

An Overview of the 2005 National Census Test¹

Jennifer Guarino Tancreto
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract

In preparation for the 2010 Census, the U.S. Census Bureau conducted the 2005 National Census Test (NCT) in the fall of 2005. The 2005 NCT was a large, multi-faceted mailout test that was designed to study questionnaire content and design as well as methods for improving response and the efficiency of data collection procedures. In this paper, I outline the five overarching objectives of the 2005 NCT, and then describe the treatments that were used to test these objectives. I explain how the treatments were intermixed into one panel design for cost saving purposes, and provide the justification behind the combinations of treatments into the experimental panels. I then describe the sample design and sample selection procedures. Finally, I report issues that occurred during the data collection period, and provide some global metrics for the test.

Keywords: Census, experiment

1. Background

In preparation for the 2010 Census, the U.S. Census Bureau conducted the 2005 National Census Test (NCT) in the fall of 2005. The 2005 NCT was a multi-faceted test that was designed to study questionnaire content and design as well as methodology to improve the efficiency of data collection procedures and response to the census.

The Census Bureau began planning the 2005 NCT in early 2004 with a call for research objectives. Subject matter and methodology experts provided their ideas for test material coupled with justifications explaining the need for the research. Test designers then developed experimental designs to encompass these test objectives in a cost-efficient manner. Cost restraints forced several iterations of the design, and many objectives were eventually eliminated from the test in an attempt to narrow the scope.

After a few months of deliberations, the Census Bureau cited five overarching test objectives.

The first objective covered data quality aspects of items that are present on the census questionnaire. Specifically, this objective aimed to improve completeness and reporting for tenure, relationship, age, date of birth, race, and Hispanic origin. Subject matter experts proposed certain variations in question wording and response categories to meet this objective.

The second objective of the test was to reduce respondent and data capture errors by improving the respondent-friendliness of both the mail and Internet data collection instruments.

The third objective was to improve population coverage, by reducing omissions and erroneous enumerations. This objective proposed alternative household roster approaches and the addition of questions to try to identify persons and households where coverage errors may be prevalent.

The fourth objective was to improve the operational feasibility of sending a follow-up mailing with a replacement questionnaire to households that had not returned a census form by a certain date.

Finally, the fifth objective was aimed at improving unit and item response by mailing a bilingual questionnaire that provided the opportunity to respond in English and Spanish.

In this paper, I will discuss the treatments that were tested to meet the five objectives of the NCT. I will explain how the treatments were intermixed into one panel design, and provide the justification behind the combinations of treatments in experimental panels. I will then describe the sample design and sample selection procedures. Next, I will mention some issues that occurred during data collection, and finally, end with a few global test metrics. Note that I will not present the experimental results of the test. For test results, please refer to Bentley *et al.* (in press), Hill *et al.* (in press), Allmang *et al.* (in press), and

¹This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical, methodological, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

Bouffard *et al.* (in press). More test results will be available in early 2007.

2. Methodology

2.1 Scope

The 2005 NCT was a mailout test only. That is, we sent questionnaires to housing units located in areas where good street addresses were available. There were no field operations in this test, such as hand delivering forms to housing units in rural areas or personal visit follow-up with nonresponding housing units. Because the 2005 NCT was a national test, it would have been too costly to hire field staff in all areas of the country to perform field operations.

The 2005 NCT went into the field in August 2005, and data capture continued through November 15, 2005. September 15th was “census day” or the reference date that households used to determine their household composition and age of their household members.

2.2 Mailings

The 2005 NCT used multiple mailings to contact sampled housing units. Every housing unit was sent an advance letter as a first contact on August 22, 2005. The advance letter informed households that they would soon receive a request to complete a questionnaire for the 2005 NCT.

The second mailing was the initial questionnaire package. Housing units received a paper questionnaire and a postage-paid return envelope. Also included in the mailing package was a letter from the Census Bureau’s Director that encouraged households to participate and provided the option of responding by Internet. The paper questionnaire also reminded respondents of the option to respond by the Internet before the first question. The initial questionnaire package was mailed one week after the advance letter, on August 29, 2005.

The third mailing was a reminder postcard. The reminder postcard reminded households to respond to the NCT if they had not already done so. It also provided instructions so that households could use the Internet to respond. The reminder postcard was sent on September 6, 2005.

