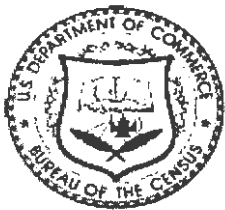# BEHAVIOR CODING ANALYSIS REPORT

Evaluating Bilingual Versions of the Non-Response Follow-Up (NRFU) for the 2004 Census Test

Prepared by: Jennifer Hunter & Ashley Landreth
U.S. Census Bureau
Statistical Research Division
Center for Survey Methods Research

TABLE OF CONTENTS

Appendices

A. INTRODUCTION

On behalf of the Planning, Research and Evaluation Division (PRED), staff from the Statistical Research Division (SRD) conducted a behavior coding analysis to pretest the 2004 Census Test Non-Response Follow-Up (NRFU) interview. PRED's interest in this evaluation centered on the performance of the Race and Hispanic Origin questions used in the 2004 Census Test, due to an Office of Management and Budget (OMB) directive which omitted the Race question's open-ended response category labeled "Some Other Race." A few additional sources of interest drove the expansion of this project's focus to include the performance of the coverage questions, the Spanish-language version of the questions, and interviewer behavior for questions asked repeatedly for each roster member in a person-based interview environment.[1] The selection of the behavior coding method for this evaluation was driven by the need to generate interviewer/respondent behavioral data to identify question administration and response difficulties.

This pretesting project pertains to the NRFU operation conducted in Queens, New York in the spring and summer of 2004. A sample of face-to-face interviews using an automated instrument on handheld computers (HHC) was taped, resulting in 220 audio-taped interviews (119 English, 72 Spanish, and 29 combination English/Spanish). These interviews were conducted with household members only; interviews with non-household members were not included in this study. Bi-lingual telephone interviewers from the Tucson Telephone Center (TTC), trained in behavior coding techniques, applied project-specific behavior codes to interviewer and respondent interactions for each survey question.

B. RESULTS

General Observations

General observations are derived from tables aggregated across all cases that summarize the results of each behavior code, by question (See Appendix C for a description of the framework of behavioral codes used for this study).

Regarding interviewer behavior:

- The automated NRFU instrument produced an interview that was less standardized than we would have hoped. The commonly accepted error threshold states that non-ideal interviewer behavior (e.g., major changes to question wording, omitting a question) should occur no more than 15 percent of the time for each behavior (Oksenberg, et. al, 1991; Fowler, 1992).
    - ➢ The questions of particular interest to this project—the Hispanic Origin and follow-up questions, the Race question, and the Coverage questions—all

---

[1] Staff from the Residence Rules Working Group expressed interest in the Coverage questions, and SRD staff was interested in the Spanish-language version of the NRFU and the effect of repeated question administration.

exceeded the error threshold. Interviewers made major changes to question wording between 26 and 66 percent of the time for these questions.

> Across all the questions, ideal interviewer behavior reached its peak at 71 percent for the question asking if anyone lived at the residence on April 1st and was at its lowest at 15 percent for the coverage question that asked if each person had another place that they also stayed.

> On average, ideal question-asking behavior across the questions that we analyzed was only 33 percent.

- Spanish-language questions had consistently lower rates of ideal interview behavior than the English-language counterparts.

- The trend for every question that we analyzed was that the question was asked as intended (i.e., exactly as worded or with slight changes) more often for Person 1 than for other persons in the household.

Regarding respondent behavior:

- Looking across all the questions, respondents had difficulty meeting the measurement objective without further probing on the part of the interviewer:

> Usual Residence (21%), Relationship (17%), and Race (36%) elicited high levels of unacceptable responses.

> After some interviewer/respondent negotiation, these problems disappeared for Usual Residence and Relationship, but inadequate answers persisted for Race (34%).

- Indications of negative respondent reactions to these questions were all but absent from the analysis; however, the subjective nature of these interactions makes capturing them difficult and unreliable.

Question-by-Question Observations about Target Questions

Hispanic Origin (#17)
The Hispanic origin question was both skipped and reworded more often for Person 2 and later in the household than it was for Person 1. When it was reworded, it was often reworded by using only one or two of the three terms to describe Hispanic origin or by adding terms, like Mestizo or Trigueno to the list of terms in the scripted question. The inadequate responses to this question were generally races or non-Hispanic countries of origin.

Follow-Up Questions: Hispanic Yes (#18) & Hispanic Other (#19)
The follow-up questions to the Hispanic origin item were also skipped more often for Person 2 and later than for Person 1. Major changes to the Hispanic origin follow-up question were often in the form of asking from which country the person came, for their nationality, reading only a few of the countries in the question, or asking if Person 2 and later were of the same origin that Person 1 was. Positive verification was present more often for Person 2 and later in the

household, indicating that the respondent had told the interviewer that everyone (or at least a subset of the household) was from the same country of origin.

Race (#20)
The Race question was asked exactly as worded or with minor changes only 16 percent of the time overall. It was skipped almost half of the time for Person 2 and later in the household. 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). Surprisingly, in some of the major change cases, the interviewer offered the respondent races that were not codeable answers, including things like Mestizo, Trigueno, Hispanic or Latino. This could reflect the interviewers' desire to overcome the problems that Hispanic respondents have with the Race question. However, it leads the respondent to answer with an uncodeable response when they might otherwise have chosen an adequate response. In over half of the cases where a respondent did not provide an adequate response, the respondent answered the Race question with a Hispanic origin.

Coverage1 (#7)
Undesirable interviewer behavior was evident almost half the time Coverage1 was administered; it was often read with major changes to question wording or omitted completely from the interview. When question wording was changed, half the time the interviewers completely omitted the main conceptual phrase, either by paraphrasing or attempting to verify information with respondents. This resulted in a different question than was originally intended: "Anyone else you forgot to mention that is part of the household?" versus "Are there people who aren't part of the household typically, but sometimes stay here?" Even though behavior coding results did not suggest there were response issues, clearly some respondents were asked quite a different question than was intended, which may affect data quality.

Coverage2 (#24)
The question contains two design features that are likely driving the high rate of unacceptable interviewer behavior and respondent interruptions. The current question stem contains a fully formed, complete question, and it is followed by a response set (required reading for interviewers) that contains a second, separate question. After hearing the first portion of the inquiry in the question stem, respondents thought they were aware of the question's intent and provided a response—hence the high rate of interruption at the question stem's conclusion. Interviewers compensated by omitting the response set entirely, paraphrased the question by incorporating one or two response options directly into the question stem itself, and avoided asking this question for Person 2 and later on the roster.

Question-by-Question Observations about Other Problematic Questions

The following questions are included here because behavior coding data indicate that they are not working in the field. Interviewers are changing the wording and intent of the questions without the benefit of subject matter or survey methodology expertise. The evidence presented below indicates problems with the question wording that should be further explored through revisions and cognitive testing of those revisions.

Usual Residence #2
The usual residence question suffered from major change more than half of the time. Most of the major change cases were because the interviewer only asked if it was the usual residence – making it a "yes/no" question instead of a "choose one." Interviewers may have done this because they felt that the other options were not necessary or that they were used so infrequently that they were not useful. At any rate, this indicates that the question is not working in the field.

Names #4
The question used to gather the roster was also read with a major change more than half of the time. In many of the major change cases, the interviewer asked the respondent to start with him or herself rather than asking for the name of a person who owns or rents the home. In other cases, the interviewer did not provide instruction for whom to start with. This could cause problems for determining relationships in the household because Person 1 may or may not be the owner or renter of the housing unit. From other cognitive research, we know that respondents sometimes forget to list themselves when giving a household roster. Perhaps interviewers had noticed this, and attempted to remedy it by asking for their names first. Alternatively, it may just be that they find it easier and more conversational to start with the person with whom they are talking.

Ownership #10
The question that asks if the housing unit is owned or rented suffered from the highest percentage of major change. This indicates that the question is awkward to be read completely as worded, possibly because it is the longest question in this survey.

Relationship #11
The question that asks how each person is related to Person 1 had an unacceptable rate of inadequate answers. In some of the cases when respondents gave an inadequate answer, they told the interviewer the people were simply related or not related. If the flashcard is not use, the question as it is currently worded can elicit the response "they are not related," which is not codeable.

Sex #12
The question on sex was skipped almost half of the time and verified another quarter of the time. This indicates that it may be a sensitive question to ask of respondents and/or interviewers have received information earlier in the interview to let them know what the sex of the person is.

Age #13
Age had a relatively high rate of requests for clarification. We hypothesize that the confusion stems from asking age as of a certain date – something that people are not used to answering.

Recommendations

The body of the report offers recommendations that address these observed problems.

ACKNOWLEGMENTS

On behalf of the Planning, Research and Evaluation Division (PRED), staff from the Statistical Research Division (SRD) conducted a behavior coding analysis to pretest the 2004 Census Test Non-Response Follow-Up (NRFU) interview. The behavior coding method was selected for its ability to detect interviewer and respondent behaviors indicative of question administration and response problems.

The impetus for pretesting the 2004 NRFU instrument came largely from a directive handed down by the Office of Management and Budget (OMB) regarding the Census Bureau's Race question. The directive required that the open-ended response option, labeled "Some Other Race," be removed. Thus, the focus of much of this behavior coding analysis is on the effect this decision had on the performance of the questions on Hispanic Origin and Race.

Other divisions and working groups also conveyed interest in this behavior coding project. This was an opportunity to learn how respondents interacted with the items collected on the decennial census form in an interviewer-administered survey. For this reason, the scope of the project was expanded beyond the questions of interest to PRED to include other questions that were of interest to other teams. Among the teams that expressed interest were those working on the coverage questions and those tasked with evaluating the use of a hand-held computer to collect census data.[2]

Staff from SRD hypothesized that two variables would affect interviewer and respondent behavior. The first was interview language (i.e., Spanish or English) and the second was "repeated question administrations," meaning, when the same question is asked for the first time, versus when it is repeated for additional members of the household. We examined the difference in language to see if there were problems that were concentrated in one language more than the other. The second variable of interest reflected the fact that the same questions were asked about multiple household members. We were interested in whether we would find more undesirable interviewer behavior when interviewers were required to repeat the survey questions for persons after Person 1. Once the interviewer has asked the question one time and repeats it for later household members, we thought the interviewer might shorten or rephrase the questions to speed up the interview or to avoid sounding redundant.

Throughout this report, there will be references to these specific areas of interest as well as a more general assessment of how the questions in the NRFU interview performed. The remainder of this report is divided into three main sections: 1) the background for the project and the NRFU instrument; 2) a discussion of the methods used to carry out the research; and 3) a presentation of 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

---

[2] Representatives from the Residence Rules Working Group conveyed an interest in the performance of the coverage questions. Members of the Decennial Statistical Studies Division (DSSD) staff were interested in a broad analysis of the feasibility of using a handheld computer to collect census data. DSSD staff members were more interested in the overall interaction patterns than in the analysis of any question or set of questions in particular.

wording, where appropriate, that may alleviate question administration and response difficulties observed in this research.

The U. S. Census Bureau conducts many small-scale tests throughout the decade to ensure the next decennial census will run smoothly. During 2004, the Census Bureau conducted a test of the NRFU operation. The NRFU operation is an attempt to gather census data from people who did not respond to the mail-out census form by interviewing them in person. The 2004 Census Test was conducted at two test sites (i.e., rural parts of Georgia and Queens, New York), selected in order to include populations that we find typically more difficult to enumerate (e.g., rural populations and recent immigrant populations).

The NRFU survey contains questions from the self-administered decennial paper form converted into an electronic format. The NRFU question wording differs slightly from the paper form because it is an interviewer-administered computer-assisted personal interview (CAPI). The NRFU survey was administered using a handheld computer (HHC)[3] for the first time during the 2004 Census Test. An HHC is an electronic device that collects and stores data much like CAPI survey would with a laptop, but the HHC is much smaller than a laptop and is operated by using a stylus instead of a keyboard. The HHC contains NRFU questions in both English and Spanish, and interviewers can select the language for the interview and "toggle" back and forth between languages, if needed, during an interview.

The NRFU interview is relatively short (i.e., 7 to 30 minutes per household depending on the number of household members) and it begins at the household level by asking questions relevant to the entire household. About mid-way through the interview, the survey switches to person-level questions, asking a series of questions for each person who lives in the household (in a person-based manner). The interview includes questions about the following:

Household-level topics:

- Usual residence or seasonal unit;

- Roster questions;

- Other household members who may live there part time; and

- Housing status (e.g., own with a mortgage, own without a mortgage, rent, etc.).

Person-level topics:

- Sex;

- Relationship to householder;

- Age/date of birth;

- Hispanic origin;

- Race; and

---

[3] HHCs are also sometimes referred to as "personal digital assistants" (PDAs) or "pocket PCs."

- Other places the person might live.

The English- and Spanish-language versions of the questions, called "Question Guides," can be found at the back of this report in Appendix A (English) and B (Spanish).

A.   STUDY DESIGN

Our goal was to capture approximately 250 face-to-face NRFU interviews onto audio-tape at the Queens, NY test site. Collection of the audio-taped interviews were monitored to ensure that roughly half the interviews taped were in English and half in Spanish to assure sufficient sample for analyzing questions of interest for each version of the survey. The sample was not a representative sample, but was rather a sample of convenience.

