Findings from that study also indicate that the term "pupilo" as a translation for "boarder" is not being understood by the vast majority of respondents, and that "parent-in-law" in English might be better written as "father-in-law or mother-in-law" if space allows on the paper questionnaire. If these terms are updated or changed in one instrument, we would recommend that they be changed across instruments.

An additional issue related to the Spanish translation of this question is that on the flashcard and in the response categories, the endings for the relationship terms are not written consistently but rather were written to follow the English version in a way that is not most logical in Spanish. Because the relationship terms need to have either a masculine or feminine ending depending on whether the person in question is a man or a woman, the instrument lists both masculine and feminine versions of the terms. There are two ways to do this that are both correct. You could write out the entire masculine and feminine terms: e.g., hijo o hija (son or daughter) or you could write the masculine term with an "a" in parenthesis: e.g., hijo(a). The current version of the question employs both of these strategies, apparently in an attempt to follow the English version: Where the English says "husband or wife," the Spanish says "esposo o esposa." Where the English says "grandchild," the Spanish says "nieto(a)." This variation does not make sense if you look at the Spanish instrument on its own. We would recommend choosing one way to write the terms in Spanish, and being consistent throughout the question. Since it is shorter to write "hijo(a)" than "hijo o hija" and both are equally valid, we recommend shortening the options whenever possible. This would not be possible for terms that vary by gender in more than just the ending, such as the "son-in-law or daughter-in-law" (yerno o nuera).

An additional issue is that it is common for recent immigrants to live with non-nuclear family members and non-relatives—categories of people who may not be listed as one of the current response options on the flashcard. Because the NRFU instrument is electronic and because it is used to enumerate large numbers of immigrant households, it might be desirable to revisit the list of categories currently provided.

Finally, many Spanish-speaking immigrants in the U.S. have lower levels of education and literacy than the general population, and reading the flashcard maybe be a difficult task for them. This may lead them to provide responses that are not listed on the card at a greater frequency.

Recommendations for Relationship

These results echo the results from 2004 (see Hunter and Landreth, 2005) demonstrating the confusion and data quality issues that arise when multiple questions try to measure various parts of the same construct. There is evidence that these questions should be combined into a single item. Tailored follow-up questions could be added to address

scenarios that are likely contributing to poor data quality. Areas that could be tested for future implementation are:

- The Spanish translation employed for this question should be tested to find out whether respondents are having difficulty understanding it when they hear it read aloud. Alternative, simplified wording should also be tested. Simpler wording could be tested in English as well. Cognitive testing would be an ideal way to examine this issue.
- Respondents often offered the exact nature of the relationship in response to Question 7, though they tended not to provide the exact information needed to properly code responses for Question 8 (e.g., the respondent says "son" instead of "biological son"). Since the list of response options is long, data quality could be increased by removing the "biological/step/foster" categories and turning them into follow-up questions if the "son/daughter" response option is selected by the interviewer (Chan, 2007 documents testing of a similar approach for use in the American Community Survey). This would ensure that interviewers administer the correct follow-up probes to collect better data and lessen respondents' cognitive burden by shortening the list by a few categories. Recommended wording for a combined Q7/Q8 Relationship Question reads as follows:

How [is/are] [you/NAME] related to [you/NAME]?

Husband or wife	Unmarried partner
Son or daughter	Foster child or foster adult
Brother or sister	Roomer or boarder
Father or mother	Housemate or roommate
Grandchild	Other non-relative
Parent-in-law	
Son-in-law or daughter-in-law	
Other relative	

[If "son/daughter," selected above, then ask a follow-up question:]
Is [NAME/he/she] your biological, adopted, or step-son or daughter?

¿Cuál es la relación entre [usted/NAME] y [usted/NAME]?

Esposo o esposa	Compañero(a) no casado(a)
Hijo o hija	Hijo(a) de crianza o adulto bajo custodia
Hermano o hermana	Inquilino o pupilo
Padre o madre	Compañero de casa o cuarto
Nieto(a)	Otro no pariente
Suegro(a)	
Yerno o nuera	
Otro pariente	

[If "son/daughter," selected above, then ask a follow-up question:]
¿Es [NAME] su hijo(a) biológico(a), su hijo(a) adoptivo(a) o su hijastro(a)?

Sex Question

This question records the gender of each household member:

Question 9. Sex

[Is NAME /Are you] male or female?

¿Es [NAME/usted] de sexo masculino o femenino?

Interviewer Behavior

Interviewer behavior was far from optimal for this question (exact reading occurred 38% of the time and the question was omitted 11% of the time; see Table 2a). The rate of verification was 25 percent (see Table 2a). Coders were instructed to assess a verification of the requested information as acceptable only when evidence on the tape indicated that interviewers had been given the information previously in the interview. Coders' notes indicate that there were many examples of gender information made available to interviewers in the preceding questions (i.e., Q7/8 Relationship). Respondents often said things like, "he's my husband," and "she's my daughter." In addition to these cues, interviewers may also have attended to gender patterns associated with names (e.g., not many females are named John) and used this information to successfully verify responses to this question instead of asking it as worded.

When interviewers changed question wording to a significant degree, coders' notes suggest they mainly posed the question in the form of "And [NAME]?" for Person 2 and beyond when there were multiple household members (see Table 7). Table 7 also shows that in a few cases interviewers buffered a question they might perceive as awkward by adding some type of normalizing statement either directly before or after the question (e.g., "I have to ask everything as it is on the screen"). Additionally, a few times in the Spanish-language interviews, interviewers dropped the phrase "de sexo" or reworded the question as "Is he a boy?" or "Is she a woman?"

Table 7. Types of Major Changes to Question Wording for the Sex Question

Type of Major Change	Observations	Percent
Rephrased question (e.g., "And John?")	37	69%
Added normalizing statement and read question exact	10	19
Reworded question as "Is he a boy?/Is she a woman?"	3	6
Dropped "de sexo" in Spanish-language interviews	3	6
Other	1	2
TOTAL	54	102%

Note: Due to rounding, total percentage does not equal exactly 100.

A slightly greater proportion of Spanish speaking interviewers seemed to be changing the wording of this question, for example dropping "de sexo" from the end of the question. While the Spanish translation for this question is standard on written questionnaires, it is not conversational and is a more complex wording than is the English. The question in Spanish literally reads "¿Es [NAME/usted] de sexo masculino o feminino?" meaning, "[Are you/is NAME] of masculine or feminine sex?" The reason that this item was not translated literally to say "male or female" as it reads in English is that the terms used to express the concepts of male and female in Spanish (macho and hembra) are typically used only to talk about animals and it would be offensive to refer to a person this way.

To say "Are you of masculine or feminine sex?" sounds about as natural and conversational in Spanish as it does in English, which probably explains why Spanish-speaking interviewers are dropping parts of this question. A more natural way to express this question in Spanish in a face to face conversation would be to ask "Es [NAME/usted] hombre o mujer?" meaning "[Are you/Is NAME] a man or a woman?" This would ideally be modified based upon the age of the person in question and of younger household members we could ask "¿Es [NAME/usted] niño o niña?" meaning [Are you/Is NAME] a boy or a girl?" This difference in wording could be programmed into the instrument as an automatic fill based upon the age of the person in question. However, in order to do this, we would need to test the possibility of asking date of birth and age prior to asking about gender. If an automatic fill is not possible because of question ordering issues, we recommend testing the use of "Es [NAME/usted] hombre o mujer?" meaning "[Are you/Is NAME] a man or a woman?" regardless of the age of the person in question. We believe that this would most likely still be an improvement over the unnatural sounding "Are you of masculine or feminine sex?" in Spanish.

As noted in the table above, the most common change that interviewers made to this question was in reading it as "and NAME?" rather than repeating the exact wording for each household resident. The problem with the more cumbersome and less conversational Spanish question wording would be solved by requiring interviewers to read the question only once and then to read "and NAME?" for persons after Person 1.

Since the NRFU instrument is often used to interview Spanish-speaking respondents of lower-than-average educational levels, it is advisable to make these questions as conversational as possible, rather than using complex and/or academic sounding terms. Further testing of this suggested rewording is advisable.

Respondent Behavior

Despite the difficulties in question administration described above, respondents were able to provide adequate answers a good deal of the time (in 79 percent of the cases; see Table 2a). Since most other respondent behavior types rarely occurred (only in 16 percent of the cases, which were inaudible), there were very few notes left by the coders. In these few cases (6), a few respondents replied "no" and a few people merely laughed instead of providing a response during the first level of exchange (data not shown). The inaudible cases are easily explained by the respondent having nodded his or her head in response to a verification question.

Recommendations for Sex

- In the 2004 study (Hunter and Landreth, 2005), the authors suggested permitting interviewers to verify gender rather than asking the question as worded for each household member. In light of the current results, this approach still seems valid.
- In addition, asking "And John?" for Persons 2 and beyond is more in line with a natural conversational style using a topic-based question administration approach—versus requiring interviewers to recite the exact wording in rapid succession for each member of the household.
- Consider testing and implementing an alternative wording for the formal Spanish translation of this question as discussed above. We recommend the more conversational "Es [NAME/usted] hombre o mujer?" meaning "[Are you/Is NAME] a man or a woman?" for Person 1, followed by "and NAME?" for later persons in the household.

Date of Birth Question

The date of birth question is aimed at capturing the month, date, and year of birth of each person in the household. It is intended to be read as:

Question 10. Date of Birth

What is [your/NAME's] date of birth?

¿Cuál es [su] fecha de nacimiento?
OR
¿Cuál es la fecha de nacimiento [de NAME]?

Interviewer Behavior

This question suffered from major change 29 percent of the time (see Table 2a). Although this is a short, straightforward question, interviewers generally made four types of major changes. These changes sometimes occurred independently, but often occurred in combination with one another. The four types of changes made by interviewers were:

1) Asking for "birthday" instead of "date of birth" in English. Similarly, Spanish-speaking interviewers sometimes said only "fecha" or "date" instead of "fecha de nacimiento" or "date of birth." This interviewer behavior did not typically produce bad respondent behavior or outcomes (although sometimes the first response and outcome were still "don't know.")
2) Asking "and NAME?" for Persons 2 and higher. This typically resulted in an adequate answer or "don't know," but occasionally resulted in a request for clarification or reread.
3) Asking "do you know" the date of birth instead of "what is" the date of birth. Similarly, the interviewer occasionally prefaced the question with "do you happen to know" or ended the question with "by any chance." Often the respondent would say "no," resulting in a "don't know" first response and outcome. Offering the respondent an "easy out" by asking if they know the information may increase "don't know"

responses. Interviewers may be doing this because in the past they have encountered respondents who do not know the dates of birth for every household resident. They may also be doing this to soften the tone of the question and to sound more polite. This happened in both English and Spanish language, 6 cases and 3 cases, respectively.

4) In a few additional Spanish-language cases, the interviewer added words or phrases to make the question sound more polite. For example, the interviewer added "Could you tell me…" prior to the scripted question or added the term "señora" or "ma'am" to the question, saying something like "And what is your mother's date of birth, ma'am?" This may indicate cultural differences in conversational politeness that interviewers perceive to be lacking in the Spanish-language instrument.

There was a significant effect of repeated question administration on good interviewer behavior for this question. For Person 1, the question was asked as intended or verified positively 85 percent of the time, while this occurred only 55 percent of the time for Persons 2 and higher (see Table 3). This was primarily due to interviewers shortening the question for Persons 2 and higher to "And NAME?" or "And NAME's birthday/birthdate?"

Respondent Behavior

Respondent "don't know" responses were fairly common (8 %; see Table 2a). The respondent would often either give no further information at all, offer an age, or compare someone else's age to the person for whom they were answering. Occasionally, respondents knew the month and date, but not the year.

Recommendations for Date of Birth

- Consider testing the addition of an optional sentence to make this question read as follows:

  "What is [your/NAME's] date of birth? (Please include the day, month and year.)"

  ¿Cuál es [su] fecha de nacimiento? or **¿Cuál es la fecha de nacimiento [de NAME]? (Por favor, incluya el día, el mes y el año)**

  This rewording might reduce the likelihood of the interviewers offering a "do you know" type of question. It may also spur respondents to answer with any part of the birthdate that they know, rather than simply saying that they don't know.
- Further study of the politeness issue is in order for this question. If respondents are relatively likely to have difficulty answering this question off the top of their heads for every household member, interviewers might be reluctant to ask it in such a straightforward and direct manner. Respondents who do not know the dates of birth of close family members most likely experience and exhibit embarrassment. Interviewers seem to be softening the question by adding phrases like "do you know?" or "could you tell me?" Some form of politeness could be added to the question text directly.