The fourth and final mailing was a targeted replacement questionnaire. A replacement questionnaire that looked identical to the initial questionnaire (i.e., contained the same experimental treatments) was sent to all housing units that had not responded by September 13, 2005. Accompanying the questionnaire was a letter from the Director urging response and providing instructions for using the Internet. The replacement questionnaire package

was delivered to the post office for mailing on September 23, 2005.

2.3 Response Modes

Every housing unit in the 2005 NCT was offered the opportunity to respond by returning a paper form by mail or using the Internet. The 2005 NCT tested two distinct Internet design paths; however, the Internet did not contain all of the experimental treatments that were present on the paper questionnaires for each panel. In this sense, households that responded by the Internet were no longer considered part of their original experimental panel since they were not exposed to the experimental treatments in the paper questionnaire.

3. Experimental Treatments

This section describes the experimental treatments used to test each 2005 NCT objective.

3.1 Objective 1 – Data Quality Aspects of Items

The first global objective involved testing variations in question wording and response categories to enhance the understandability of the tenure, relationship, age, race, and Hispanic Origin questions.

3.1.1 Tenure Question

Subject matter experts proposed three alternative wording options for the response categories in the tenure question. They proposed these changes in hopes of improving respondent comprehension, and thus, accuracy of reporting. See Hill *et al.* (in press) for a detailed description of the three experimental questions for tenure.

3.1.2 Relationship Question

Similar to the tenure item, the experimental treatment for the relationship item involved changes to the response categories. There were six proposed changes to the relationship response options. Some of the changes were suggested based on unfavorable past reactions, while others were more cosmetic in nature. See Hill *et al.* (in press) for a list of these changes. Since there were six proposed changes to the response options, none of which were expected to substantially effect the response distribution, the changes were tested together in one experimental relationship question.

3.1.3 Age and Date of Birth Questions

Subject matter experts sought two age treatments for the 2005 NCT. The first change was the addition of an instruction to provide guidance on reporting age for babies

less than a year old. This treatment was intended to curb the tendency to report age in months rather than years for this age group, as has been previously observed (Spencer *et al.*, 1998).

Secondly, they proposed reversing the order of age and date of birth to be consistent with the design of automated instruments, such as the Internet, which use this order to provide real-time age edits. Each age treatment was tested separately in the 2005 NCT.

3.1.4 Race, Hispanic Origin and Ancestry

Subject matter experts proposed testing six features of the race and Hispanic origin questions:

1. Question design: The 2005 NCT tested shortened Hispanic origin and race questions with an ancestry follow-up question. The three-question design reduced the categories to the minimum required by the 1997 Office of Management and Budget standards: “Yes” and “No” for Hispanic origin, and White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and some other race for race.
2. Examples: The 2005 NCT tested the use of race and Hispanic group examples in the shortened questions to determine if there was an impact on reporting.
3. Order: We also tested for differences in Hispanic origin identification based on changing the order of the “Spanish, Hispanic, or Latino” identifiers in the Hispanic origin question.
4. Revised race instructions: Three versions were tested. Two versions informed respondents that they could choose more than one race, while the other one highlighted the difference between Hispanic origin and race.
5. Respondents’ notes: These notes were intended to communicate the distinction between race and Hispanic origin as well as the need to answer both questions.
6. Tribal enrollment question: This question was used to determine whether there is higher reporting of specific American Indian or Alaska Native tribes using a separate question versus a collective ancestry question.

A paper containing information about the race and Hispanic origin treatments and results will be available in early 2007.

3.2 Objective 2 – Respondent-friendly Forms Design

This objective included changes to the paper and Internet forms to enhance response and data capture.

3.2.1 Respondent-friendly Paper Questionnaire Design

The 2005 NCT tested two paper questionnaire formats in an effort to improve the respondent-friendliness of the form. The first treatment tested a format that was hypothesized to be more respondent-friendly and easier for data capture systems to process.

The second treatment, dubbed the space saving design, attempted to address the growing concern about real estate on the form. This treatment included several changes intended to simplify and shorten the questions on name, sex, date of birth, age, race and ethnicity.

3.2.2 Internet Design

As part of the second global objective, the 2005 NCT included two distinct Internet design paths. The purpose of testing two paths was to determine which format was less time consuming, easier to use, and less prone to item nonresponse. For more information about these formats, please see Allmang *et al.* (in press).

3.3 Objective 3 – Population Coverage

A research and development group at the Census Bureau recommended testing two coverage factors in the 2005 NCT. The first factor, household roster, tested six approaches to roster instructions that were intended to improve coverage and the accuracy of where people are counted.