Interviews were taped throughout the field period of NRFU, from May to July of 2004. Of the 256 audio-tapes collected in the field, a total of 220 audio-tapes were usable;[4] of these 72 were in Spanish, 119 were in English and 29 were conducted in a combination of Spanish and English. Thus, sufficient sample was achieved to analyze the questions of interest in both languages.[5]

The audio-taping of face-to-face interviews required additional staff to assist with the recording equipment and consent procedures. Trained assistants, called Enumerator Taping Assistants (ETAs), accompanied field interviewers on their assignments and were tasked with gaining respondents' consent to be recorded and operating the recording equipment. One ETA typically accompanied an interviewer for the day, perhaps taping a few interviews for the same interviewer.[6] Each day the ETA was instructed to go out with a different interviewer.

In order to conduct a bilingual behavior coding project, SRD staff collaborated throughout this process with Dr. Carlos Arce, of NuStats, a contractor with the Census Bureau, who specializes in bilingual survey research. Dr. Arce's proficiency in methodological research, expertise and extensive scholarship in the study of Hispanic cultural subgroups, and his capability to work in Spanish as well as English helped ensure the development of the training and analysis of the Spanish-language data were appropriate in the context of the Spanish language and culture.

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 2004. Coders were selected based on their fluency in speaking and reading both English and Spanish, and by their reliability as interviewers as judged by their supervisors.[7] The training was designed and conducted by the authors and Dr. Arce.

To create coding assignments for each coder, the audio-taped interviews were aggregated by interview language, household size, and field interviewer and distributed into five coding

---

[4] 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.

[5] We achieved a sample that was beyond the minimum number of interviews used for other behavior coding projects undertaken by SRD (Zukerberg, Von Thurn, and Moore, 1995).

[6] We analyzed interviews from approximately 46 interviewers. The average number of interviews per interviewer was five. No single interviewer conducted more than 10 percent of the interviews.

[7] Coders interviewing experience at the telephone center ranged from one to two-and-a-half years (average 2.1 years). Spanish was the first language for three of the coders, while English was the first language for two coders. Three of the coders were Arizona natives, while the other two originated from a Mexican town bordering Arizona.

assignments, such that each coder had an approximately equal number of interviews in each language, for each household size and for each field interviewer.[8] Behavior coders applied the prescribed framework of behavioral codes to interviewer and respondent behaviors by listening to the audio-tapes and following the interview's progress by reading a written version of the questionnaire, called a "Question Guide." Two versions of the Question Guides were created, one for each interview language, and they contain survey questions that are formatted much as they would have appeared on the screen of the hand-held computer. By comparing the written document to the interviewers' recitation of the questions, coders made assessments about the interviewers' ability to read questions as they were worded. Coders also made assessments regarding whether or not responses to the questions met the objective of the measurement goal, but they did so based upon the audio-tapes only; coders did not have access to data generated by each interview so they did not know how interviewers ultimately coded a respondent's answers.

B.   LIMITATIONS

Aspects of the research design present limitations to this study and necessitate some caution in interpreting and understanding the results. The use of audio recordings to capture face-to-face interviews is the main source of these limitations, in that the audio recording limits the interactions we can code to verbal communication only. And, the process of audio-taping itself may have unintended effects on interviewer and respondent behavior.

First, audio recording restricts observable behavior to verbal communication, which misses nonverbal behavior and communication that occur naturally as part of the interviewing process in face-to-face interviews. For instance, a respondent might nod his or her head to a yes/no question indicating affirmative agreement, but this silent behavior goes undetected on an audio-tape and thus cannot be adequately captured and represented by the behavior codes. Essentially, the respondent's behavior in this situation is recorded as "inaudible" (which is in contrast to adequate), and therefore the number of adequate and codeable answers 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 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 at times 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 ability to read the question exactly as worded.

Furthermore, the act of taping the 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 an ETA may have an effect on interviewers' or respondents' behaviors (e.g., respondents may be less willing to inquire about vague terms or complex questions in the presence of two Census Bureau employees versus a one-on-one interview). Additionally, it was noted in some cases that the ETA was heard on the tape interacting during the interview. In these cases, the ETA was not a passive

---

[8] Each coder coded approximately 50 audio-taped interviews. The coders' caseloads included duplicates of tapes used for reliability purposes.

observer, but rather a third party during the interview. This behavior coding project does not capture or account for this type of interaction.

Another aspect of the behavior coding analysis that should be viewed with caution is the results from the "negative reaction" code. While coders were instructed to apply this code only in clear instances of respondent aggravation, applying this code remains extremely subjective and the resulting data are subject to reliability issues.

This analysis was limited to face-to-face interviews with a member of the household. Results are not necessarily generalizable to whole household proxy or telephone interviews.

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. 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 respond to, then our confidence grows regarding the ability of the survey instrument to meet the 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.

A framework of behavioral 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 questions that are awkwardly worded or overly complex. In addition, skipping questions that should be read might indicate 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 became more complicated. For instance, when terms are unclear, respondents may ask for clarification, or when a question is lengthy or complex, respondents may ask interviewers to reread the entire question. Additionally, refusals may indicate that respondents perceive a request for information to be too sensitive, whereas a "don't know" response 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 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 question 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). This approach, however, lacks the ability to demonstrate whether the interviewer and respondent were ultimately successful in resolving difficulties with the question-and-answer process before moving on to the next survey item. Therefore, in addition to the first-level interaction—also sometimes referred to as the first-level exchange—the "final response outcome" was also coded to determine whether an acceptable resolution was reached. Outcome codes are used to identify whether some type of acceptable or codeable answer was negotiated or whether undesirable respondent behavior persisted as the interviewer exited the question and continued with the interview. In addition, when non-ideal interactions occurred anywhere during the question administration, coders were instructed to transcribe the conversation for later qualitative analysis.

The framework of behavioral codes used for this project and an explanation of their analytical function is listed in Appendix C. The behavioral codes were designed to capture five 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"); 4) final response outcome; and 5) negative reactions to questions by respondents.

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 eight interviews, four in English and four 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 .75 indicate an excellent level of agreement across coders, while scores ranging from .40 to .75 indicate a good to fair level of agreement; scores below .40 represent poor agreement.

The following kappa scores reflect coders' agreement for three of the five behavioral variables, aggregated by interview language.[9] For interviewer behavior in both languages, and for respondent behavior and final response outcome in the English-language interviews, the kappa scores range from .70 to .48 and reflect a good to fair level of agreement. There are some issues, however, with coders' agreement for respondent behavior and final response outcome in the Spanish-language interviews (.34 and .31, respectively).

---

[9] 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' kappa scores were as follows: interviewer behavior (.68); respondent behavior (.76); and final response outcome (.75).

Table 1. Behavior Coders' Kappa Scores by Language

Interview Language	Interviewer Behavior	Respondent Behavior	Final Response Outcome
English	.63	.48	.70
Spanish	.50	.34	.31

Aspects of the data for the remaining two behavioral variables, break-ins and negative reaction, rendered the kappa test an inappropriate measure of inter-coder agreement.[10] Thus, percent inter-coder agreement was calculated, resulting in an average agreement of 93.4 percent for break-ins (92.8% and 94.0% for English- and Spanish-language interviews, respectively) and 98.5 percent for negative reaction (97.8% and 99.1% for English- and Spanish-language interviews, respectively).[11]

The low level of agreement for respondent behavior and final response outcome in the Spanish-language interviews (.34 and .31, respectively) would normally present an impediment to interpreting the Spanish-language interview results. Ordinarily we would suggest 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). This coder error likely contributed to the low kappa scores, and 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 less than those produced by the English-language interviews. There may have been an interaction between the translation of the Spanish-language instrument, the language skill of the NRFU interviewers, the language skill of the behavior coders, the comparative difficulty of the Spanish interviews versus the English interviews, and the possible legitimate assignment of multiple codes to the same

---

[10] The kappa statistic assumes the ratings have equal opportunity to be used by the coders and cannot calculate a meaningful score for tables in which the majority of the cells contain no data. Due to the fact that break-ins and negative reactions were infrequent and dichotomous (i.e., either they occurred or they did not), many of the cells in the 2 X 2 tables were empty or contained so little data that the kappa statistic results were not meaningful for the purpose of assessing inter-coder reliability.

[11] For comparison purposes, percent inter-coder agreement was also calculated for the behavioral variables above for which kappa scores were generated. The average agreement for those three variables for the English- and Spanish-language interviews was 78.0 percent (80.1% agreement on interviewer behavior, 65.7% on respondent behavior, and 88.1% on final response outcome) and 76.3 percent (73.9% agreement on interviewer behavior, 74.7% on respondent behavior, and 80.4% on final response outcome), respectively.

interaction for particular questions under certain circumstances. Without further analysis of the audio-taped interviews, however, it is impossible to determine the exact cause of this result.

E.   QUALITY CHECK

During an initial review of the dataset by the researchers, it was evident that coders had some difficulty applying the codes, and that they often coded the same situation in different ways.[12] This was particularly evident for two questions of primary interest to this research, the Hispanic Origin and Race questions. We recoded a portion of the variables to more accurately reflect the interactions when the notes were clear and justified the assignment of an alternative code.[13] For example, the coders sometimes mistakenly coded an interviewer verification (V+ or V-) as a major change (MC) when they asked respondents, "You're Hispanic, right?"

Coders also had difficulty determining when a response met the measurement objective of the question. For instance, when the question was asked with a major change (MC), yet produced an adequate answer (AA), coders sometimes miscoded acceptable responses from respondents as inadequate answers (IA), perhaps because interviewers figured the respondent was answering a different question so technically their response was invalid. An example that easily demonstrates this phenomenon occurred frequently for the Race question. If the interviewer said "Are you white?" and the respondent affirmed, that is a major change (MC) for the interviewer, but an adequate answer (AA) for the respondent because "white" meets the measurement objective and is an acceptable response according to that question's response set.

The recoded data produced a more conservative picture of the errors for both interviewer and respondent behavior. Overall, exact reading percentages increased, while major change percentages decreased. Both positive and negative verifications also increased. For respondent behavior, adequate answer percentages generally increased while inadequate answer percentages decreased. The exception to this trend was for the Race question, in which adequate answers decreased and inadequate answers increased. This was due to the fact that interviewers sometimes miscoded Hispanic origin responses as adequate answers to the Race question, which further demonstrates the difficulty with these two questions. We suspect that because the coders were also Hispanic, they may have had more difficulty conceptualizing and evaluating responses to the Census Bureau's collection of race and ethnicity data using the two-question approach. It seemed confusing to some Hispanic coders that "Hispanic" was not a legitimate response to the Race question.

---

[12] This was apparent because of the notes that accompanied all codes that indicated an interaction that was other than ideal.

[13] All questions that were analyzed and are discussed in this report were examined for recoding. The total percent of recoding for Interviewer Behavior (IB), Respondent Behavior (RB), and Outcome (OB) was 5%. Those questions that were recoded at a level of 10% or more included Relationship (IB) at 13%, Sex (IB) at 18%, Hispanic Origin (RB and OB) at 13% and 12% respectively, and Race (IB, RB, and OB) at 14%, 21%, and 16% respectively. Recoding rates were similar across language (5% for English and 7% for Spanish), with the exception of the Hispanic origin and Race questions. RB and OB for Hispanic origin was recoded at 8% for English interviews and 17% for Spanish interviews. For Race, English recoding was 13%, 15%, 11% and Spanish was 15%, 29% and 23% for IB, RB, and OB, respectively. Spanish notes were recoded from a direct translation of the notes into English.

F. LOGISTIC REGRESSION USING INTERVIEW LANGUAGE AND REPEATED QUESTION ADMINISTRATIONS

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 and all other administrations of the same question) on "good" behavior by the interviewer and the respondent for each question.[14] We looked at the effect of language and repeated administration on interviewer behavior, respondent first-level exchange behavior, and respondent outcome behavior.

---

[14] Exact wording/slight change (Code E/S) and positive verification (Code V+) were considered "good" interviewer behavior. The only behavior that was considered good respondent behavior was an adequate answer (Code AA).

A.   INTRODUCTION

For the purpose of this report, we plan to focus on a few questions of critical interest to the behavior coding evaluation of the NFRU, namely the Hispanic Origin question, along with its associated follow-up questions (#17, 18, 19), and the Race question (#20). Due to an insufficient number of observations generated by the Race follow-up questions for American Indians, Asians, and Pacific Islanders (#21, 22, 23), these questions have been excluded from the analysis. Next, the coverage questions are discussed (#7, 24). Toward the end of the report, brief results are presented concerning other questions where the behavior coding data indicated a reason for concern. Those questions are about usual residence (#2), gathering the roster (#4), ownership (#10), relationship (#11), sex (#12), age and date of birth (#13, 14). The full set of behavior coding results is available in Tables 2 and 3.[15]

Before discussing the behavior coding results in general, we will describe the two main tables to which we will refer throughout the larger results section of this report. Next we discuss the strategy used for determining when a question is causing administration and response issues. Then, we will present a brief discussion of our general impressions of the data. Finally, we will identify and discuss the logistic regression results that reveal significant differences in the performance of the English- and Spanish-language versions of the questions, as well as significant differences for interviewer behavior between the first administration of the question and subsequent administrations of the same question (referred to as "repeated administrations").

