2005 National Census Test:
Tenure, Relationship, and Age Report

Cynthia A. Rothhaas
Jason D. Machowski
Jennifer Guarino Tancreto
Joan M. Hill

Distributed for the November 30, 2006
Special Joint Meeting of the U.S. Census Bureau Advisory Committees

U S C E N S U S B U R E A U

2005 National Census Test: Tenure, Relationship, and Age Report

Cynthia A. Rothhaas
Jason D. Machowski
Jennifer Guarino Tancreto
Joan M. Hill
Decennial Statistical Studies Division

USCENSUSBUREAU
Helping You Make Informed Decisions

CONTENTS

EXECUTIVE SUMMARY iii
1. BACKGROUND 1
1.1 Tenure 1
1.1.1 What is the impact on response behavior of dropping the term "cash" from the two renter response options? 2
1.1.2 What is the impact on response behavior of adding the instruction "Include Home Equity loans" following the first owner option –owned with a mortgage or loan? 2
1.1.3 What is the impact on response behavior of dropping the term "cash" from the two renter response options AND adding the instruction "Include home equity loans" following the first owner option – owned with a mortgage or loan? 3
1.2 Relationship 3
1.2.1 Changing the wording "Natural-born son/daughter" to "Biological son or daughter" 3
1.2.2 Changing the wording of "Foster child" to "Foster child or foster adult" 4
1.2.3 Using the conjunction "or" instead of a slash "/" or a comma "," where appropriate for the relationship responses to help clarify the question 4
1.2.4 Excluding the write-in for "Other relative" 4
1.2.5 Deleting the spanner "If NOT RELATED to Person 1:" above the "nonrelative" categories 4
1.3 Age 5

2. METHODOLOGY 5
2.1 Panel Design 6
2.2 Response Mode 6
2.3 Mailing Strategy 6
2.4 Sample Design and Standard Errors 7
2.5 Calculation of Self-Response Rates 7
2.6 Calculation of Item Nonresponse Rates 9
2.7 Analytical Methods 9
2.8 Quality Assurance Procedures 10

3. LIMITATIONS 10
3.1 Relationship Panel 10
3.2 Confounding Factors 10
3.3 Test versus Census 10
3.4 Limitations with Evaluation Measures 10

4. RESULTS 11
4.1 Self-Response Rates 11
4.2 Tenure Analysis 11
4.2.1 Item Nonresponse Rates 12

4.2.2 Response Distributions 13
4.3 Relationship Analysis 14
4.3.1 Item Nonresponse Rates 15
4.3.2 Response Distributions 15
4.3.2.1 Natural-born Son/Daughter Category by Hispanic Origin 18
4.4 Age Analysis 19
4.4.1 Item Nonresponse Rates 19
4.4.2 Response Distributions 19

5. SUMMARY AND RECOMMENDATIONS 21
5.1 Tenure Conclusions 21
5.2 Relationship Conclusions 21
5.3 Age Conclusions 22

6. REFERENCES 22

List of Tables

Table 1 – Self-Response Rates by Mode 11
Table 2 – Tenure Item Nonresponse Rates for Mail Returns at the National and Stratum Level 12
Table 3 - Tenure Item Nonresponse Rate Differences and Standard Errors for Mail Returns at the National and Stratum Level 12
Table 4 - Tenure Item Nonresponse Rate Differences and Standard Errors for Mail Returns at the National and Stratum Level 13
Table 5 - Tenure Response Distributions at the National Level for Mail Returns 13
Table 6 - Tenure Response Differences and Standard Errors at the National Level for Mail Returns 14
Table 7 - Relationship Item Nonresponse Rates, differences, and Standard Errors for Mail Returns Only by Stratum 15
Table 8 – Response Distributions, Differences, and Standard Errors for Mail Returns Only 16
Table 9 – Response Distribution to Son or Daughter Category by Hispanic Origin OF PERSON 1 for Mail Returns Only 18
Table 10 – Item Nonresponse Rates for Reported Age for Mail Returns Only 19
Table 11 - Distributions for Reported Age for Mail Returns Only 20
Table 12 - Distributions of Reported Ages for those with Computed Age of Zero for Mail Returns Only 21

List of Appendices

Appendix A – Self-Response Rates by Stratum and Mode 24
Appendix B – Tenure distributions, differences, and standard errors by stratum for mail returns 25
Appendix C – Relationship Distributions, Differences, and Standard Errors for Mail Returns by Stratum 26

EXECUTIVE SUMMARY

The 2005 National Census Test was a mailout/mailback national test conducted to compare variations of questionnaire content, and various methods to increase response to the Census. This analysis focuses on the experimental panels in the 2005 National Census Test that were designed to test changes to the tenure, relationship, and age questions. The sections below provide information about the experimental treatments that were tested, as well as the results of the testing.