The second factor was a set of coverage probes. One coverage probe was intended to gauge the potential for undercoverage within a household. The second probe was included to help detect potential overcoverage for each person in the household. Two sets of coverage questions (each containing one undercount question for the household and one overcount question for each person within the household) were used in the 2005 NCT.

Additionally, we conducted a coverage follow-up interview by telephone with a sample of respondents to assess the accuracy with which respondents report household coverage using the various rosters and coverage questions. A paper containing detailed information about the coverage treatments and results of the experiment will be available in early 2007.

3.4 Objective 4 – Replacement Questionnaire Strategy

We tested five strategies for implementing a targeted replacement questionnaire mailing in the 2005 NCT. These strategies ranged from the printing and labeling of the questionnaire packages to offering the Internet option without a paper questionnaire. The goal of this testing was to improve the operational feasibility of sending a follow-up mailing to housing units that had not responded by a certain date. For more information on these treatments, see Bentley *et al.* (in press).

3.5 Objective 5 – Bilingual questionnaire

The final global objective of the 2005 NCT called for the use of a bilingual questionnaire. A bilingual questionnaire, that offered the option of responding in English or Spanish, was created for this test objective. For more information about this form, please see Bouffard *et al.* (in press).

4. Creation of Experimental Panels

There were many diverse treatments that were included in the 2005 NCT. Table 1 shows the number of experimental treatments proposed for each test objective. Note that some treatments involved multiple changes, but the changes were tested simultaneously in one treatment.

Table 1. Number of Experimental Treatments by Objective

| Objective | Number of Treatments |
|--|----------------------|
| 1 – Data quality for short form items | |
| Tenure | 3 |
| Relationship | 1 |
| Age/date of birth | 2 |
| Race/Hispanic origin | 5 |
| 2 – Respondent-friendly form design | 2 |
| 3 – Population Coverage | 12 |
| 4 - Replacement Questionnaire Strategy | 5 |
| 5 – Bilingual Questionnaire | 1 |

Our goal was to develop an experimental design in a manner that would be as cost-efficient as possible. Because of limited funding, we did not have the option of creating one experimental panel for each individual treatment, nor could we propose looking at the interactions

of all treatments in a fully-factorial design. Instead, due to the large number of treatments, we were forced to combine treatments in experimental panels.

4.1 Criteria

Given the need to combine treatments, we set forth our requirements for this process. First, we wanted to avoid combining experimental treatments that shared a common objective. For instance, we did not want to combine treatments such as the respondent-friendly questionnaire design and the bilingual form since both treatments were intended to, at least in part, increase response. Instead, we combined treatments where the objectives and main outcome measures did not overlap.

Moreover, we wanted to join treatments where there was no hypothesized interaction between the treatments. For instance, one treatment encouraged Internet response at the replacement questionnaire mailing. We felt that this treatment had the potential to impact respondent demographics since Internet access is not universal. Therefore, we did not combine that treatment with any treatment that studied response distributions of person or household level items.

Finally, we felt that the roster designs and coverage questions could lead to the inclusion or exclusion of particular household members, and thus, could affect distributions of person items. Therefore, we did not combine any coverage treatments with the relationship, age/date of birth, or race and Hispanic origin objectives.

4.2 Description of the Panels

Officially, we ended up with 20 experimental panels; however, one panel, our control panel, had 5 different components.

The control panel contained mostly content that had been previously used or tested. Moreover, the control panel used the standard mailing strategy for both the initial and replacement questionnaires. For many items, such as tenure, relationship, age and date of birth, the control panel version was taken from Census 2000². For other items, we employed versions of the questions that were shown to be superior to that used in Census 2000 through experimental testing. For instance, the race and Hispanic origin items were adapted from the 2003 National Census Test questions because these questions had some advantages over the Census 2000 version (Sheppard *et al.*, 2004). Moreover, the roster design chosen as the control version

2 In some cases, we made minor variations, such as line spacing, to these items due to space constraints on the forms.

surpassed the Census 2000 roster in the Alternate Questionnaire Experiment in Census 2000 (Gerber *et al.*, 2003). A few of the features of the 2005 control form were untested, including the invitation to respond by Internet at the top of the questionnaire, the instruction for who to list as Person 1, and the extended roster questions for persons 7 through 12.