Behavior Coding Results Tables

The aggregate results of the behavior coding for interviewer and respondent behaviors are contained in Tables 2 and 3. Table 2 contains interviewer behavior parsed across six variables per question, accounts for approximately 100 percent of interviewers' behavior, 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); and
>
> Question skipped incorrectly (Skip).

Similarly, the percentage of respondent behaviors at the first-level exchange is parsed across nine variables per question, which includes the following:

---

[15] Because of their decreased significance to the evaluation of the NRFU instrument, a few questions have been omitted from these attachments. These questions are mainly verification questions and follow-up probes to more substantive questions, for example, "Is there anyone else?" (Question 5). The omitted question numbers are 5, 6, 8, 9, 15, and 16.

Adequate answer (AA);

Inadequate answer (IA);

Uncertain answer (UA);

Qualified answer (QA);

Clarification requested (CL);

Request to re-read question (RR);

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 this table. These calculations were made based upon the total number of first-level exchange respondent behaviors for each question. Break-ins are calculated separately from the nine 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 codeable or uncodeable or they may interrupt for clarification).

Table 3 contains percentages for the respondents' final response outcome and contains the same variables included for the first-level respondent behavior, excluding question re-read (RR) and clarification request (CL). The percent of respondents' negative reaction per question is also provided in this table, and was derived by the same method as the break-in column data.

These tables represent approximately 220 households (containing almost 800 people in total) interviewed for the NRFU survey. Questions 1, 2, 3, 4, 7, and 10 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. This accounts for the large increase in the number of observations for Questions 11 through 24 (e.g., Question 20, the Race question, contains 789 observations).

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 each type of behavior (e.g., major change or inadequate answer). Though this is an arbitrary cut-point, this level of non-ideal behavior suggests a question has a "high level" of the problem that merits some attention (Oksenberg, et. al, 1991; Fowler, 1992).

B.   GENERAL OBSERVATIONS

According to the standard for interpreting behavior coding data, the interviewer behavior results are quite striking; for all the questions that were administered at least 20 times, all but one question (i.e., Sex, #12) exceeded the 15 percent threshold for major changes made to question wording (see Table 2). Interviewers altered wording to the extent that question meaning could

have been interpreted differently than intended anywhere from 26 to 67 percent of the time. For the questions of primary interest to this study, the Race and Hispanic Origin questions and the Coverage questions, the results look very similar; major change behavior ranged from 26 to 66 percent. In fact, on average, ideal question-asking behavior across the all of the questions that we analyzed was only 33 percent (it was between 15 and 51% for the particular questions of interest). This demonstrates that for most of the NRFU instrument—including those questions of substantive interest to this project—interviewers had a difficult time achieving standardized question reading at satisfactory levels.

In terms of respondent behavior, the results indicate respondents were able to provide adequate answers to most questions; however, there were some problems. Despite the lower-than-desired rates for questions read to respondents "as-worded," surprisingly few questions, three, generated a significant level of responses that failed to meet the measurement objective. The percent of inadequate answers rose to unacceptable levels in the questions on Usual Residence (21%), Relationship (17%), and Race (36%). When looking at the final response outcome (Table 3) these issues disappear for two of the three questions, but an unacceptable level of inadequate answers persists for the Race question (34%). For the Usual Residence and Relationship questions, the resolution of most of the inadequate answers indicates the responses offered by respondents did not fit neatly into the available response set at first, either due to confusion over question intent or response options. In either case, it appears that, at least for these two questions, the issue was eventually resolved through interviewer/respondent negotiation. The persistent problem in the Race question, as evidenced by the high percent of inadequate final outcome responses, indicates the concepts may be poorly communicated and/or respondents may not able to respond within the parameters of the response set provided. In either case, these results suggest a significant number of respondents were unable to negotiate the Race question to the satisfaction of this question's measurement goal, based on the answers provided to the interviewers.[16]

There were a few additional questions for which unacceptable or non-ideal respondent behaviors approached the 15 percent threshold. The Date of Birth question generated 14 percent "don't know" responses, but this probably reflects that respondents sometimes find they lack this information for others in the household. Four questions were plagued by respondent interruptions more than 10 percent of the time a response was offered during the first-level exchange: these were the Coverage questions (Coverage 1, 11% and Coverage 2, 13%); Ownership question (11%); and the Hispanic Yes follow-up question (11%). Three of these questions contain illustrative examples interviewers are either instructed to list after reading the main question (Coverage questions) or are incorporated into the question itself (Hispanic Yes question). In the Ownership question, each of the four response options is to be read in full, creating a lengthy question to which some respondents can give an answer without hearing the later options. Not surprisingly, these types of question constructions cause some respondents to interrupt once they recognize the type of response expected or they hear a particular response that applies to them.

---

[16] Though we did not have access to the data file generated by NRFU at the time this report was written, it would be worthwhile to explore how interviewers field coded/recorded these responses.

Negative Respondent Reactions

The NRFU questions were also evaluated for their ability to evoke "negative reactions" from respondents, but there were very few negative reactions noted in this study. In fact, no question exceeded 1 percent of cases exhibiting a negative reaction. There are two possible explanations for this. The first is that there simply were not many negative reactions to these questions. The second is that the behavior coding method, because it relies on the evaluation of individual coders, is not equipped to reliably detect something as subjective as negative reactions. Behavior coders were not privy to any body language that the respondent may have been providing. They had to make a judgment on emotion based only on the sound of someone's voice and the words that they said. This is, undoubtedly, a difficult, if not impossible, task for coders and may not accurately reflect the true level of anger, irritation, or frustration evoked by these questions. Therefore, the reliability of these data is highly questionable, and was mentioned earlier as a potential limitation to this study.

Logistic Regression Results

The findings from the logistic regression analysis are included in Tables 4 and 5 (for Repeated Question Administration and Language, respectively) in the back of this report.

Repeated Question Administrations
The trend for every question that we analyzed was that the question was asked exactly or with slight changes more often the first time than later times it was asked. When respondent behavior was classified as "good" at a higher rate on the first administration than later administrations, it seemed often to be due to an increased number of inaudible responses for questions about later persons in the household (data not presented in this report). As the interview progresses, respondents may become more likely to nod or shake their heads to answer yes/no questions rather than verbalizing an answer.[17] This will be discussed further at the question level. Based on the logistic regression analysis, we found significant effects of repeated question administrations on good interviewer behavior for the questions on Age (#13), Hispanic origin (#17), Race (#20) and Coverage2 (#24); on good respondent behavior for the Hispanic origin follow-up question (#18); and on good final outcome behavior for the questions on Hispanic origin (#17) and its' follow-up (#18; see Table 4). Question-level analyses of these differences will be presented in the section pertaining to each question.

Interview Language
The trend for language was that questions in English were more often administered correctly than those in Spanish. This trend was evident in each of the 14 questions of substantive interest. There was no overarching trend of language effects on respondent behavior. Based on the logistic regression analysis, we found significant effects of language on good interviewer behavior for the question on Date of birth (#14) and on good respondent behavior for the question on Race (#20). Analyses for these questions will be presented in the question-level section.

---

[17] Inaudible responses are more frequent for repeated administrations for every question (data not presented in this report).

C. QUESTION-LEVEL ANALYSIS AND RECOMMENDATIONS

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 question on race and includes an introductory statement that was carried over from the paper form designed to reduce non-response by Hispanics to the Race question. The introduction and Hispanic origin question were scripted as follows:

> **Question 17. Hispanic Origin**
>
> **"I am going to ask you two questions, one on Hispanic Origin and one on race. Please answer both questions. [Are you / Is NAME] of Spanish, Hispanic or Latino origin?"**

Interviewers were instructed to read the introductory statement (i.e., the first sentence) the first time it is administered in the household, but it was optional for later household members. This is the rule we instructed the coders to use when judging correct reading or major change.

Interviewer behavior
Overall, the Hispanic origin question was asked exactly as scripted or with minor changes only 33 percent of the time (see Table 2). This did differ by repeated administration (see Table 4). It was asked exactly or with minor changes 60 percent of the time for Person 1 and 22 percent of the time for later persons in the household (data not presented in this report; Table 4 shows data aggregated over exact wording/slight change and positive verification). It was skipped only 3 percent of the time for Person 1, but it was skipped 27 percent of the time for later people (data not presented in this report). Positive verifications were also slightly higher for later persons in the household than for Person 1 (data not presented in this report). This indicates that interviewers may have been using the response for Person 1 and applying it to all household members, either because the respondent told them to do so, or because they thought it was appropriate to do it on their own.

There was a major change in 42 percent of the cases overall (see Table 2). Review of the notes provided by the coders indicates that in almost half of the major change cases (47%; N=153), the interviewer used only one or two of the terms to ask the question. This could have been caused by the tendency for respondents to interpret the question as a request to choose one (Spanish, Hispanic, or Latino origin) rather than to answer with "yes" or "no." By giving the respondent only one term, it becomes clearer that it is a yes/no question. An alternative explanation is that interviewers find the three terms redundant, or they believe that respondents will identify with one or two of the terms more than the last.

In 16 percent of the major change cases, notes show that the interviewer changed the question by adding examples in addition to or instead of those that were scripted in the question. These were usually asking about countries that indicate Hispanic origin (e.g., "Are you Mexican?").

21

Other major changes included omitting the introductory phrase for the first person, phrasing the question negatively (e.g., "You are not Spanish, Hispanic or Latino, are you?), asking explicitly for nationality and simply asking if the next person in the household was the same as the last.

Respondent behavior

We trained the coders to assess responses based on whether or not they met the measurement objective of the question. Thus, several types of responses were considered adequate for coding purposes. "Yes/no" responses were allowed, as were responses taken directly from the origin categories offered in the question itself (e.g., "I'm Latina" or "I'm Spanish"). Responses were also considered adequate if they represented one of the more detailed origins available in the follow-up screens (e.g., Columbian, Mexican, etc.). This provides a conservative estimate of inadequate responses for several reasons. First, it does not address the issue of respondents misinterpreting the question as a "pick one," rather than a "yes/no." Secondly, it relies on the interviewer to recognize that the country mentioned as the origin is, in fact, one of the acceptable Hispanic origin countries (the coders had this information available to them in the coding booklet, interviewers did not have this information immediately available).

Overall, respondents gave codeable answers to this question 74 percent of the time on the first exchange (see Table 2). This did differ by repeated administration (see Table 4). Person 1 had codeable answers 82 percent of the time, and later persons had codeable answers only 69 percent of the time (see Table 4). The difference here is not due to a large difference in inadequate responses, but rather, due to a much higher percentage of inaudible responses for later persons in the household (data not presented in this report, but note the overall 11% I/O in the respondent behavior column in Table 2). As mentioned before, this could be because it is a yes/no question and during the interview the respondent becomes more likely to nod or shake his or her head rather than verbalizing an answer.

Behavior coder notes indicated that the inadequate responses to this question were generally races or non-Hispanic countries of origin.

Follow-up Questions to Hispanic Origin

There were two follow-up questions to the Hispanic origin item. The goal of these questions was to get countries of origin for Hispanics. The first follow-up question offers the most common countries of origin, and the second gathers any other country of origin. The text associated with the follow-up questions is found below.

The first follow-up was to be asked if respondent said "yes" to the initial Hispanic origin question:

> **Question 18. Hispanic Yes**
>
> **"(Are you / Is NAME) Mexican, Mexican American, Chicano? Puerto Rican? Cuban? Another Spanish, Hispanic or Latino origin? (For example, Argentinean, Columbian, Dominican, Nicaraguan Salvadoran, Spaniard, and so on.)"**

The second follow-up was to be asked if the respondent said "other" to the first follow-up question:

Question 19. Hispanic Other

"What is the name of the Spanish, Hispanic or Latino origin? (For example, Argentinean, Columbian, Dominican, Nicaraguan Salvadoran, Spaniard, and so on.)"

These questions are difficult to code because it is unlikely that the interviewer would need to read all three questions to get a detailed country of origin from the respondent. It is more likely that the respondent would offer the country to the first or second question. Because the last follow-up question was administered relatively few times, we did not conduct a detailed analysis of it. We will however, briefly look at the first follow-up question.

In 45 percent of all cases, the interviewer made a major change to the Hispanic origin follow-up question (see Table 2). These major changes, as noted by the coders, were often in the form of asking from which country the person came, or for his or her nationality. In other cases, interviewers read only a few of the countries in the question or asked if Person 2 and later were of the same origin that Person 1 was. Verification was common for this question for this reason. In 15 percent of cases, respondents had already offered country of origin prior to hearing this question (see Table 2, V+). Positive verification was present more often for Person 2 and later in the household, indicating that the respondent had told the interviewer that everyone (or at least a subset of the household) was from the same country of origin (data not presented in this report).

Respondents were able to give an adequate response to this question 74 percent of the time overall (see Table 2). There was a difference by repeated administration for both respondents' first interaction and for the outcome (see Table 4). This appears to be due to an increase in inaudible responses for Person 2 and later, rather than a difference in inadequate answers (data not presented in this report). Again, this may be due to the respondent nodding or shaking his or her head more often later in the interview.

Recommendation(s) for Hispanic Yes (#18) and Hispanic Other (#19)
Consider collapsing the two follow-up questions to the Hispanic origin question by asking simply for country of origin.