The housing units selected for the tenure, relationship, and age panels (30,000 housing units per treatment panel) and its corresponding control panel (30,000 housing units) were equally allocated to two strata that reflect differences in the racial and ethnic composition, and, hence, response propensity of the mailout/mailback universe. The high non-White or Hispanic concentration stratum, which encompassed roughly 32 percent of housing units in the universe, contained a high proportion of the non-White or Hispanic populations. The remaining 68 percent of the housing units fell in the low non-White or Hispanic concentration stratum.

Tenure

The 2005 test included three alternate versions of the tenure question.

1. The first experimental question dropped the term "cash" from the two renter response options, since past research indicates that respondents have difficulty with the concept of "cash rent" when rent is often paid by check.

2. The second question added an instruction to include home equity loans following the first owner option – *owned with a mortgage or loan*. This treatment was intended to help respondents understand that a home is considered "mortgaged" if a home equity loan is present.

3. Finally, the third version made both changes mentioned above in order to test both of these treatments simultaneously.

Results show:

- Dropping the term "cash" from the renter response options illustrated a positive effect in that it lowered item nonresponse.

- Adding the loan instruction resulted in fewer households marking "owned free and clear," which was a desired effect of this treatment.

Recommendation:

- Drop the word "cash" from the two renter response options.

- Add the instruction to include home equity loans following the first owner option, *owned with a mortgage or loan,* to clarify the issue of home equity loans for respondents.

Relationship

The purpose of the relationship test objective was to clarify response categories and eliminate known sources of confusion. Note that these changes were tested together in one experimental question version.

Changes included:

- Changing "Natural-born son/daughter" to "Biological son or daughter" because prior testing indicated that the term "Natural-born son/daughter" received unfavorable reaction from adoptive parents. In addition, it translates to "born out of wedlock" in colloquial Spanish.
- Changing "Foster child" to "Foster child or foster adult" to address those persons who are 18 years of age or older but live in foster-type situations.
- Replacing the slashes "/" and commas "," with the word "or" to help clarify the relationship question.
- Eliminating the write-in field for "Other relative" because past research showed a relatively large number of write-ins were non-relatives, uncodeable data, duplicates of another checkbox response category, or foreign language equivalents.
- Removing the spanner "If NOT RELATED to Person 1:" to discourage respondents from marking more than one response box.

Results show:

- No difference in item nonresponse between the control and the relationship panel.

- There were significantly more multiple responses in the control panel (0.3 percent) than in the experimental panel (0.1 percent) at the national level. This difference was also observed in the high non-White or Hispanic stratum. Therefore, the relationship question changes appear to be helpful in reducing the selection of multiple relationship categories.

- In the high non-White or Hispanic stratum, the "Other Relative" category showed higher response in the relationship panel (without a write-in option) than the control panel (2.4 percent to 1.6 percent). Note that there were no significant differences in the percent of people selecting the "other relative" category at the national level. Thus removing the write-in field appears to have no deleterious effect on the distribution.

- Overall, the remaining changes appeared to have no significant effects on distributions.

Recommendation:

- Make all the changes to the relationship question as described above.

- Conduct future research to look at the effect of adding examples to the "Other relative" category (aunt, uncle, cousin, niece, nephew, etc.) to help clarify what is meant by "other relative."

Age

The age experiment tested the addition of an instruction to direct respondents to report babies as age zero when the child is less than one year old. The instruction was tested to reduce the tendency of parents to report their baby's age in months, weeks, or days.

Results show:

- No difference in item nonresponse to the age question with or without the instruction.

- For babies aged zero (based on their date of birth), we found that including the age instruction increased the reporting of age zero and decreased the erroneous reporting of ages one through 11.

- Results also show that, for babies aged zero (based on their date of birth), item nonresponse for reported age was significantly lower in the presence of the age instruction.

- For the full age distribution, the percent of persons with a reported age of zero was significantly higher with the age instruction than without. Distributions for the other age groups did not change significantly.

Recommendation:

Add the instruction "*Please report babies as age 0 when the child is less than 1 year old"* to the age question.