Age Question

The age question is aimed at determining age for each resident in the household for whom a date of birth was not given at Question 10. It is intended to be read as:

Question 11. Age

> **What was [your/NAME's] age on April 1, 2006?**
>
> **¿Cuál era la edad [de NAME] el 1 de abril del 2006?**
> **¿Cuál era su edad el 1 de abril del 2006?**

Interviewer Behavior

This question was correctly skipped 80 percent of the time because date of birth was given for that portion of the sample (see Table 2a). Incorrect skips occurred 5 percent of the time (in these cases, a date of birth had not been given, but this question was skipped anyway). When this question was read (N=31), major change occurred about half of the time (17 out of 31). Major changes consisted of asking "how old" a person was rather than using the word "age," omitting the April 1, 2006 date, and asking if the respondent knew the age. These problems were occasionally combined. There were very few coder notes on Spanish cases, so most of the changes reported were in English-language interviews. It does appear that interviewers changed the question in similar ways across the two languages.

> ***Recommendations for Age***
> No potential solutions at this time. Emphasize in training the importance of reading the question exactly as worded, particularly the April 1, 2006 reference date. Perhaps explaining to interviewers that the Census needs Age as of April 1st rather than the day that they interview someone would help. If the date is eliminated from questions where it is not necessary, interviewers might find it easier to read it in this question.

Age Check Question

The age check question is aimed at confirming age for each resident in the household, usually calculated from date of birth. However, it is asked for every household member, even when Question 11 was asked and the respondent provided an age verbatim. It is intended to be read as:

Question 12. Age Check

> **For the census, we need to record age as of April 1, 2006. So, just to confirm – [NAME was/you were] [AGE] on April 1, 2006?**
>
> **Para el Censo, tenemos que anotar la edad que se tiene al 1 de abril de 2006. ¿Por lo tanto, solo para confirmar tenía [usted/nombre] [age] años el 1 de abril del 2006?**

Interviewer Behavior

This question suffered from major change 52 percent of the time and was incorrectly skipped 20 percent of the time (see Table 2a). Interviewers regularly omitted the entire first sentence of the question, the entire reference date or just the year of the reference date. They also occasionally replaced "2006" with "this year." They also frequently omitted "so just to confirm" in the second sentence. In addition, "so just to confirm" was sometimes reworded as "just to check" in English or "para chequear" (in order to check) in Spanish. This rewording occurred in 8 Spanish cases and 4 English cases.

Interviewers also frequently changed the verb form to present tense (e.g., "So, just to confirm, John is 24?"); this coincided with omitting the reference date and happened much more frequently in Spanish than in English (22 cases in Spanish and 5 cases in English). This is a particularly troublesome problem because it changes the meaning, and the potential answer to the question. For any household member who had a birthday between April 1 and the day of the interview (which occurred between May 1 and late June to mid-July), the correct answer to this question would have been different when posed in the present tense without the April 1 reference date.

Perhaps not surprisingly, there was also an effect of repeated question administration on good interviewer behavior for this question. It was asked as intended or verified correctly 44 percent of the time for Person 1, but only 19 percent of the time for Person 2 (see Table 3). This is likely due to the fact that it is repetitive and unnecessary to re-read the first sentence for each person in the household in the topic-based format (in direct sequence). Many interviewers were, in effect, establishing the reference date when asking about Person 1 and not repeating it for each subsequent household resident.

Respondent Behavior

There were few respondent and outcome problems, but those few included requests for clarification and "don't know" responses.

Recommendations for Age Check

- Script the first sentence only for Person 1. Having to repeat it multiple times in a row creates undue interviewer and respondent burden and most likely does not elicit additional useful information.
- Omit this question for persons for whom the respondent has given an age in response to Question 11. It seems unnecessary to first ask for age, and then immediately confirm it. Confirmation of age makes sense in cases where only date of birth has been given by the respondent.
- Consider revising the phrase "So, just to confirm" or "Por lo tanto, solo para confirmar" to make them sound more conversational and easier to read. The alternative wording brought up by interviewers: "So, just to check" or "entonces, solo para chequear" might be a valid replacement. Further cognitive testing of these new options may be in order.

Race and Ethnic Origin Series

The 2006 instrument contains a three part series aimed at measuring race and ethnicity, including a Hispanic origin question, a race question and an ancestry question. This series is intended to gather all detailed origins and races through the ancestry question, rather than having a series of follow-ups, like that used in the 2004 "2-part" series.

Hispanic Origin Question

The Hispanic origin question is designed to gather information on whether or not each person is of Spanish, Hispanic or Latino origin. It precedes the questions on race and ancestry.

The Hispanic origin question was scripted as follows:

Question 13. Hispanic Origin

"[ARE YOU / IS NAME] of Hispanic, Latino or Spanish origin?"

"¿Es [usted/NAME] de origen hispano, latino o español?"

Interviewers were instructed to read the question as worded in full for Person 1 and for every additional household resident.

Interviewer behavior

Overall, the Hispanic origin question was asked exactly as scripted or with minor changes only 50 percent of the time (see Table 2a). Interviewers made a major change to the question wording in 29 percent of the cases overall (see Table 2a). Though the difference was not significant by language, when examining the data by language, we see that this question was asked as intended in 61 percent of the English cases and in only 33 percent of the Spanish cases (see Table 4).

Correct interviewer behavior differed significantly by repeated question administration (see Table 3). The question was asked as intended 92 percent of the time for Person 1 but only 28 percent of the time for later persons in the household (Table 3). Similarly, the interviewer made a major change in question wording only 7 percent of the time when reading the question for Person 1, but made a major change 41 percent of the time when reading the question for Persons 2 and higher (full data not presented in this report). The question was skipped only 1 percent of the time for Person 1, but it was skipped 30 percent of the time for later people (full data not presented in this report). Generally, the question was administered rather well for Person 1, but was not administered as desired for the rest of the household.

A review of the notes provided by the coders indicates that the most commonly occurring change in both English and Spanish was for interviewers to shorten the question for residents higher than Person 1, saying, for example, "How about John?" rather than reading the question as worded. In other cases, interviewers asked respondents questions such as "Is Maria the same?" rather than reading the question as worded (see Table 8).

Table 8. Types of Major Changes to Question Wording for the Hispanic Origin Question

Interviewer Behavior, Major changes, N= 65	English N	English %	Spanish N	Spanish %	Total N	Total %
1. Rewords to "how about NAME?"	27	79	15	48	42	65
2. Rewords to: "is NAME the same?"	2	6	2	6	4	6
3. Omits "Spanish" (español) from question	3	9	5	16	8	12
4. Omits "Latino" from question	1	3	5	16	6	9
5. Omits "Hispanic" (Hispano) from question	0	0	1	3	1	2
6. Asks about multiple residents in 1 question	1	3	2	6	3	5
7. Other	0	0	1	3	1	2
Total	34	100	31	98	65	101

Note: In some cases an interviewer was coded as having made more than one major change to the question.
Due to rounding, all total percentages do not equal exactly 100.

Another common major change to the question occurred when interviewers only read one or two of the three terms listed in the question (Hispanic, Latino or Spanish). This occurred in 4 of the English-language cases, with interviewers most often dropping the term "Spanish" from the question. It also occurred in 7 of the Spanish-language cases, with interviewers most often dropping both the terms "Latino" and "Spanish." There was only 1 Spanish-language case in which the term "Hispanic" was dropped from the question, and in this case the interviewer also dropped "Latino" and only read "Spanish."

This type of interviewer behavior may occur in response to the tendency for some respondents to interpret the question as a request to choose an "option" when asked "are you of Spanish, Hispanic, or Latino origin?" rather than to answer with a "yes" or "no" response. There is evidence that this happens more often in Spanish-language interviews. When a respondent is asked this question in Spanish, he/she may feel that it should be obvious to the interviewer that he/she is Hispanic since the conversation is occurring in Spanish. For this reason the respondent may interpret this as a multiple choice question. There is evidence from cognitive testing studies that respondents often struggle to choose one of the three "options" in this question, and express a great deal of confusion (Goerman, forthcoming). The interviewer may compensate for this potential misunderstanding by reading only one of the examples, which makes the question a more obvious yes/no question to respondents. An alternative explanation is that interviewers may find the three terms redundant, or they might believe that respondents will identify with one or two of the terms more than the other/s.

Another major change made by just a few interviewers was asking a respondent to say whether multiple residents in the household were Hispanic with one question, saying, for example, "So, your kids are all Hispanic?" This was noted in 1 English-language case and 2 Spanish-language cases.

Respondent behavior

We trained the coders to assess responses based on whether or not they met the measurement objectives of the question. Thus, several types of responses to this question were considered adequate for coding purposes. "Yes/no" responses were considered to be correct, as were responses taken directly from the origin categories offered in the question itself (e.g., "I'm Hispanic," "I'm Latina," or "I'm Spanish"). We presume that a response such as "I'm Hispanic" would be easily coded as "yes" into the instrument by interviewers and therefore do not consider this type of answer to be a concern. However, there is some evidence that respondents experience unnecessary difficulty with this question due to confusion over what type of information is being requested. For this reason, we examined all responses to this question carefully.

Overall, respondents gave codable answers to this question 69 percent of the time on the first exchange (see Table 2a). However, this also differed significantly by language (see Table 4). In English-language cases, respondents provided a codable answer 82 percent of the time, and in Spanish-language cases they provided codable answers only 57 percent of the time (see Table 4).

There were a relatively high number of inaudible responses to this question, 6 percent of cases in English and 8 percent of cases in Spanish (data not presented in this report). This could be because it is a yes/no question, which encourages some respondents to nod or shake their heads instead of verbalizing a response.

Behavior coder notes indicated that the majority of respondents gave a yes or no response. However, 75 percent of the respondents who gave a yes/no response were non-Hispanic English speakers who responded "no." Of the 33 Spanish-speaking Hispanic respondents, only 2 responded "yes" and of the 35 English-speaking Hispanic respondents, only 5 responded "yes" (see Table 9). It was most common for Spanish-speaking Hispanics to respond by giving their nationality, saying for example, "I'm Mexican." A number of other Spanish speakers, 18 percent, misinterpreted the format of the question and picked one of the three "response options," saying for example, "I'm Hispanic" or "she's Latina." This type of response is presumably easily coded as a "yes" response by interviewers and was therefore considered to be an adequate answer.

A potentially more serious misunderstanding occurred when Spanish speakers interpreted the question to be asking about their citizenship. Five respondents replied to this question with answers such as, "I was born here" or "we're all American citizens." Given the current political climate surrounding immigration law and reform, having a question that appears to be singling out Hispanic respondents to ask about their citizenship or nationality could contribute to respondents' fears and could even lead to non-response. This is an issue that requires further research.

English-speaking Hispanic respondents tended to answer the Hispanic origin question by choosing one of the three "options," most often saying "I'm Hispanic." This was true in 51 percent of the cases. Six of the English-speaking Hispanic respondents also responded with a nationality such as "I'm from Honduras." One of these respondents also interpreted the question as one about U.S. citizenship, and two others responded to this question in a way that would be more appropriate to the ancestry question, saying "I'm Hispanic-Haitian" or "I'm Mexican-

American." Finally a couple of respondents expressed confusion, saying "I don't know how to answer that" or "I'm half."

Table 9. Types of Responses to the Hispanic Origin Question

Response type	Type of respondent: Hispanic (Span) (n=33)		Hispanic (Eng) (n=35)		Non-Hisp (n=33)		Total (n=99)	
	N	%	N	%	N	%	N	%
1. Yes/No	2	6	5	14	25	75	32	32
2. Hispanic/Latino/Spanish	6	18	18	51	0	0	24	24
3. The same/also	6	18	1	3	2	6	9	9
4. Nationality	13	39	6	17	0	0	19	19
5. U.S. Citizenship	5	15	1	3	1	3	7	7
6. Race	0	0	0	0	1	3	1	1
7. Ancestry	0	0	2	6	1	3	1	1
8. Other misinterpretation	1	3	2	6	3	9	6	6
Total	33	99	35	100	33	99	99	99

Note: Due to rounding, all total percentages do not equal exactly 100.

While non-Hispanic English-speaking respondents appeared to have the least amount of difficulty with this question, even they experienced some problems. Seventy-five percent of these respondents replied with a simple "no." The few respondents who had difficulty with the question gave a variety of different types of responses. One person said "he was born here," one gave an answer that would be more appropriate for the race question, saying "I'm black," and one gave a response that would be more appropriate for the ancestry question, saying "I'm Italian-Irish." Three respondents had various other confusions or misinterpretations. One person responded by asking "Do I look like it?" Another respondent did not understand the question and a final respondent asked "What are the choices again?"