Race Question

The question on race was scripted as follows:

Question 20. Race

"Using this list, please choose one or more races that (you / NAME) consider(s) (yourself / himself / herself) to be."

Interviewer Behavior

Overall the Race question was asked exactly as worded or with minor changes only 16 percent of the time (see Table 2). There was an effect of repeated administration on the correct administration of this question (see Table 4). For Person 1, it was asked exactly as worded or with minor changes 36 percent of the time (data not presented in this report; Table 4 shows data aggregated over exact wording/slight change and positive verification). For later household members, it was asked exactly as worded or with minor changes only 8 percent of the time. This question was skipped very often – in fact, it was the second most frequently skipped question that we analyzed (see Table 2). For Person 2 and later in the household, the Race question was skipped 49 percent of the time (data not presented in this report). 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). Overall, 18 percent of the time race was verified by the interviewer[18] (see Table 2). This breaks down to a 10 percent verification for Person 1's race and a 20 percent verification of Person 2 and later's race (data not presented in this report). This further supports the hypothesis that the interviewer uses Person 1's answer to the Race question to infer information about the race of other household members.

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 themes:

Table 6. Themes in Major Changes to Question Wording for Race
as Documented in Behavior Coders' Notes

Interviewer Behavior	Percent of All Major Change
1. "Do you consider (him/her/yourself). . .	33%
2. "Using this list, what would you consider. . ."	14%
3. Adding races not on list to question	14%
4. Asking single race only to respondent	14%
5. "Of what race. . ." or "Is the race. . . "	10%
6. Showed/read list only	7%
7. Asking for nationality	4%
8. Other	4%
Total	100%

Perhaps the most surprising finding was that in 14 percent (N=29) of these cases the interviewer offered the respondent races that were not on the list of codeable races. These often included Hispanic options, such as Mestizo, Trigueno, Hispanic or Latino. Presumably, this reflects the interviewers' desire to overcome the problems that Hispanic respondents have with the Race

---

[18] The verification statistics presented here collapse positive and negative verification. For the race item, it is the understanding of the researchers that race should always be asked and not verified. For this reason, it seemed adequate to collapse positive and negative verifications for the discussion of this item.

24

question. However, it leads the respondent to answer with an uncodeable response when they might otherwise have chosen an adequate response.

Respondent Behavior
We allowed adequate answers to the Race question to include races listed on the main screen of the Race question as well as races that were codcable from the follow-up screens (see entire lists in the Question Guide in Appendix A). For example, Fijian is not a response option on the main screen, but it can be acccssed on the "Other Pacific Islander" follow-up screen. We did not make a distinction between answers that were codeable from the first screen versus answers that were codcable from later Race follow-up screens. This results in a conservative estimate of inadequate answers because it presumes that interviewers know how to code a race that is more detailed than those available on the first Race question screen (again, the coders were provided with this information in their question guides).

Respondents gave an adequate answer to the Race question 42 percent of the time on the first exchange (see Table 2) and ended up with a codeable answer in almost half the cases by the final exchange (see Table 3). When respondents did not give an adequate answer on the first exchange, they often gave an answer that simply was not codeable based on the response options offered by the instrument (see Table 2).

The inadequate responses can be classified as follows:

Table 7. Themes in Inadequate Responses to Race as Documented
      in Behavior Coders' Notes

Response	Percentage of All Inadequate Responses	
1. Variant of Hispanic origin	59%	
a. "Triguena"		19%
b. "Mestiza"		13%
c. "Hispanic"		8%
d. "Latin"		6%
e. "Spanish"		5%
f. "Puerto Rican"		4%
g. "Ecuadorian"		2%
h. "Dominican"		2%
2. "Mixed"	11%	
3. Skin color (e.g., dark-skinned)	11%	
4. Actually stating "not on list"	7%	
5. Other	12%	
Total	100%	

The most striking finding here is that in over half of the cases where a respondent did not provide an adequate response, the respondent answered the Race question with a Hispanic origin. In most of these cases, the interviewer did not persist to negotiate an adequate response. From the behavior coding data, we cannot determine how the interviewer recorded the race, but the fact that the interviewer did not actually gather a codeable race from respondents half of the time indicates reason for concern. If the interviewer is choosing a race for the respondent, then it is not reflective of the race that the person considers him or herself to be. If the interviewer leaves the question blank or enters "don't know" or "refuse," then we are left with missing data that will need to be imputed.

There was an effect of language on respondent behavior (see Table 5). In 50 percent of English interviews, respondents gave a codeable response. However, in only 34 percent of interviews in Spanish did respondents give a codeable response. This is most likely due to the fact that over half of all uncodeable responses were describing Hispanic origin as a race.

In about 5 percent of cases, respondents gave a qualified answer (see Table 2). The most common qualified answer was "White Hispanic."[19] These respondents were telling the interviewer that they were not simply white, but white Hispanic. While this does not necessarily indicate a problem with entering data for the interviewer, it does indicate a problem for the respondent in answering the question.

We were not able to analyze the follow-up questions to the Race question because those questions were administered fewer than 20 times. Very few observations for American Indians, Other Asians, and Other Pacific Islanders were captured.

Recommendation(s) for Race (#20)
In recognition of the fact that 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 suggestions for question improvement. These data suggest, however, the artificial separation of ethnicity concepts into two distinct questions causes quite a bit of confusion and response burden for those of Hispanic descent when asked to answer the Race question. These behavior coding results suggest people of Hispanic origin often identify their race as Hispanic.

Coverage Questions

Two coverage questions were embedded in the NRFU instrument, one toward the beginning of the instrument, Coverage1 (#7), and one at the very end, Coverage2 (#24). The Coverage1 question is asked only once for the entire household, while the Coverage2 question is asked for each person listed in the roster.

The Coverage1 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 (the question uses the term "permanent residents" to convey this concept):

---

[19] This was considered qualified because they provided more information than a simple codeable race.

Question 7. Coverage1

"Besides the name[s] you gave me earlier, were there other people who lived or stayed at this place part of the time but were not permanent residents? For example, live-in employees or children in joint custody?"

The Coverage2 question appears just after the Race question and its associated follow-up questions. This 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 reference date, in addition to collecting information about the reason for this behavior:

Question 24. Coverage2

"[Do you / Does NAME / Did NAME] sometimes live or stay somewhere else?

   To attend college?
   To stay at a seasonal or second residence?
   To be closer to work?
   For a child custody arrangement?
   For any other reasons?"

The behavior coding results indicate both questions suffered from major changes to question wording, but Coverage2 faired far worse. Table 2 indicates Coverage2 was read with major changes to question wording over half the time (66%), while Coverage1 was read with major changes almost one third of the time (31%). In addition, both questions were omitted by interviewers a lot more than we would have liked (around 18% for each), and in Coverage2 this tended to happen when it should have been asked for Person 2 and later—after it had already been posed once to the respondent. Interviewer and respondent behavior did not differ significantly between the English- and Spanish-language versions of the survey (see Table 5). There was, however, a significant effect of repeated administration on desirable interviewer behavior in the Coverage2 question; during subsequent readings of this question, ideal interviewer behavior decreased and non-ideal interviewer behavior increased (see Table 4).

Despite aspects of the coverage questions that produced serious administration difficulties for interviewers, the behavior coding data did not detect any serious response issues (see Table 2). Although both questions experienced respondent interruptions that exceeded 10 percent, they still evoked responses at the first-level of exchange that met the questions' objectives most of the time (91% for Coverage1 and 85% for Coverage2). Table 3 indicates these percentages further increased after subsequent negotiation between interviewers and respondents occurred (95% and 89% for Coverage1 and Coverage2, respectively). Of those responses that were not coded as adequate, a relatively high percentage were inaudible (i.e., 3% for Coverage1 and 8% for Coverage2), presumably due to the respondent nodding or shaking his or her head. These results suggest there was a lack of overt response problems, and therefore, the remaining discussion for these questions will focus primarily on interviewer behavior and question administration issues.

Coverage1
Of the 219 households interviewed, Table 2 shows interviewers read the Coverage1 question exactly as worded or with minor changes only half of the time (51 %). A substantial amount of undesirable interviewer behavior was evident; the question suffered from both major changes to question wording (31%) and omission from the survey (18%).

According to the coders' notes, major changes to question wording were mostly due to the omission of the question's main conceptual phrase "…part of the time but were not permanent residents?" Omitting this main concept essentially changes the question's intent to "Does/Did anyone else live here?" and does not communicate to respondents that the question intends to capture people who might stay there on an impermanent basis and may not be considered as part of the household's usual makeup. Of the 67 observations for which major changes occurred, the question's main concept was omitted from the question 60 percent of the time (See Table 8 below). When interviewers read the phrase containing this main concept, but still changed the wording significantly, they generally left off the last statement containing the illustrative examples (19%).

Table 8. Themes in Major Changes to Question Wording for
Coverage1 as Documented in Behavior Coders' Notes

Interviewers' Question-Asking Behavior	Percent Major Change
Paraphrased, but read main concept	
Read 1st sentence only	19%
Left off portion of 2nd sentence	10%
Paraphrased and omitted main concept	
Read 1st and 2nd sentence	48%
Verified	12%
Other	10%
Total	100%

Interviewers accomplished the omission of the main concept in two ways: 1) the question was read as worded, except for the main conceptual phrase (48%); and 2) interviewers paraphrased the entire question by attempting to verify that no one else lived in the household, for example, "And there was no one else living here then?" (12%). A few times, interviewers even tacked on an "April 1st" reference date, changing the concept of the question from a general period of time to one particular day.

By omitting this main conceptual phrase, interviewers were asking quite a different question: "Anyone else you forgot to mention that is part of the household" versus "Are there people who aren't part of the household typically, but sometimes stay here?" Even though behavior coding

results did not suggest there were response issues, clearly some respondents were asked quite a different question than was intended, which may affect data quality.

Respondent interrupted interviewers 11 percent of the time this question was read (see Table 2). Of the 20 observations, 19 contained detailed notes for analysis. In most of these cases, respondents interrupted the interviewers as they read the first sentence; some of the other interruptions occurred during the second statement containing the illustrative examples. From past survey design research, we know that the end of a question communicates to respondents that it is time to provide a response, and respondents tend to interrupt when the main question is followed by additional information (e.g., includes/excludes, or as in this case, illustrative examples). This suggests that the structure of Coverage1 is causing respondents to interrupt the full question reading.

Recommendation(s) for Coverage1 (#7)
Interviewers may be omitting the main conceptual phrase (i.e., "…stayed part of the time but were not permanent residents") in an attempt to shorten this long question, without realizing they are changing the question's intent. To solve this problem, it would be helpful to rephrase the question to make it less convenient for interviewers to edit out the main concept. This could be accomplished any number of ways, but may make the most sense to switch the order of the conceptual phrase's subparts (i.e., "stay here" and "but do not stay here all the time"). For example:

> **Besides [you / the name(s) you just gave me], is there anyone else that had another place to stay, but stayed here often?**

Coverage2
Interviewers had great difficulty administering the Coverage2 question correctly for all persons listed on the roster (see Table 2). Of the 778 observations pertaining to this question, exact readings were extremely infrequent (15%); it was more often read with major changes (66%). In addition, the logistic regression identified an effect of repeated administration on ideal, or "good," interviewer behavior. The first time this question was asked in each household, ideal interviewer behavior (i.e., exact reading, positive verification) occurred 27 percent of the time, but dropped significantly to 12 percent for all subsequent administrations of this question (see Table 4). The main shift in behavior attributable to this finding seems to be interviewers' tendency to skip this question instead of asking it for each remaining member of the household.

The Coverage2 question was one of the four questions interviewers read with major changes over 60 percent of the time (see Table 2). When interviewers made major changes, behavior coders reported that interviewers usually skipped the response options altogether, a finding that is consistent across interview languages. Analysis of the 512 interactions for which detailed notes were available indicates omission of the entire response set occurred 70 percent of the time there was a major change. For the remaining cases (28%), interviewers either read one or two response options, or they referred to them in some manner that was not exactly the way they were scripted. When interviewers did refer to the response options, they often mentioned only one or two and they generally did not read them verbatim.

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 at that age. For example, the interview is scripted to ask all people, even children and the elderly, if they attended college and to ask everyone, even adults, if they were in joint custody arrangements. Interviewers may have tailored the question by withholding response options they felt did not apply for particular household members. In other cases, interviewers may have treated response options as illustrative examples they could choose from in order to shorten the question, providing one or two that seemed most widely applicable in order to provide enough information to respondents. The two most commonly offered were those relating to college and work.

A common interviewer strategy for shortening this question was achieved by marrying the question stem with one or two response options, which effectively eliminated the question at the end of the question stem. For example, interviewers asked, "Did you live somewhere else to attend college or to be closer to work?" Another similar strategy emerged as interviewers asked the question in the negative (e.g., "So she doesn't live anywhere else?" and "You don't stay some place else to go to school or for work?"). In many cases interviewers seemed to be verifying this information, though it is difficult to know whether sufficient information had previously been provided by the respondent (i.e., "She didn't live anywhere to go to school or anything, right?" and "He didn't live anyplace else, did he?" and "So your wife stays here all the time too, right?").