1. BACKGROUND

In preparation for the 2010 census, the Census Bureau is conducting a series of tests. In late 2005, a mailout/mailback national test was conducted using variations of questionnaire content and various methods to increase response to the Census, including replacement questionnaire methods. The test also included the Internet as an optional mode for completing the census short form. Census Day was September 15, 2005.

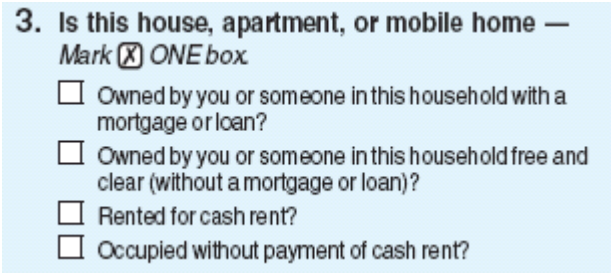
The objectives for the 2005 National Census Test (NCT) were:

- Test methods to improve completeness and accuracy of reporting for short-form items, including tenure, relationship, age and date of birth, and race and Hispanic origin.
- Test ways to reduce respondent and data capture errors, and improve respondent friendliness in mail and Internet modes.
- Test ways to improve coverage accuracy by reducing omissions and erroneous enumerations, and/or flagging potential errors for coverage followup interviews.
- Test ways to improve the operational feasibility of the second mailing.
- Test ways to improve self-response and maintain data quality by mailing bilingual questionnaires.

This analysis focuses on the experimental panels in the 2005 NCT that were designed to test changes to the tenure, relationship, and age questions. The Population Division (POP) and the Housing and Household Economic Statistics Division (HHES) proposed this research for the 2005 NCT to address the research questions listed in the following sections.

1.1 Tenure

Three experimental panels were designed to test wording variations in the tenure question, specifically changes to the response categories. These three panels were compared to the control panel that contained the Census 2000 tenure question (shown below).



3. Is this house, apartment, or mobile home —
Mark (X) ONE box.

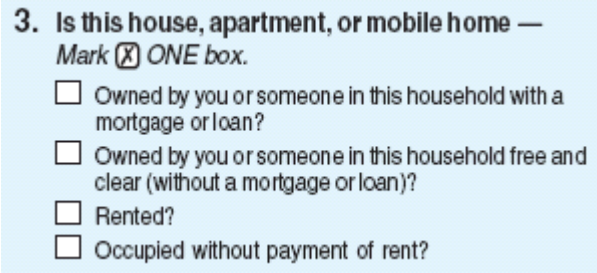
- ☐ Owned by you or someone in this household with a mortgage or loan?
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented for cash rent?
- ☐ Occupied without payment of cash rent?

The three treatment panels were designed to address the following research questions:

1.1.1 What is the impact on response behavior of dropping the term "cash" from the two renter response options?

Analysts have long advocated this change to the renter response options since research indicates that respondents have a difficult time with the concept of "cash rent" given that most rent is paid by check rather than cash (Hunter and DeMaio, 2004). Moreover, there was relatively low reliability in Census 2000 for the category "occupied without payment of cash rent" (Singer and Ennis, 2003).

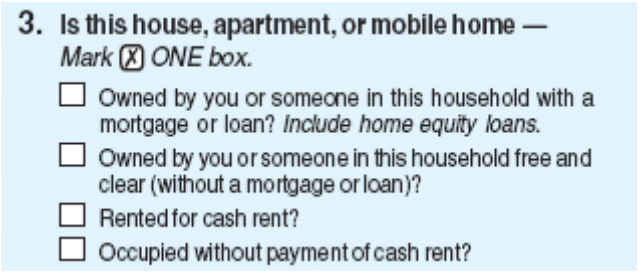
The question in this experimental panel appeared as:


3. Is this house, apartment, or mobile home —
Mark ☒ ONE box.
☐ 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?
☐ Occupied without payment of rent?

1.1.2 What is the impact on response behavior of adding the instruction "Include home equity loans" following the first owner option - owned with a mortgage or loan?

Results from the Census 2000 also showed relatively low reliability for "owned with a mortgage" and "owned without a mortgage" (Singer and Ennis, 2003). Housing data analysts have suggested that part of the confusion may stem from the fact that certain respondents may not understand that a home is considered "mortgaged" if a home equity loan is present (Hunter and DeMaio, 2004).