As previously mentioned, the control panel had five components. The first component was the true control panel described above. This component served as a baseline to which most of the experimental treatments were compared. The other four components, which consisted of four of the five replacement questionnaire strategies, shared the same content and mailing strategy as the “control” for the initial questionnaire mailing; however, each contained its own variation of the traditional replacement questionnaire mailing strategy.

The reasoning behind the placement of these five components into one panel relates to sample size and statistical power. Since all components shared the same design for the initial questionnaire, analytical comparisons involving the control panel could use the returns across all five components (with a weighting adjustment to exclude cases exposed to a replacement questionnaire experimental treatment). Our research indicated that the increase in sample size resulting from the use of all components would cut down the variance of an estimate by about one-half (Bouffard, 2005). Unfortunately, our basic assumption of similar response rates for the initial questionnaire returns among the five components was violated (due to a treatment effect in one of the components), so we could not combine sample across the components as planned. Hence, panel 1 contained five components that were analyzed separately in the test.

The methodology used to form the coverage and roster panels was pretty basic. We crossed the roster designs and coverage probes in a factorial design to enable an evaluation of their main effects and anticipated interactions. This factorial design resulted in 12 experimental treatments (6 rosters times 2 sets of coverage questions). As discussed in section 4.1, we felt it was best to isolate these 12 panels from the other treatments since the coverage treatments could impact the population that is reported on the form, and thus, could differentially affect estimates for person items.

The next panel contained the bilingual census form. Because this form had a radically different format than the English only form, in addition to the potential to alter the population responding, we did not combine the bilingual form with any other treatments.

We then created a panel to test the respondent-friendly

questionnaire design. We combined this treatment with one replacement questionnaire strategy, namely a treatment that sought to distinguish the replacement questionnaire from the initial questionnaire using a message on the replacement questionnaire. Since this messaging feature was also used in one component of the control panel, combining these treatments allowed analysts to study the effect of the respondent-friendly design by itself, and in the presence of this feature. Finally, we added one other replacement questionnaire strategy to this panel. This final strategy involved postal tracking of forms that were in the mail stream to be returned to the Census Bureau. This tracking was transparent to respondents, and thus, we expected no confounding effects with the other treatments in the panel.

Finally, we considered the question-specific treatments involving race, Hispanic origin, age, relationship, and tenure. We attempted to consolidate these treatments into five panels in a way that would minimize any bias or interaction. We began with race and Hispanic origin treatments, since these items had substantive changes (i.e., moving from two long questions to three shorter questions). We felt confident that we could combine the tenure, age, and relationship treatments with race and Hispanic origin treatments since the extensive changes in race and Hispanic origin came after the tenure, age, and relationship items. Moreover, we felt that the minor changes in the tenure, relationship, and age items were not likely to impact the race and Hispanic origin treatments since these changes were subtle.

Also included in one of these five panels was the remaining paper form design treatment, the space saving design. Since this design treatment was intended to save space on the form, we combined this design with the longest race and Hispanic origin questions to see if the design would accommodate these questions.

5. Sample

5.1 Size

Our first step in designing the sample was to determine a sample size for the panels that would facilitate the analysis of each objective in a cost-efficient manner. We used data from the 2003 National Census Test in addition to Census 2000 data to help achieve this task. We started out with estimates of unit response rates and population characteristics that were directly related to the objectives of the test. We then developed minimum difference requirements for characteristics we planned to measure between experimental panels.

For unit response rate objectives, we aimed to reach statistical significance when a two-percentage point or

higher difference existed between panels. Given the size of some of the small racial groups, we hoped to measure differences on the order of .5 percentage points for the multiple races, Asian, and American Indian/Alaska native categories, and about one percentage point for the Black and White racial groups.

We recommended using 10,000 housing unit samples in each panel where unit response rate was the main analytical measure. Because shifts in the smaller race categories were harder to detect, panels involved in the race and ethnicity testing contained three times as many housing units (30,000). Finally, we used 15,000 housing units for each population coverage panel. For the 2005 NCT, we selected a total of 420,000 housing units across 20 panels (including the control panel components).

The sample size calculations assumed 10 percent of the respondents in each panel would use the Internet to respond. Analysts planned to remove these cases from the analysis of the paper experimental treatments since the Internet did not offer the same experimental treatments.