Paraphrasing was also evident within the question stem. Interviewers commonly dropped "sometimes" and often mentioned either "live" or "stay," not both as the question is scripted. In addition, the phrase "somewhere else" was often substituted with phrases such as "some place else," "another place," or "anywhere else." Modifications to the stem of the question, however, may be less problematic for data quality compared to the omission of some or all of the response options.

Respondents' reactions to this question may have caused interviewers to take shortcuts with question wording, if they were not already doing so. Table 2 shows respondent interruptions were greatest for this question compared to all other questions (13%). It is interesting to note that all of the interruptions occurred the first time this question was posed to respondents, which probably encouraged interviewers to change or skip this question for subsequent administrations (data not presented in this report). From previous survey methodology research, we know that interruptions are predictable when the question stem itself poses a question before any response options are provided. Seventy-five out of the 102 interactions for which detailed notes were available reveal that almost half (47%) of the interruptions occurred at the conclusion of the question stem (See Table 9 below).

Table 9. Point of Respondent Interruption in Coverage2 as
Documented in Behavior Coders' Notes

Point of Interruption	Observations N=75	Cumulative Percent
Question stem		
Mid-sentence	6	8%
End of sentence	29	47%
Response options		
After/during 1st	13	64%
After/during 2nd	14	83%
After/during 3rd	9	95%
After/during 4th	4	100%

In addition to changing question wording, Table 2 indicates interviewers sometimes skipped the question entirely (18%). This behavior rarely occurred the first time the question was to be administered; in almost every case, interviewers omitted this question when it should have been repeated for other household members (see Table 4).

The Coverage2 question contains two question design features that are likely driving the high rate of unacceptable interviewer behavior and respondent interruptions The current question stem contains a fully formed, complete question, and it is followed by a response set (required reading for interviewers) that contains a second, separate question. After hearing the first portion of the inquiry in the question stem, respondents were aware of the question's intent and provided a response—hence the high rate of interruption at the question stem's conclusion. It may have seemed pointless to interviewers to continue reading the response options when an acceptable response had already been given, which may be the reason for the high rate of interviewers omitting the response set altogether. Still other interviewers, to avoid this awkwardness, found a way to paraphrase the question by incorporating one or two response options directly into the question stem itself. And finally, interviewers avoided asking this question for the remaining persons listed on the roster.

Recommendation(s) for Coverage2 (#24)
These undesirable interviewer behaviors exposed respondents to nonstandardized questions, which could potentially affect the response distribution. Adopting an alternative question structure and administration procedure would most likely encourage a more standardized question administration:

- Incorporate the first response option into the question stem; and
- Administer this question in a topic-based manner (i.e., ask the first part of the question of everyone on the roster, then ask the next portion of the question and cycle through the roster, and so on).

- Use age to filter these questions to only ask them of those people to whom they might pertain (e.g., ask the college question only to college age people, ask the joint custody question only to children)

An automated instrument like the NRFU could easily handle this type of question administration and data collection, would be more likely to result in standardized delivery of complete question stimulus to all respondents, and could look like this:

> (For ages 17-24)
> **"Do you sometimes live or stay somewhere else to attend college?**
> > **How about Manuel?**
>
> (For all ages)
> **Do you stay at a seasonal or second residence?**
> > **How about Manuel?**
>
> (For ages 16 and older)
> **Do you stay somewhere else to be closer to work?**
> > **How about Manuel?**
>
> (For ages 18 and younger)
> **Do you stay somewhere else for a child custody arrangement?**
> > **How about Manuel?**
>
> (For all ages)
> **Do you sometimes live or stay somewhere else for any other reason?**
> > **How about Manuel?"**

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. This question suffered from major change more than half of the time (see Table 2). It is intended to be read as:

Question 2. Usual Residence

"Is this house/apartment/mobile home the usual residence of someone in your household, or is it a vacation home, seasonal home, or held for occasional use?"

This question was administered with a major change in 63 percent of all cases (see Table 2). Examining the notes revealed that 64 percent of the major change cases were because the interviewer only asked if it was the usual residence – making it a yes/no question instead of a choose one.[20] They left out the options of vacation home, seasonal home or held for occasional

---

[20] Other major change reasons were asking if it was for personal use, if it was their permanent residence and omitting the reference to "someone in your household." These were less frequent than a simple shortening of the question, however.

use. Interviewers may have done this because they felt that the other options were not necessary or that they applied so infrequently that they were not useful.

This question also elicited a high rate of inadequate answers at the first level of exchange (i.e., 21%, see Table 2). Respondents indicated that the house or apartment was their "regular" or "permanent" residence or sometimes answered a question other than what was intended – by letting the interviewer know that they own or rent the place. In other cases, respondents misinterpreted the request for them to pick one of the options and answered "no." Interviewers were able to resolve the problem with the respondent and gather an acceptable response in about on third of these cases, but the problem persisted for about 13% of respondents (see Table 3). This is further support that the question is not working as intended. Even when respondents hear the question as it is scripted, they can not always answer properly without some additional help from the interviewer.

> ### *Recommendation(s) for Usual Residence (#2)*
> These findings indicate a problem with the question wording that should be further explored through revisions and cognitive testing of those revisions. Leaving the question as it is will result in interviewers changing it on their own without the benefits of subject matter and questionnaire design expertise. A revision of this question would likely improve respondents understanding as well.

Gathering the Roster (Names)

The question used to gather the household roster has two purposes. The first is to gather all of the names of all of the people living or staying in the housing unit. The second is to gather name of one of the owners or renters as Person 1. Household relationships are recorded by each person's relationship to Person 1. In census data analysis, family or non-family households are derived by these relationships. 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). This question was also read with a major change more than half of the time (see Table 2). It is scripted as follows:

Question 4. Names

"What is the name of each person who lived or stayed at this residence on April 1, 2004? Start with the name of one person who owned or rented this house/apartment/mobile home on April 1, 2004?"

This question was administered with a major change in 61 percent of cases. Of the major change cases, 39 percent (N=52) of the time behavior coders reported that the interviewer asked the respondent to start with him or herself rather than asking for the name of a person that owns or rents the home. In 11 percent of the major change cases, the interviewer did not provide instruction for whom to start with.[21] In both instances, this could cause problems for determining

---

[21] Not all of the major changes for this question involved whom to start with. Many also involved leaving out one of the dates in the question or changing "one person who owned or rented" to "the person who owned or rented" which is a small change, but a major one because that would indicate that only one person was the owner or renter and may cause problems for the respondent.

relationships in the household because Person 1 may or may not be the owner or renter of the household.

The fact that in almost a quarter of all administrations, interviewers asked the respondents to start with his or her own name indicates that there may be a problem with gathering names in the way it is scripted. From other cognitive research, we know that respondents sometimes forget to list themselves when giving a household roster. Perhaps interviewers had noticed this, and attempted to remedy it by asking for their names first. Alternatively, it may just be that they find it easier and more conversational to start with the person with whom they are talking.

Recommendation(s) for Names (#4)
Some consideration should be given to how the roster is gathered. Perhaps it would be easier to start with the household member with whom the interviewer is speaking, then ask a separate question to determine the owner or renter to be used to gather household relationship information. This approach was recently adopted by the Census Coverage Measurement Person Interview.

Ownership Question

The question that asks if the housing unit is owned or rented suffered from the highest percentage of major change (67%, see Table 2). It is scripted as follows:

Question 10. 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 (without a mortgage or loan)?**
> **Rented for cash rent?**
> **Occupied without payment of cash rent?"**

Examination of the major change notes revealed the following patterns:

Table 10. Themes in Major Changes to Question Wording for
Ownership as Documented in the Behavior Coders' Notes

Interviewer Behavior	Percent of All Major Change
1. Asking only "rented" or "rented for cash rent"	52%
2. Only asking "owned or rented"	19%
3. Asking only "owned"	3%

In 35 percent of all administrations of this question, notes indicated that the interviewer only asked if the respondent rented – not giving them the explicit options of owning or occupying

34

without payment of rent. This may be due to the location – in the Queens test site many of the housing units were apartments, and interviewers may have had the experience that most people rented. The high level of major change in this question indicates that the question is awkward to be read completely as worded. It is, in fact, the longest question in this survey.

Recommendation(s) for Ownership (#10)
Given that the NRFU is an automated instrument, this question could be shortened into two questions, first gathering whether the place is owned, rented or occupied without payment and then whether it is owned with a mortgage or not. This is another question that should be considered for revision and cognitive testing. It is apparent that it is not working in the field and interviewers are changing the intent of the question without the benefit of subject matter or survey methodology expertise.

Relationship Question

The question that asks how each person is related to Person 1 had a rate of inadequate answers (i.e., 17%) that exceeded the threshold of unacceptable behaviors (see Table 2). The question wording and codeable response options are as follows:

Question 11. Relationship

Which of these categories best describes how [NAME] is related to [you / NAME]?
Husband or wife
Son or daughter
Brother or sister
Father or mother
Grandchild
In-law
Other relative
Roomer, boarder
Housemate, roommate
Unmarried partner
Foster child
Other nonrelative receiving formal care
Other nonrelative

There is a flashcard that displays all of the response options for this question that interviewers were instructed to use; however, we have no way of knowing from an audio-tape if it was used or not.

In some of the cases when respondents gave an inadequate answer, behavior coders reported that they told the interviewer the people were simply related or not related. This was sometimes due to the fact that the interviewer asked if the people were related or not related prior to (or instead of) asking the question as worded. In other cases, the interviewer asked the question as it appeared, but the respondent said the people were not related in response to the question itself. In most of these cases, the interviewer was able to negotiate an adequate response for the outcome

35

(see Table 3). In other cases of inadequate answers, respondents provided things like "friend" or "girlfriend," which can be coded if the interviewer takes the step of placing "friend" into the "Housemate, roommate" category and "girlfriend" into the "Unmarried partner" category. Other answers given by respondents are ones that would fit into the "Other relative" category, like "niece" or "nephew." These are cases where the interviewer has to use some judgment to make the classifications.

Recommendation(s) for Relationship (#11)

The suggestion we have for this question is to consider how it applies to unrelated people, particularly when the interviewer fails to use the flashcard. Without the flashcard, the question as it is currently worded can elicit the response "they are not related," which is a response that cannot be field coded without additional probing by interviewers (e.g., "Is John your roommate or unmarried partner?"). Alternate question wording should be tested that avoids the presumption that the two people are related. Of course any new question wording should be cognitively tested prior to being implemented in the field.

Sex Question

The question used to determine gender of each household member also demonstrated significant problems.

Question 12. Sex

"[Is NAME /Are you] male or female?"

The question on sex was skipped 48 percent of the time and verified 28 percent of the time (see Table 2). It was only actually asked as worded 21 percent of the time. Interviewers verified sex in another quarter of all administrations, which indicates that they have received information earlier in the interview to let them know what the sex of the person may be. Interviewers may have felt that the answer to this question was obvious and/or it was insulting to the respondent for them to ask.

Recommendation(s) for Sex (#12)

Perhaps an instruction permitting the interviewer to verify sex would decrease the number of times that it is skipped entirely.

Age

The question designed to gather age as of Census day for each of the household members demonstrated an interesting finding.

Question 13. Age

"What was [your/ NAME's] age on April 1, 2004? (If you don't know the exact age, please estimate.)"

Respondents had to request clarification in 7 percent of all administrations of the age question (see Table 2). This is interesting because it is among the highest percentages of request for clarification and it is a question that should be easy to understand. We hypothesize the confusion stems from asking age as of a certain date – something that people are not used to answering. Eighty-six percent of people were able to provide an age, but some of these people needed "help" answering a relatively easy question because it was asked in a way that people are not used to hearing.

The other interesting finding in the Age question is that it was asked more often as intended for Person 1 than for later persons in the household (62% versus 45%; see Table 2). This was due to a difference in major change for Person 1 and later persons (33% versus 50%; data not presented in this report). From the notes, seems that for Person 2 and later, the interviewer was more likely to leave the date off the question. This is further evidence that asking age as of a specific date in the past is troublesome to interviewers.

> ### *Recommendation(s) for Age (#13)*
> Since the NRFU is an automated instrument, if we gathered date of birth before age, we could calculate and verify age as of today, rather than as of a date in the past. This would be a cognitively easier task for respondents, while not losing any data (i.e., we could still calculate age as of Census day for our records). We realize and acknowledge that it is necessary to ask age as of Census day in the absence of date of birth. In cases where date of birth is not known, the question would need to remain as it is. This revision, however, would simplify the interview for all cases where date of birth is known.

Date of Birth

The question asking for date of birth of each household member demonstrated some additional interesting findings.

> ### Question 14. Date of Birth
>
> ### "What is [your / NAME's] date of birth?"

Date of birth was asked accurately more often in English than in Spanish (see Table 5). It was asked about half the time correctly in English, but only about a third of the time in Spanish. It is not clear why this was the case. In both languages, coders noted that the reasons for major change varied from asking for birthday to requesting that the respondent begin with the month to prefacing the question with "do you know. . ."

Date of Birth had the highest rate of "don't know" responses. Fourteen percent of the times in which it was asked, the respondent replied with a "don't know" (see Table 2). This could indicate that we are asking for information that respondents simply do not have for other household members or it could be an indication that they are not willing to provide it.

> ### *Recommendation(s) for Date of Birth (#14)*
> None.

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 some cause for concern over data quality.