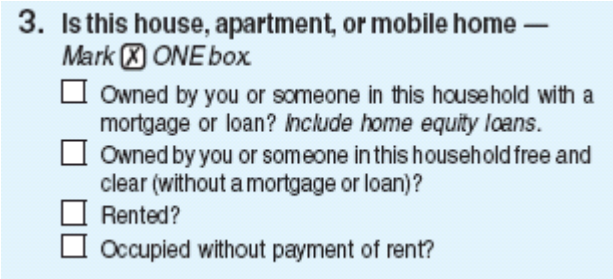
The question in this experimental panel appeared as:


3. Is this house, apartment, or mobile home —
Mark ☒ ONE box.
☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
☐ 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?

1.1.3 *What is the impact on response behavior of dropping the term "cash" from the two renter response options AND adding the instruction "Include home equity loans" following the first owner option - owned with a mortgage or loan?*

The last part of this tenure analysis looks at the results of combining the two changes into one test question to see if the combination of the changes has any impact on the results.

The question in this experimental panel appeared as:


3. Is this house, apartment, or mobile home — *Mark ☒ ONE box.*

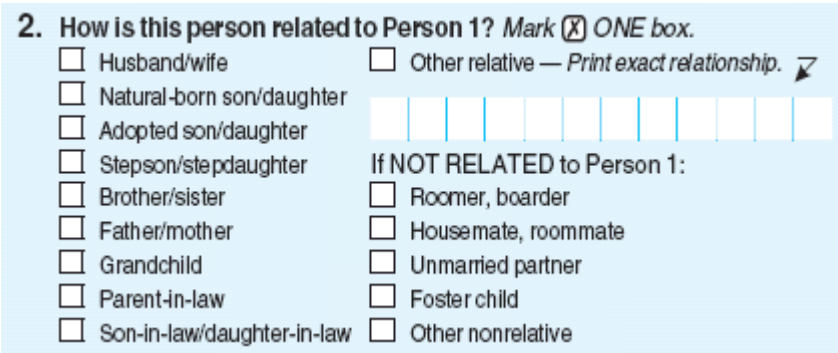
- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

1.2 Relationship

The relationship test objective proposed testing terminology and format changes to the relationship response options.

The purpose of this testing was to determine if the changes listed below affected responses to the relationship question.

The relationship question in the control panel appeared as:


2. How is this person related to Person 1? *Mark ☒ ONE box.*

- ☐ Husband/wife
- ☐ Natural-born son/daughter
- ☐ Adopted son/daughter
- ☐ Stepson/stepdaughter
- ☐ Brother/sister
- ☐ Father/mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law/daughter-in-law
- ☐ Other relative — *Print exact relationship.*

If NOT RELATED to Person 1:

- ☐ Roomer, boarder
- ☐ Housemate, roommate
- ☐ Unmarried partner
- ☐ Foster child
- ☐ Other nonrelative

The following items were tested as a set in one experimental panel.

1.2.1 Changing the wording "Natural-born son/daughter" to "Biological son or daughter"

POP analysts have noted that the phrase "natural-born" has received unfavorable reaction in the past from adoptive parents. Cognitive research showed that some respondents thought "natural-

born" meant that no drugs were involved in the birth, natural as opposed to caesarian birth, or natural conception as opposed to in-vitro fertilization (Hunter and DeMaio, 2004). In addition, it translates as "born out of wedlock" in colloquial Spanish. Other Census Bureau surveys, such as the Survey of Income and Program Participation, already use the word "biological" as opposed to "natural."

1.2.2 Changing the wording of "Foster child" to "Foster child or foster adult"

For the 2004 Census Test, the special place/group quarters team proposed adding a new category in the relationship question to identify persons receiving care in group homes. The change was intended to address the high rate of persons who were reported as both foster children and aged 18 and over (Hunter, Schwede, and Aaker, 2003). The 2004 Census Test used "Other non-relative receiving formal care."

For the 2005 NCT, subject matter experts proposed using "Foster child or foster adult" in lieu of "Other non-relative receiving formal care."

1.2.3 Using the conjunction "or" instead of a slash "/" or a comma "," where appropriate for the relationship responses to help clarify the question

This change was made to help clarify the response categories.