5.2 Frame

We used the Hundred-Percent Census Edited File to select the sample. This file contained edited data for housing units in Census 2000. The universe for this sample was limited to housing units from mailout areas in Census 2000; that is, areas containing housing units with city-style addresses. About 80 percent of all housing units in the United States had city-style addresses as of Census 2000. Puerto Rico was not part of the test.

5.3 Stratification

Since the objectives of this test were wide-ranging, we considered using different stratification schemes for the various sets of objectives to maximize efficiency. For instance, we considered using race and ethnicity indicators to create strata for the Hispanic origin and race treatments, and response rate related indicators for the other treatments. However, research revealed that we could use one universal stratification approach across all panels since there was a strong correlation between race and response to the census. Using one stratification approach also facilitated the combination of experimental treatments within the panels.

We created two strata based on the racial and ethnic composition of blocks, using data from Census 2000. To accomplish this, we first classified all census blocks into one of twelve demographic/tenure groups (6 race groups by owner/renter status) used for the Accuracy and Coverage Evaluation in Census 2000. Next, we analyzed mail return rates, sample sizes, and race and Hispanic

origin distributions for each group. We attempted to define the strata in a way that would minimize the anticipated variances. We decided to include all owners and renters in the non-White race groups as the “High non-White or Hispanic Concentration” and the others as the “Low non-White or Hispanic Concentration” since the tenure did not seem useful in stratifying the universe. Approximately 32 percent of the housing units are in the “high” stratum and 68 percent in the “low” stratum (Bentley, 2005).

5.4 Allocation

Next, we set out to determine how best to allocate the sample to the strata. Our research showed that proportional allocation would maximize variance reduction; however, since the test proposed comparing both national and stratum-level rates, we recommended equal allocation to allow adequate sample in the smaller stratum. Our research showed that the standard error on our response rate estimate was only slightly higher for equal allocation than for proportional allocation.

5.5 Sample Selection Method

We used a systematic sampling approach to select the sample from each stratum separately, taking a string of housing units at each hit. Selecting hits at each take every helped ensure that the samples in each panel were homogenous. Only every third housing unit in the string was assigned to the panels to prevent any contamination from sampling next-door neighbors.

Since we selected panels of varying sizes, not all panels were assigned at each hit. The overall sampling interval for each stratum was based on the largest panels (30,000 total housing units). The smallest panels of size 10,000 housing units were assigned at every third hit, and the panels of size 15,000 housing units were assigned at, on average, every other hit. Once all the selected housing units were flagged as sample cases, each of the selected housing units within each hit was assigned to a single panel (Bentley, 2005).

6. Test Issues

As with any large test, the 2005 NCT encountered some issues during the data collection period. Most notably, the test coincided with two catastrophic Hurricanes during the questionnaire mailings. Mail service to areas in Texas, Louisiana, Mississippi, and Alabama was disrupted and, in some cases, suspended during most of the data collection period.

6.1 Hurricane Katrina

Hurricane Katrina occurred on August 29, 2005, the Monday following the Friday evening post office delivery of the initial questionnaire mail packages. We learned that mail had been completely suspended to areas that encompassed 3200 cases in our sample. The Census Bureau decided to remove these 3200 cases from the replacement questionnaire mailing, so as not to overload the mail facilities in these areas. As it turned out, mail delivery was still suspended in most of these areas at the time of the replacement questionnaire mailing.

Analysts decided to treat all 3,200 cases in the areas where we suppressed the replacement questionnaire mailing as “undeliverable as addressed” (UAA), as they were temporarily undeliverable during data collection. By making these cases UAA, they are excluded from the analyses. Sample loss due to the hurricane was proportional across panels.

6.2 Hurricane Rita

On September 23, 2005, Hurricane Rita hit areas around the Louisiana and Texas borders. September 23 was the same day that the replacement questionnaire mailing was delivered to the postal service for mailing. Prior to the hurricane, the USPS suspended mail in areas where the hurricane was projected to make landfall. After the Hurricane, mail was suspended to some additional areas for about one day, but was quickly restored to the majority of the areas. The analysts found that response and UAA rates in the areas hit by the hurricane were within range. Therefore, the data in these areas were processed as normal in the test.

7. Global Test Measures

In this section, I report some global measures from the test. Specifically, I present information about the delivery of the questionnaires in addition to the overall unit response rate. All figures presented in this section have been weighted to account for the sample design. Standard errors for all estimates were in the range of .1 to .2 percent. Thus, all implied comparisons are statistically significant.