The results from behavior coding suggest that almost all of the survey questions in the automated NRFU instrument suffered from administration difficulties at rates that were surprisingly high. Interviewers tended to change the survey questions, sometimes quite dramatically, or omit them entirely—omissions seemed most problematic for questions repeated at the person level. Response issues were less dramatic across the board, but did tend to plague particular questions, most notably Race.

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. Respondents are also having difficulty providing adequate answers to some of the questions. In order to resolve these issues, revision and further pretesting efforts are warranted. Iterative rounds of cognitive interviewing are recommended for future versions of these questions.

These behavior coding results make clear that the Spanish version of the questions did not function as well as their English counterparts, 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.

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.

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. 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 Fransisco, CA: Jossey-Bass.

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

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.

Zukerberg, A., Von Thurn, D., and Moore, J. (1995) "Practical Considerations in Sample Size Selection for Behavior Coding Pretests." Paper presented at the American Association for Public Opinion Research Annual Meeting, Orlando, FL. U.S. Census Bureau, Washington, DC.

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

Question	Interviewer Behavior[1]							Respondent Behavior[2]										Break In[3]
	N	E/S	MC	V+	V-	I/O	Skip[4]	N	AA	IA	UA	QA	CL	RR	DK	R	I/O	
1 Live Here	218	70.6%	25.7%	0.0%	0.5%	0.0%	3.2%	212	81.1%	5.7%	0.5%	2.4%	8.0%	0.0%	0.5%	0.0%	1.9%	1.4%
2 Usual Res	219	32.0	62.6	0.0	0.0	0.0	5.5	205	67.8	20.5	1.0	2.4	3.4	1.0	0.0	0.0	3.9	5.9
3 POP count	217	56.7	39.2	1.8	0.0	0.0	2.3	213	88.7	1.4	11.4	3.3	3.3	0.0	0.0	0.0	1.9	4.2
4 Names	216	38.0	61.1	0.0	0.0	0.0	0.9	211	77.3	11.4	1.4	0.9	8.5	0.0	0.0	0.5	0.0	3.8
7 Coverage1	219	51.1	30.6	0.0	0.0	0.5	17.8	179	90.5	1.1	0.6	1.7	1.1	1.1	0.0	0.0	2.8	11.2
10 Ownership	219	28.8	66.7	0.0	1.4	0.0	3.2	212	75.9	11.3	1.4	4.2	2.8	0.0	0.0	0.0	4.2	10.8
11 Relatn	564	43.4	33.2	15.6	2.0	0.5	5.3	530	70.8	17.0	1.3	2.6	4.0	0.0	0.0	0.2	4.2	3.0
12 Sex	776	21.0	2.3	28.1	0.1	0.3	48.2	395	58.7	8.4	0.5	0.0	0.8	0.0	0.0	0.3	31.4	3.0
13 Age	776	47.9	45.5	1.0	0.3	0.3	5.0	739	70.8	3.7	6.4	4.5	7.0	0.3	6.2	0.3	0.9	1.9
14 DOB	764	41.4	44.5	2.5	0.3	1.7	9.7	699	72.4	5.4	3.7	0.1	1.6	0.0	14.0	0.6	2.1	1.7
17 Hisp	780	32.6	41.5	4.4	0.8	0.3	20.5	614	73.5	8.1	0.3	5.4	1.0	0.5	0.3	0.0	10.9	3.4
18 Hisp Yes	379	19.0	45.4	15.3	4.0	1.1	15.3	321	73.8	5.3	0.3	1.6	0.6	0.0	0.6	0.0	17.8	11.2
19 Hisp Other	31	22.6	25.8	19.4	12.9	0.0	19.4	24	79.2	12.5	0.0	0.0	0.0	0.0	0.0	0.0	8.3	4.2
20 Race	789	15.8	26.2	4.6	13.1	0.4	39.9	471	42.3	35.5	2.5	4.7	3.0	0.0	1.3	0.0	10.8	3.0
21 Amer Ind	3	0.0	33.3	0.0	0.0	0.0	66.7	1	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0
22 Asian	14	0.0	0.0	50.0	0.0	0.0	50.0	7	85.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.3	0.0
23 Pacific	1	0.0	100.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24 Coverage2	778	14.9	65.8	0.6	0.5	0.4	17.7	635	84.9	2.5	0.9	1.9	1.4	0.2	0.5	0.0	7.7	13.1

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

[2] AA = adequate answer; IA = inadequate answer; UA = uncertain answer; QA = qualified answer; CL = clarification requested; RR = question reread; D = 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.

[4] "Skip" = interviewer omitted/skipped the question when it should have been read.

Table 3. Percent Final Response Outcome and Negative Reaction for All Interviews (English and Spanish) by Question

| Question | N | Final Response Outcome[1] | | | | | | | Negative Reaction[2] |
		AA	IA	UA	QA	DK	R	I/O	
1 Live Here	212	93.9%	2.4%	0.0%	0.9%	0.5%	0.0%	2.4%	0.5%
2 Usual Res	203	83.7	13.3	0.0	1.0	0.0	0.0	2.5	0.5
3 POP count	212	95.3	0.5	0.0	2.4	0.0	0.0	1.9	0.0
4 Names	210	94.8	3.3	0.5	0.5	0.0	1.0	0.0	0.0
7 Coverage1	177	94.9	1.1	0.0	1.1	0.0	0.0	2.8	0.0
10 Ownership	212	85.4	10.4	0.0	0.9	0.5	0.0	2.8	0.0
11 Relatn	533	83.9	10.9	0.0	1.1	0.0	0.2	3.9	0.2
12 Sex	392	62.8	6.9	0.5	0.0	0.0	0.8	29.1	0.5
13 Age	734	85.6	2.6	4.0	1.5	5.3	0.3	0.8	0.0
14 DOB	697	77.9	2.9	2.3	0.0	13.9	0.9	2.2	0.3
17 Hisp	613	76.0	7.3	0.2	5.2	0.3	0.2	10.8	0.3
18 Hisp Yes	320	77.8	3.4	0.3	0.9	0.0	0.3	17.2	0.0
19 Hisp Other	24	87.5	4.2	0.0	0.0	0.0	0.0	8.3	0.0
20 Race	462	49.1	33.8	0.4	4.5	0.9	0.0	11.3	0.2
21 Amer Ind	1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
22 Asian	7	85.7	0.0	0.0	0.0	0.0	0.0	14.3	0.0
23 Pacific	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24 Coverage2	630	89.2	1.0	0.3	1.4	0.5	0.0	7.6	0.0

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

[2] Rate calculated using the number of times the question was administered (i.e., from the interviewer behavior) as the denominator.

Table 4. Percent Good Interviewer Behavior[1] for Questions that Differed by Repeated Question Administrations

Question	Interviewer behavior		Respondent behavior		Final response outcome	
	Person 1	Person 2+	Person 1	Person 2+	Person 1	Person 2+
12 Sex	56.4%	50.8%	59.1%	58.4%	63.5%	62.1%
13 Age	61.8**	45.4	70.0	71.0	88.0	84.5
14 DOB	48.3	43.6	77.5	70.6	83.0	75.9
17 Hisp	61.7**	29.3	82.3*	69.0	85.9**	71.1
18 Hisp Yes	34.8	34.5	87.6**	68.0	93.3**	71.1
20 Race	40.8**	14.6	44.1	41.5	54.4	46.7
24 Coverage2	26.8**	11.9	84.8	85.7	92.8	88.2

* Significant difference at p < .01[2]

** Significant difference at p < .001

[1] Exact wording/slight change (Code E/S) and positive verification (Code V+) were considered "good" interviewer behavior. The only behavior that was considered good respondent behavior was an adequate answer (Code AA). Findings were very similar when only exact wording/slight change was used as good interviewer behavior. The differences were that there is an additional an effect of multiple administration on interviewer behavior for Hispanic Origin Follow-up (first time=30%, later times=14%, p<.001).

[2] We conducted a total of 42 tests (14 questions and 3 dependent measures). To ensure a study-wide significance level of .05, we recommend using a Bonferroni adjustment, which lead to a significance level of p< .001, which is what we used as a guideline for interpreting results. However, we also present results significant at the p< .01 level in the table that can be interpreted with some caution.

Table 5. Percent Good Interviewer Behavior[1] for Questions that Differed by Interview Language

	Interviewer behavior		Respondent behavior		Final response outcome	
Question	English	Spanish	English	Spanish	English	Spanish
1 Live Here	77.4%*	58.3%	81.5%	82.9%	94.6%	92.9%
2 Usual Res	32.1	32.0	70.6	63.2	84.0	82.1
3 POP count	63.9	48.7	88.3	89.2	93.0	98.6
4 Names	37.8	36.0	79.7	74.3	97.6	90.3
7 Coverage1	57.1	43.4	89.1	93.4	96.3	93.3
10 Ownership	29.8	28.6	72.1	80.8	82.2	89.0
11 Relatn	61.3	59.5	73.2	67.7	85.0	82.6
12 Sex	52.4	52.4	54.1	64.8	56.6*	70.6
13 Age	55.0*	43.0	72.5	68.2	88.3	81.3
14 DOB	51.5**	35.4	75.8	67.6	81.1	73.2
17 Hisp	41.9	33.2	69.6	79.2	71.4	83.1
18 Hisp Yes	41.1	29.9	76.9	73.2	80.5	77.4
20 Race	25.7*	17.3	49.8**	33.8	55.4*	42.8
24 Coverage2	18.3	13.0	86.3	83.8	90.1	88.9

* Significant difference at p < .01[2]

** Significant difference at p < .001

[1] Exact wording/slight change (Code E/S) and positive verification (Code V+) were considered "good" interviewer behavior. The only behavior that was considered good respondent behavior was an adequate answer (Code AA). Findings were very similar when only exact wording/slight change was used as good interviewer behavior. The differences were that there is an additional significant effect of language on interviewer behavior for sex (English=20%, Spanish=27%, p<.01) and Hispanic Origin follow-up (Hisp Yes; English=28%, Spanish=14%, p< .01) and the effect of language on interviewer behavior for race is no longer significant.

[2] We conducted a total of 42 tests (14 questions and 3 dependent measures). To ensure a study-wide significance level of .05, we recommend using a Bonferroni adjustment, which lead to a significance level of p< .001, which is what we used as a guideline for interpreting results. However, we also present results significant at the p< .01 level in the table that can be interpreted with some caution.

43

SURVEY QUESTION GUIDE

English

NRFU 2004

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 may use their discretion when reading text appearing in parentheses. It is not necessary for them to do so, but interviewers may read this text if they feel it is necessary.

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, 2004?
Acceptable Responses	Yes No
Comments	*Interviewer should read a street address for [ADDRESS]*
Skip Instructions	If yes, go to Q2 (USUAL RES). If no, end interview.

Question 2	USUAL RES
Screen	**Is this [house / apartment / mobile home] the usual residence of someone in your household, or is it a vacation home, seasonal home, or held for occasional use?**
Acceptable Responses	Usual residence Vacation home Seasonal home Held for occasional use
Comments	
Skip Instructions	Go to Q3 (POP COUNT).

Question 3	POP COUNT
Screen	**How many people were living or staying in this [house/ apartment / mobile home] on April 1, 2004?**
Acceptable Responses	Number 1-49
Comments	*May read flashcard aloud - this would still be an exact reading, as long as question is worded as written above.*
Skip Instructions	Go to Q4 (NAMES).

Question 4	NAMES
Screen	**What is the name of each person who lived or stayed at this residence on April 1, 2004? Start with the name of one person who owned or rented this [house / apartment / mobile home] on April 1, 2004.**
Acceptable Responses	First name, (Middle Initial), Last name
Comments	*R should give first and last name, but middle initial and suffix are acceptable too. If R only gives first name, it is an inadequate answer.* *Code only for first person that is listed.*

Skip Instructions	Go to Q5 (ANYONE_ELSE1).

Question 5	ANYONE_ELSE1
Screen	**Is there anyone else?**
Acceptable Responses	Yes No
Comments	*Code only the first time this is said, should be said after the first person is listed, or when R stops giving names.*
Skip Instructions	After you get a "no" to this question - If number given to Q3 (POP COUNT) does not equal the number of names given in Q5, go to Q6 (RECONCILE POP). Else, skip to Q7 (COVERAGE1)

Question 6	RECONCILE_POP
Screen	**Earlier I recorded that there were [POP COUNT] people living or staying at [ADDRESS]. Just now we listed [X] people.** **Which number is correct?**
Acceptable Responses	[POP COUNT] [X]
Comments	*This will only appear if the count given in Q3 (POP COUNT) is inconsistent with the number of names given. It should be skipped in most cases.*
Skip Instructions	Go to Q7 (COVERAGE1).

Question 7	COVERAGE1
Screen	**Besides the name[s] you gave me earlier, were there other people who lived or stayed at this place part of the time but were not permanent residents? For example, live-in employees or children in joint custody.**
Acceptable Responses	Yes No
Comments	*Note any comments here about reconsidering the answer to the previous question. Like – "Oh there's also Nanny. I should have said 5 people, not 4."*
Skip Instructions	If yes, go to Q8 (ADD PERSON). If no, skip to Q10 (OWNERSHIP).