Some respondents answered this question by saying "the same thing" or "también," meaning "also." We don't have full information on the reasons for this since coders' notes were sometimes lacking in detail. In a lot of these cases this response seems to have been led to by the interviewer asking "and what about NAME?" or "is NAME the same?"

Discussion

This question seems to work best for non-Hispanic, English-speaking respondents. These people seem to interpret the question as a yes/no question with relatively little difficulty. For Hispanic respondents more confusion arises. When these respondents are either having a conversation in Spanish or speaking with an interviewer face-to-face, they may think that it is obvious that they are Hispanic. This context may lead them to interpret the question as a multiple choice question. It sounds to many of them like the survey is asking if they are *either* Hispanic, Latino or Spanish. While living in Spanish-speaking countries, people do not refer to themselves as "Hispanic" or "Latino" as these are distinctly U.S. terms and concepts. Depending upon people's level of acculturation, they may or may not have even heard of these terms. Within the U.S. context, immigrants often refer to themselves in everyday life using their nationality, introducing

themselves as Mexican or Salvadoran. To further complicate this issue, the term "Spanish" is a nationality, which might lead respondents to think that the question is asking about nationality. On the whole, the way this question is worded is confusing for Hispanic respondents, particularly Spanish speakers.

We presume that most interviewers are able to successfully code the responses "I'm Hispanic" or "I'm Latino" as "yes" responses to this question. However, this assumption should not be so easily made when a respondent provides a nationality in response to the question. It is asking a great deal of interviewers to know the names of every Spanish-speaking country in the world, especially when this information is not part of their training. Even among the small number of cases in this study, there was evidence that an interviewer incorrectly input a "no" response into the instrument when a respondent said "I'm Mexican" in response to this question. The interaction (translated from Spanish) was as follows:

> Interviewer: "Is Juan of Hispanic, Latino or Spanish origin?"
> Respondent: "He's Mexican."
> Interviewer: "He's not of Hispanic, Latino or Spanish origin?"
> Respondent: No.
> Interviewer: No? Ok.

This respondent went on to list herself as "Mexican" in response to the Hispanic origin question as well and later answered that she was "Hispanic" in response to the race question. At that point, the interviewer chastised her for having answered "no" in response to the Hispanic origin question earlier in the interview. It was unclear whether the interviewer went back and changed the response in the instrument.

Without specialized training, interviewers cannot be expected to decide in all cases whether a respondent should be recorded as being Hispanic, when many respondents do not know these terms or refer to themselves in this way. If some interviewers are not aware that Mexico is a Spanish-speaking country, there will surely be others who do not know about countries with fewer immigrants in the U.S., such as Uruguay or Panama.

> ***Recommendations for Hispanic Origin***
> We recommend that further testing of this question be done with both Spanish and English-speaking Hispanic respondents to identify a better way to phrase this question to make the yes/no nature of the question more obvious and to avoid the potential for interpreting the question as referring to citizenship status. Cognitive interviewing would be an ideal method for this type of testing. Both rewording of the question and additional interviewer training are probably in order.

Race Question

The race question was the second in the three part series. The question was scripted as follows:

Question 14. Race

"What is [YOUR/NAME'S] race? (You may choose one or more races. For this census, Hispanic origins are not races. White or Caucasian, Black, African American or Negro, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Some other race?)

"¿Cuál es [su/la] raza [de NAME]? (Usted puede elegir una o más razas. Para este censo, origen hispano no es una raza. Blanca o Caucásica, Negra o africana americana, India americana o nativa de Alaska, Asiática, Nativa de Hawaii u otra de las islas del Pacífico, Alguna otra raza)?

Interviewer Behavior

Overall the race question was asked exactly as worded or with minor changes only 23 percent of the time (see Table 2a). Major changes were made 50 percent of the time. The question was asked as intended at similar rates in English and Spanish.

As previously noted, there was an effect of repeated administration on the correct administration of this question (see Table 3). For Person 1, the question was asked exactly as worded or with minor changes 34 percent of the time. For later household members, it was asked exactly as worded or with minor changes only 19 percent of the time, despite the fact that the two instructional sentences and the response categories only needed to be read for the first household member.

This question was incorrectly skipped very often–in fact, it was the second most frequently skipped question that we analyzed (see Table 2a). For Person 1, the race question was skipped less than 2 percent of the time, but for Person 2 and later in the household, it was skipped 30 percent of the time (data not shown in table). Presumably this indicates the tendency for the interviewer to input the same race for all household members either because the interviewer assumes that this is the case, or because the respondent told them so earlier in the interview.

Based on the notes provided by the behavior coders, we looked for commonalities in the way interviewers changed the question when there was a major change. We found several types of major changes as noted in Table 10. The most commonly made change to the wording of the race question occurred when interviewers rephrased the question to say "And how about Mary?" for persons higher than Person 1, rather than asking "What is Mary's race?" as they should have done. This accounted for 28 percent of the major changes made in English and 25 percent of the major changes made in Spanish.

Interviewers made a number of major changes to the question wording even when reading it for the first time in reference to Person 1. For example, in 16 percent of the major change cases overall, the interviewer omitted all of the response options for Person 1. This accounted for 17 percent of the major changes made in English and 13 percent of the major changes made in Spanish. In 15 percent of the English-language cases, interviewers omitted just some of the response options for Person 1. This did not occur in the Spanish-language cases examined here.

Table 10. Types of Major Changes to Wording of the Race Question

Interviewer Behavior, Major changes, N= 110	English N	English %	Spanish N	Spanish %	Total N	Total %
Person 1-specific problems						
1. Omits all response options for P1	15	17	3	13	18	16
2. Omits some response options for P1	13	15	0	0	13	12
3. Omits "Hispanic origins are not races" for P1	12	14	0	0	12	11
4. Omits the phrase "one or more" for P1	6	7	0	0	6	5
General problems						
5. Rephrases to "and how about NAME?"	24	28	6	25	30	27
6. Omits parts of response options, e.g. Negro, Caucasian	5	6	5	21	10	9
7. Asks about multiple residents in 1 question	1	1	3	13	4	4
8. Adds explanation of Question	1	1	2	8	3	3
9. Offers Hispanic or Latino as a race	2	2	0	0	2	2
10. Adds transition "here are the choices"	0	0	2	8	2	2
11. Rephrases Q after reading as worded	0	0	2	8	2	2
12. Repeats or emphasizes 1 response option	0	0	1	4	1	1
13. Other	7	8	0	0	7	6
Total	86	99	24	100	110	100

Note: In some cases an interviewer was coded as having made more than one major change to the question.
Due to rounding, all total percentages do not equal exactly 100.

However, in 6 percent of the English cases and 21 percent of the Spanish cases coded as having a major change, interviewers dropped some of the examples pertaining to a single response option. For example, rather than saying, "White or Caucasian," an interviewer would simply say "White" and rather than reading "Black, African American, or Negro" as worded, an interviewer read only "Black."

In 14 percent of the English cases, the interviewer omitted the instruction "For this census, Hispanic origins are not races." This instruction was not omitted in any of the Spanish cases. In 6 of the English cases, the interviewer omitted the instruction "You may choose one or more races." This did not occur in the Spanish-language cases recorded here.

There were a number of changes that were made less frequently. For example, in 3 Spanish-language cases and 1 English-language case, the interviewer asked about the race of multiple household members in one question, saying for example, "and all of your kids too?" rather than going through each household resident separately as instructed.

In 2 English-language cases, the interviewer offered "Hispanic" or "Latino" as response options, perhaps because he or she had experienced respondents offering this as a response to the question in past interviews.

There were a number of Spanish-language cases in which the interviewer read the question as worded and then immediately rephrased it or added an explanation of the reason for the question, saying, for example, "What race do you consider yourself?" or providing a definition for the term "race." One interviewer even added, "They want you to put white." This behavior may have been due to the fact that the interviewer had been met with confusion from Spanish-speaking respondents in the past and wanted to be helpful or to avoid a lengthy discussion. This type of situation occurred in 5 Spanish-language cases and in 1 English-language case.

Respondent Behavior

Respondents gave an adequate answer to the Race question on the first exchange 38 percent of the time (see Table 2a) and ended up with a codable answer in just over half the cases by the final exchange (see Table 2b). This means that out of the 59 cases where an inadequate response was initially provided, the interviewer only persisted to achieve an adequate response by the final outcome in 14 cases. From the behavior coding data, we cannot determine how the interviewer recorded the respondent's race in the instrument, but the fact that the interviewer did not actually gather a codable race from respondents almost half of the time indicates reason for concern. If the interviewer is choosing a race for the respondent, then it may not be not reflective of the race that the person considers himself or herself to be. If the interviewer enters "don't know" or "refuse," then we are left with missing data that will need to be imputed.

As previously noted, there was a significant effect of language on respondent behavior (see Table 4). In 54 percent of English interviews, respondents gave a codable response on the first exchange. However, this occurred in only 16 percent of interviews in Spanish. This is most likely due to the fact that a large percentage of the uncodable responses came about when a respondent provided some type of Hispanic origin as his/her answer to the question (see Table 11 which classifies inadequate responses).

Table 11. Types of Inadequate Responses to the Race Question

	English		Spanish		Total	
Respondent behavior: Inadequate, N=87	N	%	N	%	N	%
1. Providing nationality (Hispanic)	11	24	10	24	21	24
2. "Hispanic," "Latino," or "Chicano"	10	22	7	17	17	20
3. Replying only "the same thing"	7	16	5	12	12	14
4. Responding with Don't Know	5	11	4	10	9	10
5. Discussion in place of response	2	4	6	14	8	9
6. Describing oneself as "American"	1	2	3	7	4	5
7. Answering for multiple persons	3	7	1	2	4	5
8. Providing nationality (non-Hispanic)	2	4	0	0	2	2
9. Responds with other term e.g. Moreno/dark	0	0	1	2	1	1
10. Other	4	9	5	12	9	10
	45	99	42	100	87	100

Note: In some cases a respondent was coded as having given more than one type of inadequate response to the question. Due to rounding, all total percentages do not equal exactly 100.

The most striking finding here is that the vast majority of inadequate responses to this question were related to confusion on the part of Hispanic respondents. In 24 percent of the cases where a respondent did not provide an adequate response, he/she answered the race question with a nationality corresponding to a Hispanic origin, saying for example, “I’m Mexican.” This happened in equal numbers of English- and Spanish-language cases. (In 2 additional English-language cases, a respondent provided a non-Hispanic nationality, saying “I’m Lebanese” or “I’m Haitian.”)

In 17 additional cases, a respondent answered the question by saying either “I’m Hispanic,” “I’m Latino,” or “I’m Chicano.” This happened in 10 English-language cases and 7 Spanish-language cases. In 4 additional cases, Hispanic respondents seemed to interpret the race question as a question about U.S. citizenship and replied “I’m American.” One Spanish speaker responded that he was “moreno” or dark-skinned, apparently understanding that the race question refers to a physical characteristic but not finding a response option on the list that he felt described himself appropriately.

This question elicited a discussion rather than a response in 2 of the English-language cases and 6 of the Spanish-language cases. In all of these cases the respondent was confused by the question and did not know how to respond. These were all Hispanic respondents. One English-speaking respondent said “I don’t know” and went on to explain that he was Mexican American. A Spanish-speaking respondent asked “What do you put for Mexican Americans?” which introduced a discussion between himself and the interviewer. There were 5 English speakers and 4 Spanish speakers who responded with confusion or a flat “don’t know” response.

There were 3 English speakers and 1 Spanish speaker who responded about multiple household residents at once, saying for example, “I can’t answer that for them. There are too many choices,” referring to other household residents, or by saying “both of us” in response to the question about one person.

Finally, there were a number of respondents whose responses exhibited other difficulties, such as one woman who picked one response option but added Hispanic, saying “soy blanca hispana” (I’m white Hispanic). Another respondent replied only “I’m not white.” Yet another person, after hearing all of the choices replied “the first one you said” rather than repeating the response option. This might indicate that the first response option (White or Caucasian) are not the words that the respondent would normally use to describe him/herself, or that the list of categories was so long that the respondent could not remember the exact terms listed at the beginning.

On the whole, most of the difficulties discussed above were experienced by Hispanic respondents, both English and Spanish speakers. This is probably due in large part to the fact that the response options listed on the race question are not the terms that these respondents normally use to describe themselves. Judging from the patterns in response, Hispanic respondents typically think of themselves in terms of their nationality or Hispanic origin, rather than one of the race categories listed in this instrument. This point is made very clear by the respondent who asked an interviewer, “What do you put for Mexican Americans?”