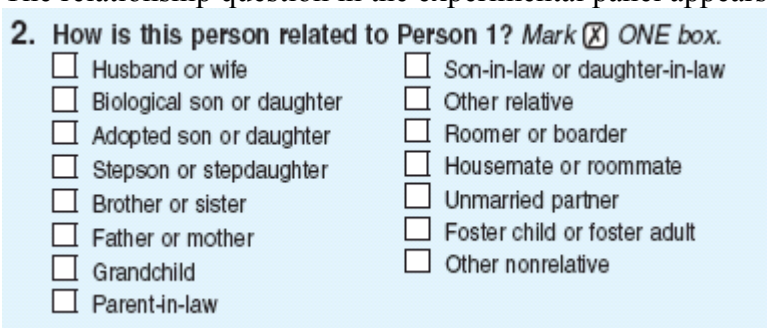
1.2.4 Excluding the write-in for "Other relative"

Census 2000 included an "Other relative" write-in to the relationship question. POP analysts noted that a large number of write-ins contained responses that were not relatives, nonsense or information that could not be coded, duplicates of response categories listed on the form, and foreign language equivalents (e.g., "hermano" instead of "brother"). The 2005 NCT tested the exclusion of the write-in field to determine what effect excluding the write-in option has on the distributions.

1.2.5 Deleting the spanner "If NOT RELATED to Person 1:" above the "nonrelative" categories

This change was made to try to reduce the number of multiple responses.

The relationship question in the experimental panel appears as:


2. How is this person related to Person 1? *Mark* ☒ *ONE box.*

- ☐ Husband or wife
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Unmarried partner
- ☐ Foster child or foster adult
- ☐ Other nonrelative

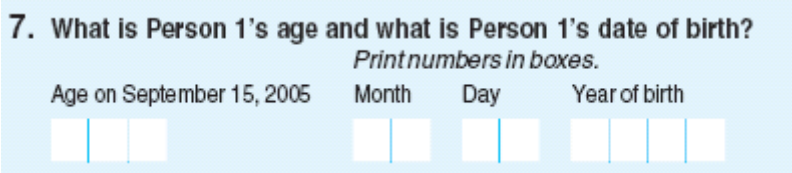
1.3 Age

The age objective proposed adding an instruction to direct respondents to report babies as age zero when the child is less than one year old.

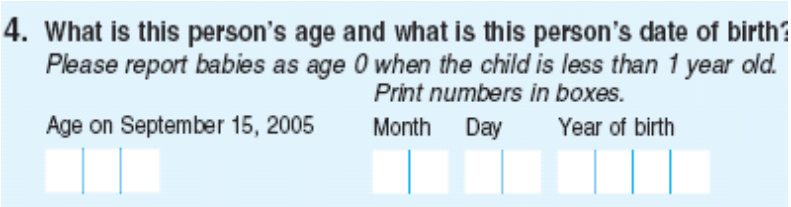
In earlier censuses, census takers were instructed to put age in monthly fractions for infants below age one (U.S. Census Bureau, 2002). In recent censuses, it was expected that babies less than one year old be reported as age zero, since age distributions are based on years.

The age zero population historically has had their ages reported in months instead of years. As Spencer and Perkins (1998) note, "There is a serious problem with reported age 0. It is only 25% of the size it should be, apparently because parents are coding their babies' age as 1 to 11 'months' rather than answering 0 'years'." It appears that people tend not to think of their babies as having lived zero years, but instead as having lived a certain number of months (e.g., one month, six months, or 18 months, etc.), or even in weeks or days.

The age question on the Census 2000 form and on the 2005 NCT control panel appeared as:



In an effort to correct this problem, one of the panels in the 2005 NCT included the instruction to direct respondents to report babies less than one year old as age zero. The age question on this experimental panel appeared as:



Note that another report from the 2005 NCT focuses on the experimental treatments that tests the effect of reversing the age and date of birth questions. Please see the Martin (2006) forthcoming report for the results of that testing.

2. METHODOLOGY

This section provides a description of the 2005 NCT design and contact strategies, as well as a description of response rate calculations and analytical methods.

2.1 Panel Design

The 2005 NCT was comprised of 20 experimental panels[1]. For this analysis we studied the control panel and four treatment panels. Three of these treatment panels contained the tenure treatments: one for removing the word "cash," one for adding the home equity loan instruction, and one for removing the word "cash" and adding the home equity loan instruction. A fourth treatment panel used in this analysis included the changes to the relationship question and the instruction for the baby's age. Each panel used in this analysis consisted of 30,000 sampled housing units. Also note that the experimental version of the tenure, relationship, and age items appeared in the same questionnaire treatment panels as an experimental version of the race and Hispanic Origin series. Each experimental panel tested multiple treatments to increase survey design efficiency and contain costs. We have no reason to hypothesize an interaction between the race and Hispanic Origin series and the age, relationship, and tenure questions.