Question 8	ADD_PERSON
Screen	**What is the name of the person who lived or stayed at this place but is not already on the list of names?**
Acceptable Responses	First name, (Middle Initial), Last name
Comments	*This will only be asked if the answer to Q7 (COVERAGE_1) is "yes".*

	Code only for first person listed. *R should give first and last name, but middle initial and suffix are acceptable too. If R only gives first name, it is an inadequate response.*
Skip Instructions	Go to Q9 (ANYONE ELSE2).

Question 9	**ANYONE_ELSE2**
Screen	**Is there anyone else?**
Acceptable Responses	Yes *No*
Comments	*This is only asked if Q8 (ADD_PERSON) is asked.* *Code only the first time this is said, should be said after the first person is listed, or when R stops giving names.*
Skip Instructions	After you get a "no," go to Q10 (OWNERSHIP).

Question 10	**OWNERSHP**
Screen	**Is [this / your] [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 (without a mortgage or loan)?** **Rented for cash rent?** **Occupied without payment of cash rent?**
Acceptable Responses	Owned with a mortgage or loan Owned free and clear Rented for cash rent Occupied without payment
Comments	*Interviewer must read all response options.*
Skip Instructions	Go to Q12 (SEX) for first person.

Starting here, all questions are asked for each person.

Question 11	RELATN
Screen	**FOR PERSON 1** **This Q is not read.** **FOR PERSON 2 +** **Which of these categories best describes how [NAME] is related to [you / NAME]?**
Acceptable Responses	Husband or wife Son or daughter Brother or sister Father or mother Grandchild In-law Other relative Roomer, boarder Housemate, roommate Unmarried partner Foster child Other nonrelative receiving formal care Other nonrelative
Comments	*This question will be skipped for the first person in the household.*
Skip Instructions	Go to Q12 (SEX).

Question 12	SEX
Screen	**[Is NAME / Are you] male or female?**
Acceptable Responses	Male Female
Comments	*This should be explicitly asked for each person.*
Skip Instructions	Go to Q13 (AGE).

Question 13	AGE
Screen	**What was [your / NAME=s] age on April 1, 2004?** **(If you don=t know the exact age, please estimate.)**
Acceptable Responses	Number 0-150
Comments	
Skip Instructions	Go to Q14 (DOB).

Question 14	DOB
Screen	What is [your / NAME=s] date of birth?
Acceptable Responses	Month, Day, Year
Comments	*Full date of birth is required for an acceptable response. Month and day is inadequate.*
Skip Instructions	If the calculated age is not the same as the reported age, go to Q15 (AGECHECK). Else, skip to Q17 (HISP).

Question 15	AGECHECK
Screen	**For the Census, we need to record age as of April 1, 2004. Based on the date I just entered, [your / NAME=s] age was [AGEC].** **Earlier I recorded [your / his / her / NAME=s] age as [AGE].** **Which age is correct as of April 1, 2004?**
Acceptable Responses	[AGE] [AGEC]
Comments	*This will only be asked if the reported AGE (Q13) is inconsistent with the age as calculated from the date of birth. This will be skipped for most Rs.*
Skip Instructions	If the answer corresponds to the given age in Q13 (AGE), go to Q16 (DOBCHECK). If the answer is the calculated age (AGEC), skip to Q17 (HISP).

Question 16	DOBCHECK
Screen	**Since [your / NAME=s] age as of April 1, 2004 was [AGE], can you help me correct [your / NAME=s] date of birth?** **I have [MM / DD / YYYY]. What should it be?**
Acceptable Responses	Month, Day, Year
Comments	*This is only asked when Q15 (AGECHECK) is asked, and the answer is the reported age instead of the age calculated from the date of birth. This will be skipped for most Rs.*
Skip Instructions	Go to Q17 (HISP).

Question 17	HISP
Screen	<u>FOR PERSON 1</u> I am going to ask you two questions, one on Hispanic origin and one on race. Please answer both questions. [Are you / Is NAME] of Spanish, Hispanic or Latino origin? <u>FOR PERSON 2 +</u> (I am going to ask you two questions, one on Hispanic origin and one on race. Please answer both questions.) [Are you / Is NAME] of Spanish, Hispanic or Latino origin?
Acceptable Responses	Yes No
Comments	*Interviewer should read optional intro statement for the first person. This is required for an exact reading for the first person, but is optional for an exact reading for later persons in the household.* *Rs must give a group as worded above, or below in Q18 (HISP ORIGIN) for an adequate answer. Check both lists before you categorize the answer as adequate or inadequate.*
Skip Instructions	If yes, go to Q18 (HISP_YES). If no, go to Q20 (RACE).

Question 18	HISP_YES
Screen	[Are you / Is NAME] **Mexican, Mexican American, Chicano?** **Puerto Rican?** **Cuban?** **Another Spanish, Hispanic or Latino origin?** **(For example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.)**
Acceptable Responses	Mexican, Mexican American, Chicano Puerto Rican Cuban Another Spanish, Hispanic or Latino origin Argentinean Bolivian Central American Indian Chilean Colombian Costa Rican Dominican Ecuadorian Guatemalan Honduran Nicaraguan Panamanian

	Paraguayan
	Peruvian
	Salvadoran
	South American Indian
	Spaniard
	Spanish American
	Uruguayan
	Venezuelan
Comments	*Rs may give more than one response.*
Skip Instructions	*If R answers "other" go to Q19 (HISP_OTHER)* *Else, go to Q 20 (RACE).*

Question 19	**HISP_OTHER**
	What is the name of the Spanish, Hispanic, or Latino origin? (For example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.)
Acceptable	*Argentinean*
	Bolivian
	Central American Indian
	Chilean
	Colombian
	Costa Rican
	Ecuadorian
	Guatemalan
	Honduran
	Nicaraguan
	Panamanian
	Paraguayan
	Peruvian
	Salvadoran
	South American Indian
	Spaniard
	Spanish American
	Uruguayan
	Venezuelan
Comments	*Rs may give one or two, but no more than two, responses.* *To have an acceptable response, Rs must give exactly one of the responses shown above.*
Skip Instructions	Go to Q 20 (RACE).

Question 20	RACE
Screen	**Using this list, please choose one or more races that [you consider yourself to be / NAME considers himself to be / NAME considers herself to be].**
Acceptable Responses	White Black, African American, or Negro American Indian or Alaska Native [Go to Q21] Asian Indian Chinese Filipino Japanese Korean Vietnamese Other Asian (For example: Hmong, Laotian, Thai, Pakistani, Cambodian) [Go to Q22] Native Hawaiian Guamanian or Chamorro Samoan Other Pacific Islander (For example: Fijian, Tongan) [Go to Q23]
Comments	*Rs may provide more than one answer (example - I am Japanese and Native Hawaiian)* *Rs must give a group as worded above, or as listed in Q21 (AMER IND), or Q22 (ASIAN), or Q23 (PACIFIC) for an adequate answer.* *Check all 4 lists before you categorize the answer as adequate or inadequate*
Skip Instructions	If "American Indian or Alaska Native" go to Q21 (AMER_IND). If "Other Asian," go to Q22 (ASIAN). If "Other Pacific Islander," go to Q23 (PACIFIC). If a combination of these items is given, go to Questions 21-23 in that order. Else, go to Q24 (COVERAGE2).

Question 21	AMER_IND
Screen	**What is the name of [your / NAME=S] enrolled or principal tribe?**
Acceptable Responses	Alaskan Athabascan Apache Blackfeet Cherokee Cheyenne Chickasaw Chippewa Choctaw Comanche Delaware Hopi Pueblo Laguna Pueblo Lumbee Mexican American Indian Mohawk Muscogee Creek Oglala Sioux Navajo Rosebud Sioux Seminole Sioux Turtle Mountain Band of Chippewa Tohono O=Odham Yaqui
Comments	*This question only appears if the R reported "American Indian or Alaska Native" in the Race question (Q20).* *Rs may give one or two, but no more than two, responses.* *Rs must give a group as worded above for an adequate answer. If they say any group not listed above, it is an inadequate answer, even if it is a tribe.*
Skip Instructions	Unless another "other" category (i.e., "other Asian" or "other Pacific Islander") was reported in the Race question (Q20), go to Q24 (COVERAGE2).

Question 22	Asian
Screen	What is the name of [your / NAME=s] Other Asian race?
Acceptable Responses	Bangladeshi Bhutanese Burmese Cambodian Hmong Indo-Chinese Indonesian Iwo Jiman Laotian Malaysian Maldivian Mongolian Nepalese Okinawan Pakistani Singaporean Sri Lankan Taiwanese Thai Yello
Comments	*This question only appears if the R reported "Other Asian" to the Race question (Q20).* *Rs may give one or two, but no more than two, responses.* *Rs must give a group as worded above for an adequate answer. If they say any group not listed above, it is an inadequate answer, even if it is an other Asian group.*
Skip Instructions	Unless another "other Pacific Islander" category was reported in Q20 (RACE), go to Q24 (COVERAGE2).

Question 23	PACIFIC
Screen	What is the name of [your / NAME=s] Other Pacific Islander race?
Acceptable Responses	Chuukese Fijian Kosraean Marshallese Micronesian Palauan Papua New Guinean Pohnpeian Polynesian Saipanese Tahitian Tokelauan Tongan Yapese

Comments	This question only appears if the R reported "Other Pacific Islander" to the Race question (Q20).
	Rs may give one or two, but no more than two, responses.
	Rs must give a group as worded above for an adequate answer. If they say any group not listed above, it is an inadequate answer, even if it is an other Pacific Islander group.
Skip Instructions	Go to Q24 (COVERAGE2).

Question 24	COVERAGE_2
Screen	[Do you / Does NAME / Did NAME] sometimes live or stay somewhere else? **To attend college?** **To stay at a seasonal or second residence?** **To be closer to work?** **For a child custody arrangement?** **For any other reasons?**
Acceptable Responses	To attend college To stay at a seasonal or second residence To be closer to work For a child custody arrangement For any other reasons No
Comments	Interviewer must read all response options. *If "any other reason" is given, write notes verbatim.
Skip Instructions	Go back to Q11 (RELATN) for next person. Else, end.

SURVEY QUESTION GUIDE

Spanish

NRFU 2004

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 may use their discretion when reading text appearing in parentheses. It is not necessary for them to do so, but interviewers may read this text if they feel it is necessary.

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	¿Vivía usted, o algún miembro de su hogar, en la dirección *[ADDRESS]* el 1 de abril del 2004?
Acceptable	Sí No
Comments	*Interviewer should read a street address for [ADDRESS]*
Skip Instructions	If yes, go to Q2 (USUAL RES). If no, end interview.

Question 2	USUAL RES
Screen	¿Es esta [casa / apartamento / casa móvil] La residencia usual de alguien en su hogar, O es una casa vacacional, de temporada o de uso ocasional?
Acceptable	Residencia usual Casa vacacional de temporada o de uso ocasional
Comments	
Skip Instructions	Go to Q3 (POP COUNT).

Question 3	POP COUNT
Screen	¿Cuántas personas vivían o se quedaban en esta [casa / apartamento / casa móvil] el 1 de abril del 2004?
Acceptable Responses	Anote el número de personas (1-49)
Comments	*May read flashcard aloud - this would still be an exact reading, as long as question is worded as written above.*
Skip Instructions	Go to Q4 (NAMES).

Question 4	NAMES
Screen	¿Cuál es el nombre de cada persona que vivía o se quedaba en [esta / esa] residencia el 1 de abril del 2004? Comience con el nombre de una persona que era dueña o que alquilaba [esta / esa] (casa / apartamento / casa móvil) el 1 de abril del 2004.
Acceptable Responses	Nombre, (Inicial), Apellido
Comments	*R should give first and last name, but middle initial and suffix are acceptable too. If R only gives first name, it is an inadequate answer.*

	Code only for first person that is listed.
Skip Instructions	Go to Q5 (ANYONE_ELSE1).

Question 5	**ANYONE_ELSE1**
Screen	**¿Hay alguien más?**
Acceptable Responses	Sí No
Comments	*Code only the first time this is said, should be said after the first person is listed, or when R stops giving names.*
Skip Instructions	After you get a "no" to this question - If number given to Q3 (POP COUNT) does not equal the number of names given in Q5, go to Q6 (RECONCILE POP). Else, skip to Q7 (COVERAGE1)

Question 6	**RECONCILE_POP**
Screen	**Anteriormente anoté que [había / habían] [POP COUNT] [persona / personas] vivían o se quedaban en la dirección *[ADDRESS]*. Pero ahora acabamos de anotar a [X] [persona / personas].** **¿Cuál es el número correcto?**
Acceptable Responses	[POP COUNT] [X]
Comments	*This will only appear if the count given in Q3 (POP COUNT) is inconsistent with the number of names given. It should be skipped in most cases.*
Skip Instructions	Go to Q7 (COVERAGE1).

Question 7	**COVERAGE1**
Screen	**Aparte [del nombre / de los nombres] que me acaba de dar, ¿habían otras personas que a veces vivían o se quedaban en [este/ ese] lugar pero que no residían [aquí / ahí] permanentemente? Por ejemplo, empleados que viven en el hogar o hijos bajo custodia conjunta.**
Acceptable Responses	Sí No
Comments	*Note any comments here about reconsidering the answer to the previous question. Like – "Oh there=s also Nanny. I should have said 5 people, not 4."*
Skip Instructions	If yes, go to Q8 (ADD PERSON). If no, skip to Q10 (OWNERSHIP).