Recommendations for Race

Since the Office of Management and Budget (OMB) mandates the approach federal agencies may use for collecting Hispanic origin and race data, we refrain from offering lengthy recommendations for question improvement. In the event that OMB regulations change and allow for greater flexibility for the wording of this question series, additional research into terms that Hispanic respondents use to describe themselves is warranted.

These data do suggest a number of issues that warrant further research. First of all, this question was originally designed for use on a paper questionnaire. When all of the response options are read aloud to a respondent, the question is extremely lengthy. For Hispanic respondents who do not hear an answer choice that includes a term they would use to describe themselves, this question can cause a great deal of confusion and discomfort.

Ancestry Question

The question on ancestry was the third in the three-part series and it was scripted as follows:

Question 15. Ancestry

"(People in the United States are from many countries, tribes, and cultural groups.) What is [YOUR/NAME'S] ancestry or tribe? (For example, Italian, African American, Dominican, Aleut, Jamaican, Chinese, Pakistani, Salvadoran, Rosebud Sioux, Nigerian, Samoan, Russian, etc.)"

"(Las personas de los Estados Unidos provienen de muchos países, tribus y grupos culturales.) ¿Cuál es [su] ascendencia o tribu? (Por ejemplo, italiana, africana americana, dominicana, aleuta, jamaicana, china, paquistaní, salvadoreña, Rosebud Sioux, nigeriana, samoana, rusa, etc.)

Interviewer behavior

Overall, the ancestry question was asked exactly as scripted or with minor changes only 23 percent of the time (see Table 2a). There was no significant difference by language. Interviewers made a major change to the question wording in 46 percent of the cases overall (see Table 2a).

The interviewer made a major change in question wording 72 percent of the time when reading the question for Person 1 (data not shown in table). For Person 1, the interviewer was required to read the introduction, the question and all of the examples. Clearly this was not an easy task for the interviewers. When reading the question for Persons 2 and higher, the interviewer was only required to read the question text, "What is [YOUR/NAME'S] ancestry or tribe?" In these cases, interviewers made a major change to the wording somewhat less frequently, in 50 percent of the cases. This rate is still extremely high and is cause for serious concern. The ancestry question was never incorrectly skipped for Person 1, but it was incorrectly skipped 25 percent of the time for later persons (data not shown in table). On the whole, this question was not administered as intended most of the time.

A review of the notes provided by the coders indicates that the most commonly occurring change in both English and Spanish was for interviewers to omit some of the examples that should have been read for Person 1 (see Table 12). This happened in a similar percentage of Spanish and English cases (26 percent and 23 percent of major change cases respectively). This most commonly occurred when interviewers started reading the question as worded but trailed off or were interrupted by the respondent after having read some of the examples. Many interviewers stopped after reading the examples "Dominican" or "Jamaican." Other interviewers stopped after reading various other examples. Some interviewers read some of the examples and skipped over others, omitting terms such as Aleut, Pakistani and Rosebud Sioux. In 11 percent of the cases, the interviewer omitted all examples when reading the question in reference to Person 1. In a smaller number of cases, the interviewer omitted the first sentence for Person 1 (People in the United States are from many countries tribes and cultural groups). This occurred in only 2 English cases and 1 Spanish case.

Table 12: Types of Major Changes to Wording of the Ancestry Question

	English		Spanish		Total	
Interviewer Behavior, Major changes, N= 124	N	%	N	%	N	%
1. Omits some of examples for P1	20	26	11	23	31	25
2. Rephrases to "and how about NAME?"	15	20	7	15	22	18
3. Omits all examples for P1	12	16	2	4	14	11
4. Asks "is NAME the same?"	5	7	6	13	11	9
5. Added to list of examples	6	8	4	8	10	8
6. Asks about multiple residents in 1 Q	4	5	4	8	8	6
7. Omits "tribe"	7	9	0	0	7	6
8. Adds an explanation of the Q	3	4	4	8	7	6
9. Leads R to choose a response	1	1	6	13	7	6
10. Omits first sentence for P1	2	3	1	2	3	2
11. Asked about a specific country	1	1	1	1	2	2
12. Other	0	0	2	4	2	2
Total	76	100	48	99	124	101

Note: In some cases an interviewer was coded as having made more than one major change to the question. Due to rounding, all total percentages do not equal exactly 100.

Another commonly occurring change to the question wording occurred when interviewers asked "And how about NAME?" rather than reading the question as worded for persons higher than Person 1. This happened in both Spanish and English cases, in 18 percent of cases overall. Similarly rather than reading the question as worded for persons higher than Person 1, some interviewers asked respondents "is NAME the same?"

Other interviewers added to the list of examples, most commonly including "Mexican," "Mexican American" or "Hispanic" as new examples. Interviewers also less frequently added examples of other Spanish-speaking countries such as Nicaragua, Panama and Honduras. In one case an interviewer added "African" as an example. This occurred in both English and Spanish cases and might be due to the fact that an interviewer already had information that the particular

respondent was of Mexican origin, based on his or her accent, based on the region in the U.S. in which he or she lived, or based on information supplied earlier in the interview, particularly in response to the Hispanic origin or race questions. In addition, interviewers may have been adding examples in response to having been met with confusion on the part of respondents in the past when they did not hear their own group named in such a long list of examples.

Interviewers made a number of other changes to the question wording, including asking about multiple residents in one question, saying for example, "Both of you?" or "For everyone?" Another change made in 9 percent of the English cases was omitting the word "tribe" from the initial sentence. In both Spanish and English cases the interviewer sometimes read the question as worded and then added an explanation at the end, saying for example, "it just means like where your family came from" or "Here is where we could put, for example, I could put Mexican American or Mexican." In 6 of the Spanish cases and 1 English case, the interviewer appeared to lead the respondent to choose a particular response, most commonly Mexican. In one case the interviewer encouraged a Spanish-speaking respondent to choose "American" to describe her child who had been born in the United States and another interviewer encouraged a respondent to say "Hispanic."

Respondent behavior

Respondents provided an adequate answer on the first exchange in 65 percent of cases for this question (see Table 2a). This did not differ significantly by language. Interviewers persisted and achieved an adequate answer as the final outcome in 79 percent of cases combining languages (see Table 2b).

On the whole, this question clearly caused confusion amongst respondents and was difficult for interviewers to administer exactly as worded. Also, the types of errors were generally different in English and Spanish interviews.

Coders did not provide a lot of notes on respondent behavior for this question, but it appears that the most common trend in inadequate initial responses was for Hispanic respondents to describe themselves as "American," essentially interpreting this as a question about citizenship or national origin (see Table 13). This happened in 5 Spanish-language cases and 2 English-language cases. Five other respondents described themselves as Hispanic or Latino in response to this question, apparently not understanding the difference between the intent of this question and that of the Hispanic origin question. Five Spanish-speaking respondents were unable to answer the question after first hearing it and expressed confusion. One of these respondents said "it's just that there are so many questions that one just doesn't…" Another respondent began discussing the city in Mexico where her father had been born. Other respondents said that they were "mixed." A small number of English-speaking respondents also said "none" in response to this question.

Table 13: Types of Inadequate Responses to the Ancestry Question

	English		Spanish		Total	
Respondent behavior: Inadequate, N=27	N	%	N	%	N	%
1. Describing oneself as "American"	2	13	5	45	7	26
2. Describing oneself as "Hispanic" or "Latino"	4	25	1	9	5	19
3. Confusion	0	0	5	45	5	19
4. Describing oneself as "Mixed"	3	19	0	0	3	11
5. Saying "none"	2	13	0	0	2	7
6. Discussion about residence in place of response	1	6	0	0	1	4
7. Describing oneself as "European"	1	6	0	0	1	3
8. Other	3	19	0	0	3	11
	16	101	11	99	27	100

Note: In some cases a respondent was coded as having given more than one type of inadequate response to the question. Due to rounding, all total percentages do not equal exactly 100.

Spanish-Specific Problem

There is an error in the translation of this question. The English version reads "People in the United States…" The Spanish version reads "People from the United States…" This type of discrepancy could have an impact on whether respondents interpret the question as asking only about native-born U.S. citizens or about whether it applies to immigrants as well.

Recommendations for the Ancestry question

- One problem evident in this question is that it is extremely long and difficult to administer in CAPI format. We recommend researching alternative ways to word this question. Cutting down on the number of examples provided would be a strong first step in simplifying the question.
- Since this question appears to cause greater difficulty and confusion for Hispanic and Spanish-speaking respondents than it does for English speakers and non-Hispanics, and since the largest Spanish-speaking group in the U.S. is of Mexican origin, we recommend including "Mexican" as one of the examples in this question.
- For greater ease of administration, we recommend testing this question using the "what about NAME?" format for persons higher than Person 1.
- While recognizing that the Office of Management and Budget (OMB) mandates the approach federal agencies may use for collecting Hispanic origin and race data, we recommend an examination of the entire 3-part series of Hispanic Origin, Race and Ancestry. We recommend testing to identify ways in which this question series might better capture the self identification of both Spanish- and English-speaking Hispanic respondents in the U.S. The current question series is not "working" for this group of respondents on the whole.

Overcount Question

The Overcount question is the last substantive question in the series of person-based questions and aims to determine if any enumerated individuals stayed at another place around the time of the census, in addition to collecting information about the reason for this behavior. It is interesting to note that, although interviewer behavior was very poor during the 2004 Census test (see Hunter and Landreth, 2005), this question was lengthened rather than shortened for 2006. It is now scripted as follows:

Question 16. Overcount

[Do you/Does NAME] sometimes live or stay somewhere else:
To attend college?
To be closer to work?
While in the military?
To stay at a seasonal or second residence?
For a child custody arrangement?
While in jail or prison?
While in a nursing home?
For another reason?

Algunas veces, ¿vive o se queda [usted/NAME] en algún otro lugar?
Para asistir a la universidad?
Para estar más cerca del trabajo?
Mientras está en el servicio militar?
Para quedarse en una residencia [estacional/de temporada] o una segunda residencia?
Por un arreglo de custodia de niños?
Mientras está en la cárcel o prisión?
Mientras está en un hogar de ancianos?
Por alguna otra razón?

For this question, the interviewer was required to read all response options for the first person, and then read them as necessary for later people. Only the question stem was required to be read in asking about later members of the household.

Interviewer Behavior

This question had a very poor exact reading rate. It was only administered exactly as worded or with minor changes 23 percent of the time (see Table 2a). In 59 percent of the cases, major changes to question wording were made and 19 percent of the time the question was skipped. There was not a significant effect of language on question administration behavior and behavior was similar in English and Spanish cases.

Though there was not a significant difference by repeated question administration, required interviewer behavior was different for Person 1 than for other persons in the household. For this reason, we will examine the types of major-change behavior separately.

For Person 1, the interviewer was required to read all response options. This was problematic, as evidenced by Table 14.

Table 14. Types of Major Changes of the Wording to Overcount for Person 1

Type of Major Change	Observations N=49	Percent of MC for P1
Omitted all response options	10	20%
Read only 1 or 2 response options	11	22%
Read only 3 or 4 response options	3	6%
Read more than 4, but less than 8	6	12%
Changed some of response option text	17	35%

NOTE: These numbers are not intended to sum to 100 percent. There are additional ways to make a major change, including omitting words of the question stem and creating a household level question. Additionally, the last category in this table "Changed some of response option text" is not mutually exclusive of 3 of the first 4 categories.

While interviewers only left out all response options in 20 percent of major-change cases for Person 1, in an additional 40 percent of major-change cases, they left out some of the response categories. In 35 percent of all major change cases, interviewers modified the response category text. This indicates that interviewers felt they could not administer the question in its entirety, possibly due to the fact that it is a very long question and that many response options would not apply to most respondents.

Though we cannot know for certain why interviewers made these changes, perhaps they selected the response options based upon what seemed to make the most sense within the context of the interview (e.g., "You said she went to school…Did she stay somewhere else for school?"). All response options are intended to be asked of all household members, regardless of whether or not they apply to people of different ages. For example, the interview is scripted to ask of all people, even children and the elderly, if they attended college and to ask everyone, even adults, if they were in child custody arrangements. Interviewers may have tailored the question by withholding response options they felt did not apply to particular household members based upon age or other characteristics. It is also possible that interviewers were treating the response options as illustrative examples that they could choose from in order to shorten the question, providing only the first one or two response options as examples.