2.2 Response Mode

Housing units in each experimental panel in the 2005 NCT were invited to respond by a paper questionnaire or the Internet. The Internet form did not, however, contain each of the experimental treatments. 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 embedded in the paper questionnaire. Therefore, this analysis focuses on responses returned via the paper questionnaires, and excludes any households that responded by the Internet.

2.3 Mailing Strategy

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

The second mailing was an initial questionnaire package. Housing units received a paper questionnaire and a first-class postage-paid return envelope. Also included in the mailing package was a letter from the Census Bureau's Director that encouraged households to respond and provided the option of responding by Internet.

The third mailing was a reminder postcard. The reminder postcard included a statement reminding households to respond to the census test if they had not already done so. It also provided instructions so that households could use the Internet to respond.

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 prior to September

1 The control panel had five components to it. The analyses in this report used the first component only.

13, 2005. Accompanying the questionnaire was a letter from the Director urging response and providing instructions for using the Internet.

Note that there was no telephone or personal visit followup for nonresponding households in the 2005 NCT.

2.4 Sample Design and Standard Errors

The housing units in the 2005 NCT were selected from mailout/mailback areas of the country. The housing units selected for the tenure, relationship, and age panels (30,000 housing units per treatment panel) and its corresponding control panel (30,000 housing units) were equally allocated to two strata that reflect differences in the racial and ethnic composition, and, hence, response propensity of the mailout/mailback universe. The high non-White or Hispanic concentration stratum, which encompassed roughly 32 percent of housing units in the universe, contained a high proportion of the non-White or Hispanic populations. The remaining 68 percent of the housing units fell in the low non-White or Hispanic concentration stratum. For more information about the creation of the strata, please see Bentley (2005). All estimates in this report are weighted to account for the oversampling of the high non-White or Hispanic stratum.

We computed standard errors for all estimates using a stratified jackknife replication procedure. This computation method accounted for the stratification in the sample, which we expect to lower the standard errors compared to a simple random sample. Clusters of housing units (or housing units selected at each hit) were assigned sequentially to one of 250 replicates. This assignment approach also accounted for the clustering of persons within a household in computing errors for person-level estimates, since persons within households were contained in the same replicate.

2.5 Calculation of Self-Response Rates

The self-response rate is a measure of respondent behavior with regard to responding to the census test. The formula for the self-response rate is presented below.

$$\textit{Self-response rate} = \frac{\textit{\# of nonblank, primary returns}}{\textit{panel sample sizes - UAA for the panel}} \quad \textit{x 100}$$

The denominator is the number of sample housing units minus those cases identified by the United States Postal Service (USPS) as "undeliverable as addressed" (UAA).[2] Cases that were UAA were defined as those housing units where there was no response (paper or internet), and both the initial questionnaire and replacement questionnaire mailings were flagged as UAA. Any housing units determined to be UAA were considered ineligible units.

The numerator is the number of sample housing units for which we received a nonblank[3],

2 For information on the definition of Undeliverable As Addressed, please see Rothhaas (2005a).

3 For information on the identification of blank forms, please Rothhaas (2005b).

primary paper return[4]. A census return was denoted as "blank" if fewer than two of the following items were completed: tenure, household count, name, relationship, sex, age or date of birth, Hispanic origin, race, and ancestry.

We selected a primary return when multiple responses were received for a given housing unit, using the following rules:

- When more than one paper return was received from a single household (i.e., we received both an initial and replacement questionnaire return), we selected the first nonblank form received. In the rare case that two nonblank paper forms were checked in on the same date and in the same batch, we selected the initial questionnaire return.

- When multiple Internet returns were received from a single household, we selected the first nonblank return.

- When paper and Internet returns were received for a single household, we selected the first nonblank return based on date received (i.e., check-in date/submitted date). If a nonblank paper and nonblank Internet return were received on the same day from the same household, we selected the paper return as it was probably mailed before the Internet was submitted.

The self-response rate was weighted to account for the sample design.

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 use this self-response rate because it is not subject to variation in UAA rates. Specifically, the denominator of the self-response rate excludes cases for which eligibility cannot be determined, such as units that are UAA. Therefore, any variation in the UAA rates across panels will not contribute to differences in the self-response rates.