Question 8	ADD_PERSON
Screen	¿Cuál es el nombre de la persona que vivía o se quedaba en [este / ese] lugar pero que todavía no está en la lista de nombres?
Acceptable Responses	Nombre, (Inicial), Apellido
Comments	*This will only be asked if the answer to Q8 (COVERAGE_1) is AYes".* *Code only for first person listed. R should give first and last name, but middle initial and suffix are acceptable too. If R only gives first name, it is an inadequate response.*
Skip Instructions	Go to Q9 (ANYONE ELSE2).

Question 9	ANYONE_ELSE2
Screen	¿Alguien más?
Acceptable Responses	Sí *No*
Comments	*This is only asked if Q8 (ADD_PERSON) is asked.* *Code only the first time this is said, should be said after the first person is listed, or when R stops giving names.*
Skip Instructions	After you get a Ano," go to Q10 (OWNERSHIP).

Question 10	OWNERSHP
Screen	¿Es [Su / Esta] [casa / apartamento / casa móvil] ɘdad suya o de alguien en este hogar con una hipoteca o préstamo? ɘdad suya o de alguien en este hogar libre y sin deuda (sin una hipoteca o préstamo)? Alquilada por pago de alquiler en efectivo? Habitada sin pago de alquiler en efectivo?
Acceptable Responses	Propiedad con una hipoteca o préstamo Propiedad libre y sin deuda Alquilada por pago de alquiler en efectivo Habitada sin pago
Comments	*Interviewer must read all response options.*
Skip Instructions	Go to Q12 (SEX) for first person.

Starting here, all questions are asked for each person.

Question 11	RELATN
Screen	**FOR PERSON 1** This Q is not read. **FOR PERSON 2 +** ¿Cuál de estas categorías mejor describe cómo está [usted / <u>NAME</u> / la Persona X] relacionado(a) con [usted / NAME]?
Acceptable	Esposo o esposa Hijo o hija Hermano o hermana Padre o madre Nieto o nieta Pariente político Otra clase de pariente Inquilino Compañero de casa Pareja (no casada) Hijo o hija foster Otra persona sin parentesco que recibe cuidado formal Otra persona sin parentesco
Comments	*This question will be skipped for the first person in the household.*
Skip Instructions	Go to Q12 (SEX).

Question 12	SEX
Screen	¿Es *[usted / <u>NAME</u> / la Persona X]* de sexo masculino o feminino?
Acceptable	Masculino Feminino
Comments	*This should be explicitly asked for each person.*
Skip Instructions	Go to Q13 (AGE).

Question 13	AGE
Screen	¿Cuál era la edad [suya / de NAME] el 1 de abril del 2004? (Si no sabe la edad exacta, por favor deme un estimado.)
Acceptable Responses	0-150
Comments	
Skip Instructions	Go to Q14 (DOB).

Question 14	DOB
Screen	¿Cuál es [su / la] fecha de nacimiento [de *NAME* / la Persona X]?
Acceptable Responses	MM\ DD\ AÑO
Comments	*Full date of birth is required for an acceptable response. Month and day is inadequate.*
Skip Instructions	If the calculated age is not the same as the reported age, go to Q15 (AGECHECK). Else, skip to Q17 (HISP).

Question 15	AGECHECK
Screen	Para el Censo, necesitamos anotar la edad a partir del 1 de abril del 2004. A base de la fecha que acabo de anotar, [usted / NOMBRE / la Persona X] tenía *[AGEC]*. Anteriormente había anotado que [su edad / la edad de NAME / la edad de la Persona X] era *[AGE]*. ¿Cuál es la edad correcta a partir del 1 de abril del 2004?
Acceptable Responses	[AGE] [AGEC]
Comments	*This will only be asked if the reported AGE (Q13) is inconsistent with the age as calculated from the date of birth. This will be skipped for most Rs.*
Skip Instructions	If If the answer corresponds to the given age in Q13 (AGE), go to Q16 (DOBCHECK). If the answer is the calculated age (AGEC), skip to Q17 (HISP).

Question 16	DOBCHECK
Screen	Como [usted / *NOMBRE* / la Persona X] tenía *[AGE@X]* años a partir del 1 de abril del 2004, ¿podría ayudarme a corregir [su / la] fecha de nacimiento [de *NOMBRE* / de la Persona X]? *Tengo anotado que la fecha de nacimiento es DOB@X, ¿qué debería ser?*
Acceptable Responses	MM\ DD\ AÑO
Comments	*This is only asked when Q15 (AGECHECK) is asked, and the answer is the reported age instead of the age calculated from the date of birth. This will be skipped for most Rs.*
Skip Instructions	Go toQ17 (HISP).

Question 17	HISP
Screen	**FOR PERSON 1** **Voy a hacerle dos preguntas, una sobre origen hispano y otra sobre raza. Por favor conteste las dos preguntas.** **¿Es [usted / NOMBRE / la Persona X] de origen español, hispano o latino?** **FOR PERSON 2 +** **(Voy a hacerle dos preguntas, una sobre origen hispano y otra sobre raza. Por favor conteste las dos preguntas.)** **¿Es [usted / NOMBRE / la Persona X] de origen español, hispano o latino?**
Acceptable Responses	*No* *Sí*
Comments	*Interviewer should read optional intro statement for the first person. This is required for an exact reading for the first person, but is optional for an exact reading for later persons in the household.* *Rs must give a group as worded above, or below in Q18 (HISP YES) for an adequate answer. Check both lists before you categorize the answer as adequate or inadequate.*
Skip Instructions	If "yes," go to Q18 (HISP_YES). If "no," go to Q20 (RACE).

Question 18	HISP_YES
Screen	**¿Es [usted / NOMBRE / la Persona X]** **Mexicano, mexico-american, chicano?** **Puertorriqueño?** **Cubano?** **De algún otro origen español, hispano, o latino? (Por ejemplo, argentino, colombiano, dominicano, nicaragüense, salvadoreño, español, etc.)**
Acceptable	*Mexicano, mexico-americano, chicano* *Puertorriqueño* *Cubano* *De algún otro origen español, hispano, o latino* *Argentino* *Boliviano* *Indio centroamericano* *Chileno* *Colombiano* *Costarricense* *Dominicano* *Ecuatoriano* *Guatemalteco* *Hondureño* *Nicaragüense* *Panameño* *Paraguayo* *Peruano*

	Salvadoreño Indio sudamericano Español Español americano Uruguayo Venezolano
Comments	Rs may give more than one response. To have an acceptable response, Rs must give exactly one of the responses shown above. If respondent gave an answer that is an acceptable to this question for the previous question Q17 (HISP), this question can be skipped and coded as a correct skip.
Skip Instructions	If R answers "other" go to Q19 (HISP_OTHER) Else, go to Q 20 (RACE).

Question 19	HISP_OTHER
Screen	**¿Cual es el nombre del grupo de origen español, hispano o latino? (Por ejemplo, argentino, colombiano, dominicano, nicaragüense, salvadoreño, español, etc.)**
Acceptable	Argentino Boliviano Indio centroamericano Chileno Colombiano Costarricense Dominicano Ecuatoriano Guatemalteco Hondureño Nicaragüense Panameño Paraguayo Peruano Salvadoreño Indio sudamericano Español Español americano Uruguayo Venezolano
Comments	Rs may give one or two, but no more than two, responses. To have an acceptable response, Rs must give exactly one of the responses shown above.
Skip Instructions	Go to Q 20 (RACE).

Question 20	RACE
Screen	*Usando esta lista, por favor, escoja una o más razas para indicar de qué raza se considera [usted / <u>NOMBRE</u>].*
Acceptable	*Blanca* *Negra o Africana-Americana* *India Americana o Nativa de Alaska* [Go to Q21] *India Asiática* *China* *Filipina* *Janponesa* *Coreana* *Vietnamita* *Otra raza asiática (Por ejemplo: hmong, laosiana,* * tailandesa, pakistana, camboyana)* [Go to Q22] *Nativa de Hawaii* *Guameña o Chamorro* *Samoana* *De otra isla del Pacífico (Por ejemplo: Fiyiana, Tongana)* [Go to Q23]
Comments	*Rs may provide more than one answer (example - I am Japanese and Native Hawaiian)* *Rs must give a group as worded above, or as listed in Q21 (AMER IND), or Q22 (ASIAN), or Q23 (PACIFIC) for an adequate answer.* *Check all 4 lists before you categorize the answer as adequate or inadequate*
Skip Instructions	If "American Indian or Alaska Native," go to Q21 (AMER_IND). If "Other Asian," go to Q22 (ASIAN). If "Other Pacific Islander," go to Q23 (PACIFIC). If a combination of these items is given, go to Questions 21-23 in that order. Else, go to Q24 (COVERAGE2).

Question 21	AMER_IND
Screen	*¿Cuál es el nombre de la tribu en la cual está [usted / <u>NOMBRE</u> / la Persona X] [inscrito / inscrita] o la tribu principal?*
Acceptable Responses	*Alaskan Athabascan* *Apache* *Blackfeet* *Cherokee* *Cheyenne* *Chickasaw* *Chippewa* *Choctaw* *Comanche* *Delaware* *Hopi Pueblo* *Laguna Pueblo* *Lumbee* *Indio mexicano-americano* *Mohawk*

	Muscogee Creek Oglala Sioux Navajo Rosebud Sioux Seminole Sioux Turtle Mountain Band of Chippewa Tohono O=Odham Yaqui
Comments	*This question only appears if the R reported "American Indian or Alaska Native" in the Race question (Q20).* *Rs may give one or two, but no more than two, responses.* *Rs must give a group as worded above for an adequate answer. If they say any group not listed above, it is an inadequate answer, even if it is a tribe.*
Skip Instructions	Unless another "other" category (i.e., "other Asian" or "other Pacific Islander") was reported in the Race question (Q20), go to Q24 (COVERAGE2).

Question 22	**Asian**
Screen	*¿Cuál es el nombre de [su / la] otra raza asiática [de <u>NOMBRE</u> / de la Persona X]?*
Acceptable	*Bengalí* *Butanesa* *Burmesa* *Camboyana* *Hmong* *Indochina* *Indonesa* *De Iwo-Jima* *Laosiana* *Malaya* *Maldiva* *Mongola* *Nepalí* *Okinawense* *Paquistana* *Singapurense* *Ceilandesa (de Sri Lanka)* *Taiwanesa* *Tailandesa* *Yello*
Comments	*This question only appears if the R reported "Other Asian" to the Race question (Q20).* *Rs may give one or two, but no more than two, responses.* *Rs must give a group as worded above for an adequate answer. If they say any group not listed above, it is an inadequate answer, even if it is an other Asian group.*
Skip Instructions	Unless another "other Pacific Islander"category was reported in Q20 (RACE), go to Q24 (COVERAGE2).

Question 23	PACIFIC
Screen	¿Cuál es el nombre de [su / la] otra raza de las islas del Pacífico [de <u>NOMBRE</u> / de la Persona X]?
Acceptable	Chuukesa Fiyiana Kosraena Marshalesa Micronesa Paluana Papua Nueva Guinea Ponpeya Polinesia Saipanesa Tahitiana Tokelauana Tongana Yapesa
Comments	This question only appears if the R reported "Other Pacific Islander" to the Race question (Q20). Rs may give one or two, but no more than two, responses. Rs must give a group as worded above for an adequate answer. If they say any group not listed above, it is an inadequate answer, even if it is an other Pacific Islander group.
Skip Instructions	Go to Q24 (COVERAGE2).

Question 24	COVERAGE_2
Screen	¿ [Vive / vivía o se queda / quedaba] [usted / <u>NOMBRE</u> / la Persona X] a veces en algún otro lugar? Para ir a la Universidad? Para estar en una residencia vacaional o en una segunda residencia? Para estar más cerca del trabajo? Por un arreglo de custodia de niños? Por alguna otra razóna?
Acceptable	Para ir a la Universidad a quedarse en una casa de temporado o o en una segunda residencia? a estar más cerca del trabajo Por un arreglo de custodia Por alguna razónno esta en la lista No
Comments	Interviewer must read all response options. *If "any other reason" is given, write notes verbatim.
Skip	Go back to Q11 (RELATN) for next person.

Instructions	Else, end.

Framework of Behavioral Codes and an Explanation of their Analytical Function

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 UA: Uncertain Answer, respondent expresses uncertainty about the response provided and may be unsure about the accuracy of the information

Code QA: Qualified Answer, 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, respondent requests that a concept or entire question be stated more clearly (expressing uncertainty about meaning)

Code RR: Question Re-Read, respondent asks interviewer to reread the question

Code DK:	Don't Know, respondent states they do 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. In addition, we attempted to document cases where respondents seemed to be overtly annoyed, angered, or frustrated by any questions through the NR (Negative Reaction) code:

| 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). |
| Code NR: | Negative Reaction, respondent expresses frustration, irritation, or anger in response to a question. Coders were given instructed to use this code for any negative reactions evident during the question administration, regardless of the level of interaction at which it occurred. |

Final Response Outcome Codes (ultimate answer)

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 following codes were omitted: question reread (Code RR) and clarification (CL). These codes were 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), UA (uncertain answer), QA (qualified answer), DK (don't know), R (refusal), and I/O (inaudible/other).