For Persons 2 and higher, by far the most common major-change behavior was shortening the question to "How about NAME?" In 72 percent of major-change cases for Persons 2 and higher, this was the case (data not shown).

Other problems included making this into a household-level question (or at least asking about 2 or more people with a single question), which occurred in about 10 percent of all major-change cases.

Paraphrasing was also evident within the question stem. Interviewers occasionally dropped the word "sometimes" and also omitted the word "live." Modifications to the stem of the question, however, may be less problematic for data quality than is the omission of some or all of the response options.

Respondent Behavior

As was evident for some of the other questions, despite poor interviewer behavior, respondents were able to give adequate answers 84 percent of the time in response to this question (see Table 2a). Again, this does not mean that respondents would have answered the same way if they had been presented with all response options.

Spanish -Specific Problem

In Spanish the question, when read as worded, is asking whether someone lives or stays somewhere else presently and whether they did so in the past. The instrument is in effect asking Spanish-speaking respondents to think of both the present and past but with no specific reference period. The English question only asks about the present time. This will most likely cause a differing interpretation and response on the part of respondents.

This question also contains two translations for the phrase "seasonal residence," listing both the terms "estacional" and "de temporada." See the discussion above on the Usual Residence question for a more detailed discussion of this issue.

Recommendations for Overcount

- Create a series of questions for interviewer-administered instruments that combine information collected in this question with information collected and presented in the residence rules and undercoverage questions. See Childs et al. (2007) for an example of how this might be done.
- Use a flashcard for interviewer-administered in-person modes if the full response set is maintained.
- Consider using age filters to probe about only age-relevant possibilities.
- Correct Spanish translation problems to make the question equivalent to the English version.

Part IV CONCLUSION

Successful administration of survey questions using standardized interviewing techniques helps reduce non-sampling error in data collection. When interviewers deviate from scripted questions, respondents are exposed to non-standardized question stimuli, potentially influencing responses in unintended ways, resulting in cause for concern over data quality. Despite changes that were made in the automated NRFU instrument between 2004 and 2006, the results from this behavior coding suggest that almost all of the survey questions continue to suffer from administration difficulties at surprisingly high rates. Interviewers tended to change the survey questions, sometimes quite dramatically, or omit them entirely. Omissions seemed most problematic for questions repeated for each person in the household. It is evident from these results that interviewers are changing the wording and intent of the questions without the benefit of subject matter or survey methodology expertise. Although topic-based structure was used for this instrument, the script was not designed to take advantage of the benefits that topic-based administrations allows. One of the reasons topic-based administration is preferred is that after reading the question in full for the first person, it can be shortened for later people (using the "How about NAME?" structure). Interviewers frequently did this on their own, but not in a standardized fashion. If we script the interview in a topic-based-friendly way, then interviewers will know when it is appropriate to shorten questions and when it is not. We understand that the 2008 NRFU instrument will employ the "How about NAME?" structure and we look forward to analyzing behavior coding data from that operation.

Additionally, we understand that the revised question structure for the residence rules and coverage questions recommended in Childs et al. (2007) has been adopted for the 2008 NRFU. We also look forward to behavior coding a sample of those interviews to assess whether the new, recommended approach is applied in a more consistent manner by interviewers.

Response issues were less dramatic across the board, but did tend to plague particular questions, most notably the Relationship and Race questions. Respondents are also having difficulty providing adequate answers to some of the questions. In order to resolve these issues, we understand the 2008 NRFU instrument will employ a single relationship question, with a follow-up probe for children in the telephone version of the instrument. Revision and further pretesting efforts of the Race question are warranted. Iterative rounds of cognitive interviewing are recommended for future versions of this question.

These behavior coding results make clear that the Spanish version of the questions did not function as well as its English counterpart, which suggests they also need further revision and pretesting. Prior to any cognitive pretesting, though, further consideration should be given to the translation used for the Spanish-language questions. This type of review may turn up inconsistencies in conceptual equivalence that may be causing some of the administration and response difficulties.

Additionally, and more critically, the review of the Spanish question text used in the 2006 instrument revealed errors in programming that caused interviewers to be shown grammatically incorrect questions to read to respondents (see Appendix C for all of the errors). Beyond all other

recommendations contained in this report, we recommend that ***both*** the English and the Spanish versions of the 2008 and 2010 NRFU be carefully proofread to avoid the embarrassment of the 2006 Spanish question text.

Tables 2a and 2b: Complete Behavior Coding Results

Table 2a. Percent Interviewer and Respondent Behavior for All Interviews (English and Spanish) by Question

	Interviewer Behavior[1]								Respondent Behavior[2]								Break In[3]
Question	N	E/S	MC	V+	V-	I/O	IS	CS	N	AA	IA	QA	CL	DK	R	I/O	
1 Live Here	72	51	40	0	0	1	7	0	67	75	9	3	6	0	0	8	3
2 Usual Res	71	51	42	0	0	0	7	0	67	75	21	1	0	0	0	3	4
3 POP count	71	52	45	0	0	0	3	0	69	94	3	1	0	0	0	1	4
4 Names	72	35	64	0	0	0	1	0	71	93	6	0	1	0	0	0	4
5 Undercount	72	35	64	0	0	0	1	0	71	80	3	10	3	0	0	4	9
6 Ownership	72	28	64	1	0	0	7	0	67	70	19	4	1	0	0	4	7
7 Relationship1	152	58	14	5	1	0	12	11	118	53	38	0	3	1	0	6	2
8 Relationship2	121	12	22	3	0	0	52	11	56	59	36	0	0	0	0	5	0
9 Sex	209	38	25	25	1	0	11	0	184	79	3	0	2	0	0	16	0
10 DOB	206	65	29	0	0	0	6	0	196	80	6	4	2	8	0	1	0
11 Age	204	5	8	1	0	0	5	80	29	76	3	10	0	10	0	0	0
12 Confirm Age	206	28	52	0	0	0	20	0	166	83	6	2	1	0	0	8	0
13 Hisp	203	50	29	0	1	0	20	0	162	69	17	4	3	1	0	7	0
14 Race	202	23	50	1	0	0	27	0	150	38	39	3	5	3	0	12	5
15 Ancestry	202	23	46	1	5	1	24	0	154	65	8	5	4	5	0	14	3
16 Overcount	204	23	59	0	0	0	19	0	165	84	8	1	1	1	0	4	7

[1] E/S = exact/slight; MC = major change; V+ = correct verification; V- = incorrect verification; and I/O = inaudible or "other"; IS = incorrect skip; CS = correct skip

[2] AA = adequate answer; IA = inadequate answer; QA = qualified answer; CL = clarification requested; DK = don't know; R = refusal; and I/O = inaudible or "other."

[3] Break-In = respondent interrupted the interviewer while during the administration of the question. Denominator taken from the Respondent Behavior N, but is calculated separately from the respondent behaviors.

Table 2b. Percent Final Response Outcome for All Interviews (English and Spanish) by Question

	Final Response Outcome[1]						
Question	N	AA	IA	QA	DK	R	I/O
1 Live Here	67	85	7	1	0	0	6
2 Usual Res	67	82	13	0	0	0	4
3 POP count	69	100	0	0	0	0	0
4 Names	71	100	0	0	0	0	0
5 Undercount	71	93	3	1	0	0	3
6 Ownership	67	88	9	0	0	0	3
7 Relationship1	118	56	39	0	0	0	5
8 Relationship2	56	68	29	0	0	0	4
9 Sex	184	82	4	0	0	0	15
10 DOB	196	88	3	1	6	2	1
11 Age	29	79	0	10	10	0	0
12 Confirm Age	166	89	2	1	0	0	8
13 Hisp	162	73	14	4	1	0	8
14 Race	150	55	30	3	3	0	9
15 Ancestry	154	79	5	3	3	0	10
16 Overcount	166	90	5	0	1	0	4

[1] AA = adequate answer; IA = inadequate answer; QA = qualified answer; DK = don't know; R = refusal; I/O = inaudible or "other."

Table 3. Results of Logistic Regression by Repeated Question Administration: Percent Good Behavior[1] by Repeated Question Administration

	Good Interviewer Behavior		Good Respondent Behavior		Good Final Response Outcome	
Question	1st admin	Later admin	1st admin	Later admin	1st admin	Later admin
7 Relationship1[2]	81	62	63	51	65	55
8 Relationship2[2]	24	12	70	54	75	65
9 Sex	78	56	89	96	95	96
10 DOB	85*	55	85	78	93	86
11 Age	56	26	78	75	78	80
12 Confirm Age	44*	19	90	90	97	97
13 Hispanic	92*	28	74	74	85	76
14 Race	34	18	41	45	65	56
15 Ancestry	27	23	71	80	83	93
16 Overcount	28	20	89	87	95	93
Overall†	50*	32	79	77	89	83

*The percentages listed are data-driven rather than model-based. Associated regression coefficient significant at the $p < .002$ level. [3]

† Overall significance was determined by fitting the model to the data as a whole, rather than per question.

[1] Good Interviewer Behavior was defined as exact wording/slight change (Code E/S) and correct verification (Code V+). The correctly skipped category (Code CS) was omitted from the analysis because it is neither "good" nor "bad" behavior. The only behavior that was considered good respondent behavior was an adequate answer (Code AA). Inaudible respondent and outcome behaviors (Code I/O) were removed from this analysis because they are neither "good" nor "bad" behaviors.

[2] For the relationship questions, the first administration is for Person 2 and later administrations are Person 3+. This is due to the fact that there is no relationship between Person 1 and him or herself.

[3] We conducted a total of 48 tests (16 questions and 3 dependent measures). To ensure a study-wide significance level of .10, we recommend using a Bonferroni adjustment, which lead to a significance level of $p< .002$, which is what we used as a guideline for interpreting results (see http://home.clara.net/sisa/bonfer.htm to replicate this adjustment).

[4] The results of the statistical tests performed for this study are intended to be used for heuristic purposes only. The tests were performed as if the data were collected in a simple random sample, without replacement, which was the case in this study.

Table 4. Results of Logistic Regression by Language: Percent Good Behavior[1] by Language

	Good interviewer behavior		Good respondent behavior		Good Final Response Outcome	
Question	English	Spanish	English	Spanish	English	Spanish
1 Live Here	59	28	85	64	90	93
2 Usual Res	55	39	79	72	91	72
3 POP count	55	41	96	94	100	100
4 Names	42	16	98	79	100	100
5 Undercount	35	33	86	76	96	94
6 Ownership	30	28	78	65	96	76
7 Relationship1	75	64	76*	29	78*	33
8 Relationship2	25	5	72	25	77	50
9 Sex	68	57	93	94	95	96
10 DOB	74	49	81	79	86	93
11 Age	41	9	73	100	77	100
12 Confirm Age	35	16	93	83	98	94
13 Hispanic	61	33	82*	57	90*	57
14 Race	24	23	54*	16	69*	39
15 Ancestry	28	19	74	79	86	93
16 Overcount	25	18	91	82	95	92
Overall	46*	31	82*	69	89*	79

* The percentages listed are data-driven rather than model-based. Associated regression coefficient significant at the $p < .002$ level. [2]

[†] Overall significance was determined by fitting the model to the data as a whole, rather than per question.

[1] Good Interviewer Behavior was defined as exact wording/slight change (Code E/S) and correct verification (Code V+). The correctly skipped category (Code CS) was omitted from the analysis because it is neither "good" nor "bad" behavior.The only good respondent behavior and outcome is an adequate answer (Code AA). Inaudible respondent and outcome behaviors (Code IO) were removed from this analysis because they are neither "good" nor "bad".

[2] We conducted a total of 48 tests (16 questions and 3 dependent measures). To ensure a study-wide significance level of .10, we recommend using a Bonferroni adjustment, which lead to a significance level of $p< .002$, which is what we used as a guideline for interpreting results (see http://home.clara.net/sisa/bonfer.htm).

[3] The results of the statistical tests performed for this study are intended to be used for heuristic purposes only. The tests were performed as if the data were collected in a simple random sample, without replacement, which was not the case in this study.

REFERENCES

Burgess, M. and Paton, D. (1993) "Coding of Respondent Behaviours by Interviewers to Test Questionnaire Wording." Proceedings of the Survey Methods Research Section. Alexandria, VA: American Statistical Association.

Cannell, C., Fowler, F., and Marquis, K. (1968) "The Influence of Interviewer and Respondent Psychological and Behavioral Variables on Reporting in Household Interviews." Vital and Health Statistics, Series 2 (26). Washington, DC: US Government Printing Office.