In this test, we call our unit response rate a “self-response rate” since it includes responses received by Internet and mail. The self-response rate is then an overall measure of respondent behavior with regard to responding to the census test.

The denominator of the self-response rate is the number of sample housing units minus those cases identified by the United States Postal Service as “undeliverable as

addressed” (UAA). Housing units were considered UAA if there was no response (paper or Internet), and both the initial questionnaire and replacement questionnaire mailings were returned as UAA. Any housing units determined to be UAA were considered ineligible units. We observed a 7.8 percent UAA rate for test overall. Please note that this figure include the 3,200 cases where mail was suspended due to Hurricane Katrina.

The numerator of the self-response rate is the number of sample households for which we received a non-blank return. A census return was “non-blank” if at least two items were completed on the form.

Also, we limited the numerator to primary returns. We selected a primary return when we received multiple responses for one household, using the following rules:

- When we received more than one paper return from a single household, we selected the first non-blank form received. In the rare case that two non-blank paper forms were received on the same date, we selected the initial questionnaire return.
- When we received multiple Internet returns from a household, we selected the first nonblank return.
- When we received paper and Internet returns for a household, we selected the first non-blank return received. If a paper and Internet return were received on the same day, we selected the paper return as it was likely mailed before the Internet was submitted.

The formula for the self-response rate is:

$$\frac{\text{Number of non-blank, primary returns}}{\text{Sample size - UAAs}}$$

Please note that this self-response rate corresponds to the rates used in previous census tests, including the 2003 National Census Test, the Census 2000 experiments, and the 1992 and 1993 Census Tests. We used the self-response rate because it is not subject to variation in the UAA rates. Specifically, the denominator excludes cases for which eligibility cannot be determined, such as units that are UAA. Therefore, any variation in the UAA rates across panels does not contribute to differences in the self-response rates.

The self-response rate across all experimental panels for the test was 60.7 percent. It was 44.4 percent in the High Non-White or Hispanic concentration stratum, and 68.1 percent in the Low Non-White or Hispanic concentration stratum. The large difference between the two strata suggests that our stratification efforts were successful.

Overall, 7.3 percent of the overall response rate came from Internet responses, while 53.3 percent came from paper responses. More specifically, Internet response accounted for 12.1 percent of all responses.

Please see Allmang *et al.* (in press) for results related to the use of the Internet as a response mode. The paper by Bouffard *et al.* (in press) contains the results of the bilingual form experiment. Please see Hill *et al.* (in press) for a discussion of the results related to the age, relationship and tenure questions. Finally, the paper by Bentley *et al.* (in press) describes the results of the self-response options.

Results related to the race and Hispanic origin treatments, the space saving questionnaire design, and the coverage-related treatments will be available in early 2007.

Relationship, and Tenure Items from the 2005 National Census Test”, 2006 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.

Sheppard, D., Bentley, M., and Woltman, H. (2004) “2003 National Census Test: Analysis of Alternative Race and Ethnicity Questions Based on Panel Comparisons,” Decennial Statistical Studies Division 2003 Memorandum Series #B-3 and Planning, Research, and Evaluation Division TXE-2010 Memorandum Series: DCO-NCT-F-05, March 3.

Spencer, G. and Perkins Jr., R. (1998) “The Effect of Different Question Formats on Age and Birthdate Statistics From the 1996 National Content Survey,” Internal Census Bureau document, January 23.

8. References

Allmang, K., and Zajac, K. (in press) “Effect of Internet Response Mode Designs on Data Quality and Ease of Use,” 2006 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.

Bentley, M. (2005) Internal Census Bureau memorandum from Killion to Longini entitled “Sample Specifications for the 2005 National Census Test,” version 5.0, March 9.

Bentley, M., and Tancreto, J. (in press) “Analysis of Self-Response Options and Respondent-Friendly Design from the 2005 National Census Test,” 2006 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.

Bouffard, J. and Tancreto, J. (in press) “Experimental Treatment Results of the Bilingual Census Form from the 2005 National Census Test,” 2006 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.

Bouffard, J. (2005) Internal memorandum from 2005 National Census Test Analysis Team to Imel entitled “Sampling Weights for the 2005 National Census Test – Revised,” July 26.

Gerber, E., Dajani, A., and Scaggs, M.A. (2003) “An Experiment to Improve Coverage Through Revised Roster Instructions,” Report in the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, July 15.

Hill, J., Tancreto, J., and Rothhaas, C. (in press) “Experimental Treatment Results for the Age,