Lastly, please note that the self-response rate defined here is not comparable to the Census 2000 mail response or mail return rates. The self-response rate is not a return rate in the sense that we do not definitively know the occupancy status of housing units included in the denominator or the status of cases that are excluded as UAAs.

4 When more than one return is received per ID, a primary return was selected for analysis using the criteria specified in Rothhaas (2004).

2.6 Calculation of Item Nonresponse Rates

Item nonresponse rates were computed as indicators of potential data quality issues. Item nonresponse refers to the percentage of records with missing data for a particular item. The analysis of item nonresponse rates was restricted to nonblank, primary paper returns for this analysis. The item nonresponse rates were calculated according to the following definition:

$$\textit{Item Nonresponse Rate} = \frac{\textit{number of records with missing data}}{\textit{total number of records from primary paper returns}} \times 100$$

Note that the definition of records depended on whether the analysis was at the household level or the person level. The tenure analysis was at the household level, while the age and relationship analyses were at the person level. For the household level item, the total number of records was defined as the number of household records from all nonblank, primary paper returns. For the person-level items, the term "records" refers to data-defined person records on all nonblank, primary paper returns. A data-defined person record had at least two entries that met specified criteria for the following items: name, relationship, sex, age/date of birth, Hispanic origin, race, ancestry. Please see Reiser (2005) for more information about the criteria used to determine data-defined status.

2.7 Analytical Methods

We evaluated the effectiveness of the experimental tenure, relationship, and age treatments using pairwise comparisons.

That is, we used pairwise comparisons to estimate differences of self-response rates, response distributions and item nonresponse rates between test panels. When an experiment, such as the 2005 NCT, is designed to compare the effects of several experimental treatments (including the "control"), it is necessary to utilize a Multiple Comparison Procedure (MCP) to ensure that a series of individual inferences (pairwise comparisons) does not compound the probability of committing an error by incorrectly stating that a difference is statistically significant when it is not.

To control the overall error rate among simultaneous comparisons, we created "families" of pairwise comparisons. A family is a collection of inferences for which it is meaningful to take into account some combined measure of error (Hochberg and Tamhane, 1987). Use of an MCP ensures that statements about the entire family of pairwise comparisons are made such that we are 90 percent confident all inferences in the family are correct. The use of an MCP is appropriate when a final decision requires that all inferences have a high probability of being correct. Specifically, MCPs require that larger differences exist between individual comparisons to be considered significant, when many comparisons are being simultaneously compared.

Computed differences were compared to critical values using one-sided or two-sided tests, depending on the hypotheses. One-sided tests were used when we had a clear hypothesis about the direction of the difference between the panels. Two-sided test were used when we

hypothesized no differences between the panels. For this analysis, we used either the Dunn MCP or the Dunnett MCP. The Dunn procedure was used for analyses in which experimental panels were compared to the control or to each other. Dunnett procedure was used to gain additional power when comparing each of several experimental panels only to the control. The comparisons were driven by the relevant hypotheses and test objectives; not all possible comparisons were made.

2.8 Quality Assurance Procedures

Quality assurance procedures were applied to the analysis and preparation of this report. The procedures encompassed data processing, data verification, factual content, technical writing, relevance, technical review, and clearance, as appropriate. A description of the procedures used is provided in the "Handbook for the Quality Process for 2010 Census Test Evaluations."

3. LIMITATIONS

3.1 Relationship Panel

Since the changes to the relationship question all occur in the same panel, we were not able to isolate the reason for any differences between this panel and the control panel.

3.2 Confounding Factors

Each experimental panel in this analysis tested multiple experimental treatments across various items on the questionnaire. In the creation of the experimental design, every effort was made to combine treatments such that confounding was limited. Thus, for the purposes of this experiment, we assume that combined treatments had no effect on each other.

3.3 Test versus Census

Note that results from a census test may differ from results in an actual Decennial Census due to differences in media attention, advertising and scope. We cannot determine whether public reaction to the age, relationship, and tenure questions would be different in a true Decennial Census environment.