Caspar, R., Goerman, P., McAvinchey, G., Quiroz, R., and Sha, M. (2006) "Census Bilingual Questionnaire Research: Final Round 1 Report." Contract interim project report submitted to the United States Census Bureau.

Caspar, R., Goerman, P., McAvinchey, G., Quiroz, R., and Sha, M. (forthcoming) "Census Bilingual Questionnaire Research: Final Report." Contract project report submitted to the United States Census Bureau.

Chan, A. (2007) "Report on Topic-based Mode Consistency CATI/CAPI Pretesting Study for the American Community Survey (ACS)." Statistical Research Division Research Report Series, *Survey Methodology #2007-10.*

Childs, J. H., Gerber, E. R., Carter, G., and Beck, J. (2006) Cognitive Test of the 2006 NRFU: Round 2. SRD Research Report Series SSM2006/05.
http://www.census.gov/srd/papers/pdf/ssm2006-05.pdf

Childs, J. H., Carter, III., G. R., Norris, D., Hanaoka, K., and Schwede, L. (2007) Cognitive Test of the NRFU Round 3: Revised Questions. SRD Research Report Series SSM2007/09.
http://www.census.gov/srd/papers/pdf/rsm2007-09.pdf

U.S. Census Bureau. (2006) Mode Consistency Guidelines. Draft. May 3. 2006. Internal unpublished document.

Edwards, S. , Zahnd, E. , Willis, G. , Grant, D. , Lordi, N. and Fry, S. (2004) Behavior Coding across Multiple Languages: The 2003 California Health Interview Survey as a Case Study. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, Arizona.

Esposito, J., Rothgeb, J., and Campanelli, P. (1994) "The Utility and Flexibility of Behavior Coding as a Methodology for Evaluating Questionnaires." Paper presented at the American Association for Public Opinion Research Annual Meeting, Danvers, MA.

Fleiss, J. (1981) Statistical Methods for Rates and Proportions. John Wiley & Sons, New York.

Fowler, F. (1992). "How Unclear Terms Affect Survey Data." *Public Opinion Quarterly, 56*: 218-231.

Fowler, F. and Cannell, C. (1996) “Using Behavioral Coding to Identify Cognitive Problems with Survey Questions.” In N. Schwarz and S. Sudman (Eds.), Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco, CA: Jossey-Bass.

Goerman, P.L. (forthcoming) “Cognitive Interview Methodology Revisited: Development of Best Practices for Pretesting Spanish Survey Instruments.” Report from postdoctoral research study based on the CAPI version of the Spanish American Community Survey.

Hunter, J. (2005) Cognitive Test of the 2006 NRFU: Round 1. SRD Research Report Series SSM2005/07.http://www.census.gov/srd/papers/pdf/ssm2005-07.pdf

Hunter, J. and Landreth, A. (2005) Behavior Coding Analysis Report: Evaluating Bilingual Versions of the Non-Response Follow-Up (NRFU) for the 2004 Census Test. Statistical Research Division Study Series: Survey Methodology #2006-7.

Love, K., and Byrne, R. (2005). *Investigating Measurement Errors Due to Interviewer Misclassification and Mode.* U.S. Census Bureau, DSSD 2010 Census Coverage Measurement Memorandum Series #2006-D7-03, Unpublished internal report, May 23, 2005.

Oksenberg, L., Cannell, C., and Kalton, G. (1991) “New Strategies for Pretesting Survey Questions.” *Journal of Official Statistics* 7 (3): 349-394.

Rappaport, M., Davis, D., and Allen, S. (2006) Final Report on an Observational Study of Census Non-response Follow-up Interviews during the 2006 Census Test in Travis Conty, Texas Spring and Summer 2006. Report submitted by Development Associates to the Census Bureau, December 4, 2006.

Smiley, R. and Keeley, C. (1997) “Behavior Coding of the ICM Computer Assisted Personal Interview.” 1996 Community Census Results Memorandum Number 23. U.S. Census Bureau, Washington, DC.

Sykes, W. and Morton-Williams, J. (1987) “Evaluating Survey Questions.” *Journal of Official Statistics* 3 (2): 191-207.

APPENDIX A:

SURVEY QUESTION GUIDE

NRFU 2006

There are three conventions used in the NRFU Question Guide to help you (coders) determine whether a question was administered and answered correctly, and those are:

Text in **bold** -- Interviewers should read text appearing in bold. This includes questions as well as response categories, where it is appropriate to read them as part of the question.

Text in (parentheses) -- Interviewers must read this text for the first person in the household, but it is optional reading for later household members.

Text in [brackets] -- Interviewers must “fill” a question correctly when text appears in brackets. Interviewers should not read all of the bracketed text at once, but choose the appropriate word or phrase within it for certain questions.

These questions will be asked once for each household.

Question 1	LIVE HERE
Screen	**Did you or anyone in your household live at [Address] on April 1, 2006?** **¿Vivía usted u otro miembro de su hogar en *[ADDRESS]* el 1 de abril de 2006?**
Acceptable Responses	Yes Sí No No
Comments	*Interviewer should read a street address for [ADDRESS].* *Key phrases –* ***English: "****you or anyone in your household"; "[ADDRESS]"; and "April 1, 2006"* ***Spanish:*** *"usted u otro miembro de su hogar"; "[ADDRESS]"; and"el 1 de abril de 2006"*
Skip	If yes, go to Q2 (USUAL RES.) If no, conduct interview for the address at which person lived on April 1, 2006; go to Q2 (USUAL RES).

Question 2	USUAL RES.
Screen	**Is this [house/apartment/mobile home] a vacation home, seasonal residence, held for occasional use, or does someone in this household usually live here?** **¿Es [esta/este] [casa/apartamento/casa móvil] una vivienda para vacaciones, una vivienda [estacional/de temporada] o una vivienda que se mantiene para uso ocasional, o vive aquí usualmente algún miembro de este hogar?**
Acceptable Responses	Vacation home — una vivienda para vacaciones Seasonal residence — una vivienda [estacional/de temporada] Held for occasional use — una vivienda que se mantiene para uso ocasional Someone in the household usually lives here — vive aquí usualmente algún miembro de este hogar
Comments	*Key phrases –* ***English and Spanish:*** *<u>All</u> response options must be read <u>as worded</u>*
Skip	Go to Q3 (POP COUNT).

Question 3	**POP COUNT**
Screen	**The census must count every person living in the United States on April 1, 2006. We want to count people where they usually live and sleep. For people with more than one place to live, this is the place where they sleep most of the time. How many people were living or staying in this [house/apartment/mobile home] on April 1, 2006?** **El censo tiene que contar a todas las personas que viven en los Estados Unidos el 1 de abril del 2006. Queremos contar a las personas en el lugar donde viven y duermen habitualmente. Para las personas que tienen más de un lugar para vivir, éste sería el lugar donde duermen la mayor parte del tiempo. ¿Cuántas personas estaban viviendo o quedándose en [esta/este] [casa/apartamento/casa móvil] el 1 de abril de 2006?**
Acceptable Responses	Number 1-49
Comments	*Key phrases-* ***English:*** *"April 1, 2006";" usually live and sleep"; "sleep most of the time"; and "living or staying at this [h/a/mh]"* ***Spanish:*** *"el 1 de abril del 2006"; "viven y duermen habitualmente"; "duermen la mayor parte del tiempo"; and "estaban viviendo o quedándose en esta [c/a/cm]"*
Skip	Go to Q4 (NAMES).
Question 4	**NAMES**
Screen	**What is the name of each person who lived or stayed at this [house/apartment/mobile home] on April 1, 2006? Start with the name of one person who owned or rented this [house/apartment/mobile home].** **¿Cómo se llama cada una de las personas que vivía o se quedaba en esta [casa/apartamento/casa móvil] el 1 de abril de 2006? Comience por el nombre de la persona que era [el/la] [propietario/propietaria] o quien alquilaba esta [casa/apartamento/casa móvil].**
Acceptable Responses	First name, (middle initial), Last name
Comments	*R should give first and last name. Middle initial and suffix (such as Jr.) are acceptable but not required for the answer to be acceptable. If R only gives first name, it is an "IA." In Spanish, 1 or 2 nombres or apellidos are acceptable answers.* *In Spanish cases, please note any discussion/confusion (on the part of the I or R) about whether R should give 1 or 2 names or last names and any discussion/confusion related to middle initial.* *Key phrases –* ***English: "****lived or stayed at this [h/a/mh]"; "April 1, 2006"; and "person who owned or rented this [h/a/mh]"* ***Spanish:*** *"vivía o se quedaba en esta [c/a/cm]"; "el 1 de abril de 2006"; and "la persona que era [el(la)] propietario(a) o quien alquilaba esta [c/a/cm]"*
Skip	Go to Q5 (UNDERCOUNT).

**There may be an intervening question or two here - one to reconcile number of people listed versus number of people given at the POP Count Question and another to verify the respondent.

Question 5	UNDERCOUNT
Screen	**We do not want to miss any people who might have been staying here on April 1, 2006. Were there any additional people staying here that you did not include, for example: Children, such as newborn babies or foster children? Relatives, such as adult children, cousins, or in-laws? Nonrelatives, such as roommates or live-in baby sitters? People staying here temporarily?** **No queremos dejar de contar a ninguna persona que pudiera haber estado quedándose aquí el 1 de abril de 2006. ¿Hubo alguna persona adicional que se quedaba aquí a la cual usted no incluyó, por ejemplo: Niños, tales como recién nacidos o hijos de crianza? Parientes, tales como jóvenes adultos, primos o parientes políticos? Personas no parientes, tales como compañeros[as] de cuarto o niñeras que viven en el hogar? Personas que se quedaban aquí temporalmente**
Acceptable Responses	Yes Sí No No
Comments	*Note any comments here about R's reconsidering their answer to the previous question, like, "I should have said 5 people, not 4."* *Key phrases –* ***English:** "April 1, 2006"; "people staying here that you did not include"; and all response options must be read as worded* ***Spanish:** "el 1 de abril de 2006"; "alguna persona adicional que se quedaba aquí a la cual usted no incluyó"; and all response options must be read as worded*
Skip	Go to Q6 (OWNERSHIP.)

**There may be an intervening question here. If R said "yes" to the above question, interviewer will ask for names of those people.

Question 6	OWNERSHIP
Screen	**Is this [house/apartment/mobile home]** **Owned by you or someone in this household with a mortgage or loan?** **Owned by you or someone in this household free and clear?** **Rented for cash rent?** **Occupied without payment of cash rent?** **¿Es [esta/este] [casa/apartamento/casa móvil]** **Propiedad suya o de alguien en este hogar con una hipoteca o préstamo?** **Propiedad suya o de alguien en este hogar libre y sin deudas?** **Alquilada por pago de dinero en efectivo?** **Ocupada sin pago de dinero en efectivo?**
Acceptable Responses	Owned with a mortgage or loan Propiedad con una hipoteca o préstamo Owned free and clear Propiedad libre y sin deudas Rented for cash rent Alquilada por pago de dinero en efectivo Occupied without payment Ocupada sin pago de dinero en efectivo
Comments	*Key phrases –* ***English and Spanish:*** *<u>All</u> response options must be read <u>as worded</u>*
Skip	Go to Q7 (RELATIONSHIP1).

Starting here, all questions are asked for each person.

Question 7	RELATIONSHIP 1
Screen	**[Are you/is name] related to [NAME]?** **¿ Está [NAME/usted] relacionado(a) con [NAME/usted]**
Acceptable Responses	Yes Sí No No
Comments	*This question will be skipped for the first person in the household.* *Key phrases –* ***English:*** *"related"* ***Spanish:*** *"relacionado(a) con"*
Skip	Go to Q8 (RELATIONSHIP2).

Question 8	RELATIONSHIP 2
Screen	**Which one of these categories best describes how [YOU ARE/NAME IS] related to [NAME]?** **OR** **Which one of these categories best describes [YOUR/NAME'S] relationship to [NAME]?** **¿Cuál de estas categorías es la que mejor describe cómo está [usted /NAME] relacionado(a) con [usted/NAME]?** **OR** **¿Cuál de estas categorías es la que mejor describe la relación [de NAME] con [usted/NAME]?**
Acceptable Responses	Husband or wife Esposo o esposa Biological son or daughter Hijo o hija biológico(a) Adopted son or daughter Hijo o hija adoptivo(a) Stepson or stepdaughter Hijastro o hijastra Brother or sister Hermano o hermana Father or mother Padre o madre Grandchild Nieto(a) Parent-in-law Suegro(a) Son-in-law or daughter-in-law Yerno o nuera Other relative Otro parentesco Roomer, boarder Inquilino o pupilo Housemate, roommate Compañero de casa o cuarto Unmarried partner Compañero(a) no casado(a) Foster child or foster adult Hijo(a) de crianza o adulto bajo custodia Other nonrelative Otro no pariente
Comments	*This question will be skipped for the first person in the household.* *Key phrases –* ***English:*** *"one of these categories best describes"; and "related to" OR "relationship to"* ***Spanish:*** *"Cuál de estas categorías es la que mejor describe"; and "relacionado(a) con" OR "la relación con"*
Skip	Go to Q9 (SEX).

Question 9	SEX
Screen	**[Is NAME/ARE YOU] male or female?** **¿Es [NAME/usted] de sexo masculino o femenino?**
Acceptable Responses	Male Masculino Female Femenino
Comments	*This question should be explicitly asked for each person.* *Key phrases –* ***English:*** *"male" and "female"* ***Spanish:*** *"masculino" and "femenino"*
Skip	Go to Q10 (DOB).

Question 10	DOB
Screen	**What is [your/NAME'S] date of birth?** **¿Cuál es [su] fecha de nacimiento?** **OR** **¿Cuál es la fecha de nacimiento [de NAME]?**
Acceptable Responses	Month/Day/Year
Comments	*Full date of birth is required for an "AA." Month and day only is "IA." Please note any discussion/confusion about the order of month/day/year, this may happen more often for Spanish speakers* *Key phrases –* ***English:*** *"date of birth"* ***Spanish:*** *"Fecha de nacimiento"*
Skip	If answer is DK or REF, go to Q11 (AGE), else skip to Q12 (AGECHECK).

Question 11	AGE
Screen	**What was [YOUR/NAME'S] age on April 1, 2006?** **¿Cuál era la edad [de NAME] el 1 de abril del 2006?** **¿Cuál era su edad el 1 de abril del 2006?**
Acceptable Responses	Number 1-125
Comments	*This will only be asked if date of birth is DK or REF. This will be skipped for most Rs.* *Key phrases –* ***English:*** *"age"; and "April 1, 2006"* ***Spanish:*** *"edad"; and "el 1 de abril del 2006"*
Skip	Go to Q12.

Question 12	AGECHECK
Screen	**For the census, we need to record age as of April 1, 2006. So, just to confirm - [NAME was/you were] [AGE] on April 1, 2006?** **Para el Censo, tenemos que anotar la edad que se tiene el 1 de abril de 2006. ¿Por lo tanto, solo para confirmar tenía [usted/nombre] [age] años el 1 de abril del 2006?**
Acceptable Responses	Yes Sí No No
Comments	*This is asked of all respondents.* *Key phrases –* ***English:*** *"For the census"; "age as of April 1, 2006"; and "[age] on April 1, 2006"* ***Spanish:*** *"Para el Censo"; "la edad que se tiene al 1 de abril de 2006"; and "[age] el 1 de abril del 2006"*
Skip	Go to Q13 (HISP)

**There may be intervening questions here if age is incorrect.

Question 13	HISP
Screen	**[ARE YOU/IS NAME] of Hispanic, Latino, or Spanish origin?** **¿Es [usted/ NAME] de orígen hispano, latino o español?**
Acceptable Responses	Yes Sí Hispanic Hispano Latino Latino Spanish origin Español No No
Comments	*IN ALL CASES, WRITE VERBATIM RESPONSE IN NOTES.* *Key phrases –* ***English:*** *"Hispanic, Latino, or Spanish origin"* ***Spanish:*** *"origen hispano, latino o español"*
Skip	Go to Q14 (RACE)

Question 14	**RACE**
Screen	**What is [YOUR/NAME'S] race? (You may choose one or more races. For this census, Hispanic origins are not races. White or Caucasian, Black, African American or Negro, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Some other race)** **¿Cuál es [su/la] raza [de NAME]? (Usted puede elegir una o más razas. Para este censo, orígen hispano no es una raza. Blanca o Caucásica, Negra o africana americana, India americana o nativa de Alaska, Asiática, Nativa de Hawaii u otra de las islas del Pacífico, Alguna otra raza)**
Acceptable Responses	White or Caucasian Blanca o Caucásica Black, African American, or Negro Negra o africana americana American Indian or Alaska Native India americana o nativa de Alaska Asian Asiática Native Hawaiian or Other Pacific Islander Nativa de Hawaii u otra de las islas del Pacífico Some other race Alguna otra raza
Comments	*IN ALL CASES, WRITE VERBATIM RESPONSE IN NOTES.* *Sentences and response options in parentheses are required reading for the first person and optional for other persons.* *Rs may provide more than one answer*. *Key phrases –* ***English: "****race" (For Person 1 - "one or more races"; "Hispanic origins are not races"; and <u>all</u> response options must be read <u>as worded</u>)* ***Spanish:** "raza" (For Person 1 - "una o más razas"; "orígen hispano no es una raza"; and <u>all</u> response options must be read <u>as worded</u>)*
Skip	Go to Q15 (ANCESTRY)

Question 15	ANCESTRY
Screen	**(People in the United States are from many countries, tribes, and cultural groups.) What is [YOUR/NAME'S] ancestry or tribe? (For example, Italian, African American, Dominican, Aleut, Jamaican, Chinese, Pakistani, Salvadoran, Rosebud Sioux, Nigerian, Samoan, Russian, etc.)** **(Las personas de los Estados Unidos provienen de muchos países, tribus y grupos culturales.) ¿Cuál es [su] ascendencia o tribu? (Por ejemplo, italiana, africana americana, dominicana, aleuta, jamaicana, china, paquistaní, salvadoreña, Rosebud Sioux, nigeriana, samoana, rusa, etc.)**
Acceptable Responses	*Use your discretion for adequate answers. Inadequate includes things like "human," "Martian," or "Californian."*
Comments	*IN ALL CASES, WRITE VERBATIM RESPONSE IN NOTES.* *Sentences in parentheses are required reading for the first person and optional for other persons.* *Key phrases –* ***English:** "ancestry or tribe" (For person 1: "many countries, tribes and cultural groups";* *and <u>all</u> the examples must be read <u>as worded</u>)* ***Spanish:** "ascendencia o tribu" (For person 1: " muchos países, tribus y grupos culturales";* *and <u>all</u> the examples must be read <u>as worded</u>)*
Skip	Go to Q16 (OVERCOUNT)

Question 16	OVERCOUNT
Screen	**[Do you/Does NAME] sometimes live or stay somewhere else:** **(To attend college? To be closer to work? While in the military? To stay at a seasonal or second residence? For a child custody arrangement? While in jail or prison? While in a nursing home? For another reason?)** **Algunas veces, ¿vive o se queda [usted/NAME] en algún otro lugar?** **(Para asistir a la universidad? Para estar más cerca del trabajo? Mientras está en el servicio militar? Para quedarse en una residencia [estacional/de temporada] o una segunda residencia? Por un arreglo de custodia de niños? Mientras está en la cárcel o prisión? Mientras está en un hogar de ancianos? Por alguna otra razón?)**
Acceptable Responses	Yes, To attend college — Sí, Para asistir a la universidad Yes, To be closer to work — Sí, Para estar más cerca del trabajo Yes, While in the military — Sí, Mientras está/estaba en el servicio militar Yes, To stay at a seasonal or second residence — Sí, Para quedarse en una residencia estacional o una segunda residencia Yes, For a child custody arrangement Yes, While in jail or prison — Sí, Por un arreglo de custodia de niños Yes, While in a nursing home — Sí, Mientras está en la cárcel o prisión Yes, While in a nursing home — Sí, Mientras está en un hogar de ancianos Yes, For another reason — Sí, Por alguna otra razón No — No
Comments	*Interviewer must read all response options for the first person. Response options are optional reading for later people.* *If "any other reason" is given, write notes verbatim.* *Key phrases –* ***English: "****sometimes live or stay somewhere else"* ***Spanish:*** *"Algunas veces, ¿vive o se queda en algún otro lugar?"* ***All the response options must be read as worded*** for PERSON 1
Skip	End!

APPENDIX B:
Framework of Behavior Codes

Interviewer Behavior Codes **(first-level interaction)**

Code E/S	Exact Wording/Slight Change: interviewers read question exactly as worded or with slight change that did not affect question meaning or omit/change terms representing main concepts
Code MC	Major Change in Question Wording: interviewer changes to the question that either did or possibly could have changed the meaning of the question (e.g., altered verb tense, omission of reference period, paraphrasing text or substituting similar words for main concepts)
Code V+	Correct Verification: respondent provided information earlier that interviewer correctly verified and respondent accepts
Code V-	Incorrect Verification: interviewer assumes or guesses at information not previously provided (even if correct) or misremembers information when verifying and respondent disagrees
Code I/O	Inaudible Interviewer/Other: interviewer exhibits some other behavior not captured under established codes or is impossible to hear
Skipped Q	Particular questions and introductions were required reading during each administration of the survey (no skip patterns present that would cause it to be omitted), and were recorded when interviewers omitted them during the interview

Respondent Behavior Codes **(first-level interaction)**

Code AA	Adequate Answer: respondent provides response that meets the objective of the question and/or can is easily classified into one of the existing precodes
Code IA	Inadequate Answer: respondent provides a response that does not meet the objective of the question, or cannot easily be classified into one of the existing precodes—often requiring interviewer to probe for more information
Code QA	Qualified or Uncertain Answer: respondent expresses uncertainty about the response provided and may be unsure about the accuracy of the information OR respondent places modifies response by placing conditions around their response (e.g., if you mean this, then answer is that, or under this condition then answer is X, under another condition answer is Y)

Code CL	Clarification or Question Re-Read: respondent requests that a concept or entire question be stated more clearly (expressing uncertainty about meaning) OR respondent asks interviewer to reread the question
Code DK	Don't Know: respondent states he/she does not have the information
Code R	Refusal: respondent refuses to provide a response
Code I/O	Inaudible Respondent/Other: respondent exhibits some other behavior not captured under established codes or is impossible to hear

The following codes were also used to capture respondent behavior, but these aspects of the interaction were coded separately, and in addition to, the actual nature of the response/feedback. In the case of Code BI (break-in) this was done to ensure the actually nature of the response was captured, along with the interruption.

Code BI	Break-In: respondent interrupts the reading of a question or introduction (during the first-level interaction only, in other words, during the initial question-asking behavior).

Final Response Outcome Codes

The set of final response outcome codes are the same as the respondent codes used for the first-level interaction, with the exception that the clarification or question reread code (CL) was omitted. This code was excluded from the “outcome” possibilities because we suspected these behaviors would only surface during the initial question reading and any persistent problems would center around the type of answer respondents ultimately provided. Thus, the possible “outcome” codes include: AA (adequate answer), IA (inadequate answer), QA (qualified answer), DK (don’t know), R (refusal), and I/O (inaudible/other).

APPENDIX C:
Errors in the Spanish Translation in the NRFU Instrument

The following is a list of the 2006 NRFU questions which contain typos, translation mistakes and fill errors in the Spanish version. The English and Spanish intended question text is listed, followed by a description of the problems with the Spanish text for each question. These problems may be due to a lack of thorough review/proof-reading of the Spanish translation (possibly during the programming phase). In most cases they will cause interviewers not to be able to read questions exactly as worded. In some cases there is a strong likelihood that English and Spanish speakers will interpret and answer the questions differently, which will most certainly impact data quality.

Usual Residence Question

Question 2. Usual Residence

Is this [house/apartment/mobile home] a vacation home, seasonal residence, held for occasional use, or does someone in this household usually live here?

[] Vacation home, seasonal residence, held for occasional use
[] Usual Residence

¿Es [esta/este] [casa/apartamento/casa móvil] una vivienda para vacaciones, una vivienda estacional/de temporada o una vivienda que se mantiene para uso ocasional, o vive aquí usualmente algún miembro de este hogar?**

[] Una vivienda para vacaciones, una vivienda estacional/de temporada, o una vivienda que se mantiene para uso ocasional
[] Residencia habitual

The Spanish version of the question contains a typo and an incomplete sentence. The question was meant to read as written above, but the actual instrument contains the following formatting:

¿Es esta (este) [casa/apartamento/casa móvil] una vivienda para vacaciones, una vivienda estacional/de temporada o una vivienda que se mantiene para uso ocasional, o vive aquí usualmente algún miembro de este hogar¸?

[] Una vivienda para vacaciones, una vivienda estacional/de temporada, o una vivienda que se (MISSING PHRASE)
[] Residencia habitual

The second error contained in the Spanish version of the instrument is that there is a comma directly before the question mark in this question.

Finally, the response boxes that are visible to the interviewer on the screen contain an incomplete sentence. They should read as worded above but actually read:

[] Una vivienda para vacaciones, una vivienda estacional/de temporada, o una vivienda que se (MISSING PHRASE)
[] Residencia habitual

Translation:

[] Vacation home, seasonal/temporary residence or a dwelling that (MISSING PHRASE)
[] Usual Residence

Gathering the Roster (Names)

Question 4. Names

What is the name of each person who lived or stayed at this [house/apartment/mobile home] on April 1, 2006? Start with the name of one person who owned or rented this [house/apartment/mobile home].

¿Cómo se llama cada una de las personas que vivía o se quedaba en esta [casa/apartamento/casa móvil] el 1 de abril de 2006? Comience por el nombre de la persona que era [el/la] [propietario/propietaria] o quien alquilaba esta [casa/apartamento/casa móvil]. **

The Spanish version of the question contains a fill problem and some grammatical or translation errors. The question was meant to read as written above, but the actual instrument contains the following wording:

¿Cómo se llama cada una de las personas que vivía o se quedaba [casa/apartamento/casa móvil] en esta el 1 de abril de 2006? Comience por el nombre de la persona que era el (la) propietario(a) o quien alquilaba esta [casa/apartamento/casa móvil].

Literal translation:

What is the name of each of the people who lived or stayed [house/apartment/mobile home] in this on April 1, 2006? Start with the name of the person who was the owner or who rented this [house/apartment/mobile home].

First of all, the phrase “en esta” should be located before the [house/apartment/mobile home] fill. If read as worded, this question would not make sense. This question also only provides the feminine form of the word “this,” which would be incorrect if the person lives in an apartment. It should be a fill that reads [esta/este] and ideally, the instrument should fill both the correct form of “this” in Spanish and the type of housing unit in either language for the interviewer.

Secondly, there is a small translation error in that the English version: "…the name of one person who owned or rented…" was translated as "the name of the person who…" This could result in a change in interpretation by respondents, with Spanish speakers being told that there must be only one owner of the property whereas English speakers are hearing a message more like "one of the" owners.

Finally, there is a grammatical problem in the way the term "owner" and the preceeding article are written in the instrument. The term is translated with brackets containing both a masculine and feminine form of the word "owner." While appropriate for a written questionnaire, this is not a construction that can be read aloud as worded. This will require FRs to make a modification in order to read the question.

The Census Bureau Mode Consistency Guidelines state that interviewer administered instruments should contain questions written the way they would be read aloud by an interviewer. This will sometimes differ from the way something maybe written or abbreviated on a paper instrument (U.S. Census Bureau, Mode Consistency Guidelines, 2006).

The way the term "owner" is currently written, interviewers will have a number of choices as to how to handle it. To avoid wordiness they might choose only one term. They might read only the masculine form of the word, which would be appropriate for use in referring to either a man or a woman in Spanish when the sex of the person is unknown. They might read only the feminine term, which would not be appropriate as this form assumes a female owner and could only be used to refer to a woman. Finally, they might choose to read both the masculine and feminine forms, but in order to do this they would have to make a change to the question wording. They would have to say "el propietario o la propietaria." This phrase would give extra emphasis to the fact that an owner could be either a man or woman.

Rather than leaving this to the discretion of each interviewer, the ideal would be to have us make a decision on this and incorporate it into the instrument. We recommend using just the masculine term (el propietario) to refer to either a man or a woman, since this would make the question less wordy and complex. If there is a concern that this will cause people to leave women owners off of the form, cognitive testing of the two options may be in order, but since using the masculine to refer to either a man or woman is standard in Spanish (as long as the sex of the person is unknown), this does not seem necessary.

Again, the way it is currently written is appropriate for a written questionnaire, but it is not written in a way that it can be read aloud exactly as worded by an interviewer.

Ownership Question

Question 6. Ownership

Is this [house / apartment / mobile home]…

Owned by you or someone in this household with a mortgage or loan?
Owned by you or someone in this household free and clear?
Rented for cash rent?
Occupied without payment of cash rent?

¿Es [esta/este] [casa/apartamento/casa móvil] **

Propiedad suya o de alguien en este hogar con una hipoteca o préstamo?
Propiedad suya o de alguien en este hogar libre y sin deudas?
Alquilada por pago de dinero en efectivo?
Ocupada sin pago de dinero en efectivo?

The last two options present the terms rented and occupied (alquilada and ocupada) only in feminine form. If the respondent lives in an apartment, those adjectives would need to be masculine. This issue could be resolved by having a house/apartment/mobile home fill throughout the instrument. The terms "rented" and "occupied" would also need to be fills in Spanish to account for masculine or feminine forms of the adjectives. They would read [Alquilado/Alquilada] and [Ocupado/Ocupada].

Relationship Question

Question 8. Relationship2

b. Which one of these categories best describes [your/NAME's] relationship to [NAME]?

Roomer, boarder
Housemate, roommate
Unmarried partner
Foster child or foster adult
Other nonrelative

b. ¿Cuál de estas categorías es la que mejor describe la relación [de NAME] con [usted/NAME]?**

Inquilino o pupilo
Compañero de casa o cuarto
Compañero(a) no casado(a)
Hijo(a) de crianza o adulto bajo custodia
Otro no pariente

Question 8b. contains an error in the Spanish version. The question should read as worded above but the instrument actually contains the following wording:

8b. ¿Cuál de estas categorías es la que mejor describe [de NAME] relación con [usted/NAME]?

Literal translations:

The correct version reads:

> **Which of these categories is the one that best describes the relationship of [NAME] with [NAME]?**

The incorrect version reads:

> **Which of these categories is the one that best describes of [NAME] relationship to [NAME]?**

Essentially, the fill was placed incorrectly in this question and should be located after the word "relación" and not before it.

Sex Question

> **Question 9. Sex**
>
> **[Is NAME /Are you] male or female?**
>
> **¿Es [NAME/usted] de sexo masculino o femenino?****

When using the NAME fill for third person questions in Spanish, the instrument is correct but when referring to the second person "you" in Spanish, the instrument reads the same way as it does in the third person questions: For example, rather than reading:

> **¿Es [usted] de sexo masculino o femenino?**
> **[Are you] male or female?**

The instrument reads:

> **¿Es [María] de sexo masculino o femenino?**
> **[Is María] male or female?**

The incorrect fill with a name appears in Spanish when it should read: "you."

Date of Birth Question

> **Question 10. Date of Birth**
>
> **What is [your/NAME's] date of birth?**
>
> **¿Cuál es [su] fecha de nacimiento?**
> **OR**
> **¿Cuál es la fecha de nacimiento [de NAME]?****

The Date of Birth question itself (listed above) reads correctly in the instrument, however, the Date of Birth Check question (DOBCHECK(1)), for which we did not conduct behavior coding in this study, contains a major error in Spanish. It contains an incomplete sentence.

The correct Date of Birth question in English reads:

> **Since your age as of April 1, 2006 was [#], can you help me correct your date of birth?**
>
> **I have [date]. What should it be?**

The incorrect Spanish version of the question reads:

> **Como usted tenía [#] años de edad el 1 de abril de 2006, usted [MISSING PHRASE]**
>
> **Tengo anotado [date]. ¿Cuál debería ser?**

Translation:

> **Since you were [#] years old on April 1, 2006, you [MISSING PHRASE]**
>
> **I have listed [date]. Which one should it be?**

This is an extremely large error and leaves FRs unable to read the question as worded.

Age Check Question

Question 12. Age Check

> **For the census, we need to record age as of April 1, 2006. So, just to confirm – [NAME was/you were] [AGE] on April 1, 2006?**
>
> **Para el Censo, tenemos que anotar la edad que se tiene el 1 de abril de 2006. ¿Por lo tanto, solo para confirmar tenía [usted/nombre] [age] años el 1 de abril del 2006?****

This question contains an incorrect preposition in the Spanish version. Correct question wording is listed above. In the instrument, the question actually reads:

> **Para el Censo, tenemos que anotar la edad que se tiene al 1 de abril de 2006. Por lo tanto, solo para confirmar, ¿tenía [usted/nombre] [age] años <u>en</u> el 1 de abril de 2006?**

It is grammatically incorrect to express a date using the construction “en el.” The preposition “en” should be dropped.

Race Question

Question 14. Race

“What is [YOUR/NAME’S] race? (You may choose one or more races. For this census, Hispanic origins are not races. White or Caucasian, Black, African American or Negro, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Some other race?)

“¿Cuál es [su/la] raza [de NAME]? (Usted puede elegir una o más razas. Para este censo, origen hispano no es una raza. Blanca o Caucásica, Negra o africana americana, India americana o nativa de Alaska, Asiática, Nativa de Hawaii u otra de las islas del Pacífico, Alguna otra raza)?**

When asking the question of a respondent and using the term “your,” the fill for the Spanish question is employed incorrectly and the FR sees:

“¿Cuál es su raza [de NAME]?” meaning “What is your [NAME’s] race?”

If read as worded, this would be grammatically incorrect, so this encourages interviewers to deviate from the question wording in the instrument.

The fills are employed correctly in Spanish for third person questions, such as “¿Cuál es la raza de [NAME]?” meaning “What is [NAME’s] race?”

There is an additional error in the Spanish interviewer instruction for this question. The interviewer instruction is missing a word and it reads: “Vaya al panel de las [MISSING WORD]” meaning “Go to the panel on [MISSING WORD].” The English instruction reads: “Go to the answers panel,” so clearly the word “answers” or “respuestas” is missing from the Spanish instruction.

Ancestry Question

Question 15. Ancestry

“(People <u>in</u> the United States are from many countries, tribes, and cultural groups.) What is [YOUR/NAME’S] ancestry or tribe? (For example, Italian, African American, Dominican, Aleut, Jamaican, Chinese, Pakistani, Salvadoran, Rosebud Sioux, Nigerian, Samoan, Russian, etc.)”

“(Las personas <u>de</u> los Estados Unidos provienen de muchos países, tribus y grupos culturales.) ¿Cuál es [su] ascendencia o tribu? (Por ejemplo, italiana,

africana americana, dominicana, aleuta, jamaicana, china, paquistaní, salvadoreña, Rosebud Sioux, nigeriana, samoana, rusa, etc.)**

There is an error in the translation of this question. The English version reads "People in the United States…" The Spanish version reads "People from the United States…" This type of discrepancy could have an impact on whether respondents interpret the question as asking only about native-born U.S. citizens or about whether it applies to immigrants as well.

Overcount Question

Question 16. Overcount

[Do you/Does NAME] sometimes live or stay somewhere else:
To attend college?
To be closer to work?
While in the military?
To stay at a seasonal or second residence?
For a child custody arrangement?
While in jail or prison?
While in a nursing home?
For another reason?

Algunas veces, ¿vive o se queda [usted/NAME] en algún otro lugar?
Para asistir a la universidad?
Para estar más cerca del trabajo?
Mientras está en el servicio militar?
Para quedarse en una residencia [estacional/de temporada] o una segunda residencia?
Por un arreglo de custodia de niños?
Mientras está en la cárcel o prisión?
Mientras está en un hogar de ancianos?
Por alguna otra razón?**

The Spanish version should read as written above. In the instrument it reads correctly for second person "you" question. However, when asking the question in the third person it reads:

Algunas veces, ¿vive o se queda/vivía o se quedaba [NAME] en algún otro lugar?

Translation:

Does/ did [NAME] sometimes live or stay somewhere else?

In Spanish the question, when read as worded, is asking whether someone lives or stays somewhere else presently and whether they did so in the past. The English question only asks about the present time.

The Spanish question continues with incorrect fills and reads:

Para asistir a la universidad?
Para estar más cerca del trabajo?
Mientras está/estaba en el servicio militar?
Para quedarse en una residencia estacional/de temporada o una segunda residencia?
Por un arreglo de custodia de niños?
Mientras está/estaba en la cárcel o prisión?
Mientras está/estaba en un hogar de ancianos?
Por alguna otra razón?

Translation:

To attend college?
To be closer to work?
While (you) are/were in the military?
To stay at a seasonal/seasonal or second residence?
For a child custody arrangement?
While(you) are/were in jail or prison?
While (you) are/were in a nursing home?
For another reason?

The instrument is in effect asking Spanish-speaking respondents to think of both the present and past, but with no specific reference period. The English-language question uses only present tense verbs. This will most likely cause a differing interpretation and response on the part of respondents.

This question also contains two translations for the phrase “seasonal residence,” listing both the terms “estacional” and “de temporada.” Please see the discussion above on the Usual Residence question for a more detailed discussion of this issue.