3.4 Limitations with Evaluation Measures

We cannot assess the accuracy or reliability of responses for any of the items tested. Better measures of response error would emerge from a well-designed survey reinterview, however, this method was eliminated from the 2005 NCT design, in part, because of mode comparability and cost considerations. In general, the content treatments for the 2005 NCT were intended to improve clarity and/or presence of a response for each of the items. This study of alternative age, relationship, and tenure content items focused on retaining or improving item response and distributional accuracy. One aspect of data quality was measured by comparing item response rates between the experimental treatment and the control version. Another measure of data

quality was the comparison of the response distributions of the treatments to that of the control. Although this measure does not directly address response bias or reliability, distributional differences that follow the hypothesized trend may indicate a successful item alternative.

4. RESULTS

4.1 Self-Response Rates

Table 1 below contains the self-response rates for the control panel, each of the tenure panels, and the relationship/age panel at the national level (see appendix A for stratum-level rates). Additionally, the rates are further decomposed by response mode. While we have no reason to expect that these treatments would influence self-response rates, we present these rates to check on the assumption of equivalent response to each panel.

As expected, no significant differences were found between the control and each of the treatment panels at the national level. Note, however, that in the low non-White or Hispanic stratum (see appendix A) we found that the control panel had slightly higher self-response compared to the "drop cash" panel. We have no reason to believe that these relatively minor changes to the tenure response categories could impact unit response. Therefore, it may be possible that the race and Hispanic treatments in these panels resulted in this difference, since they contained substantial changes.

Table 1. Self-Response Rates (percents) by Mode

Panel	Total	Paper	Internet
Control	61.2	53.8	7.3
Tenure: "drop cash"	60.4	53.3	7.1
Tenure: "include loan instruction"	61.1	53.7	7.4
Tenure: "drop cash" and "include loan instruction"	60.8	53.3	7.5
Relationship/Age	60.7	53.7	7.0

Source: 2005 NCT housing unit analysis file

4.2 Tenure Analysis

Recall that the tenure analysis looks at three treatments: dropping the term "cash" from the two renter response options, adding an instruction to include home equity loans following the first owner option, *owned with a mortgage or loan*, and combining the two treatments above by dropping the term "cash" AND adding an instruction for home equity loans. Overall these changes to the tenure question were expected to show no significant change to item nonresponse and no significant changes to the response distributions when compared to the control.

4.2.1 Item Nonresponse Rates

Table 2 below contains tenure item nonresponse rates for the control panel and the tenure panels, both nationally and by stratum. Rates ranged from 1.2 percent to 3.3 percent.

Table 2. Tenure Item Nonresponse Rates (percents) for Mail Returns at the National and Stratum Level

Panel	National	Stratum	
		High Non-White or Hispanic Concentration	Low Non-White or Hispanic Concentration
Control	2.0	3.3	1.6
Tenure "drop cash"	1.4	1.8	1.3
Tenure "include loan instruction"	1.9	2.8	1.6
Tenure "drop cash/include loan instruction"	1.5	2.3	1.2

Source: 2005 NCT housing unit analysis file

Table 3 below illustrates pairwise comparisons of the control panel to the three experimental panels. The results show that the "drop cash" panel had significantly lower item nonresponse to the tenure item compared to the control panel, both nationally (0.6 percentage points lower) and particularly within the high non-White or Hispanic stratum (1.5 percentage points lower). These results may suggest that dropping the word "cash" from the renter response options clarified the tenure question for some respondents.

We observed the same trend between the control panel and the combined "drop cash and loan instruction panel." That is, the combined panel had significantly lower item nonresponse than the control on the order of 0.5 percentage points nationally, and 1.0 percentage point in the high non-White or Hispanic stratum.

Table 3. Tenure Item Nonresponse Rate Differences (percentage points) and Standard Errors for Mail Returns at the National and Stratum Level

Panel	National	Stratum	
		High Non-White or Hispanic Concentration	Low Non-White or Hispanic Concentration
Control – "drop cash"	**0.6*** **	**1.5*** **	0.3
	(0.2)	(0.3)	(0.2)
Control – "include loan instruction"	0.1	0.5	0.0
	(0.2)	(0.3)	(0.2)
Control – "drop cash/include loan instruction"	**0.5*** **	**1.0*** **	0.3
	(0.2)	(0.3)	(0.2)

Source: 2005 NCT housing unit analysis file

* Indicates difference is statistically significant with family-wise error rate at α=0.1 (using a two-tailed test).

** Indicates difference is statistically significant with family-wise error rate at α=0.05 (using a two-tailed test).

*** Indicates difference is statistically significant with family-wise error rate at α=0.01 (using a two-tailed test).

Because the combined drop cash/include loan instruction panel showed significantly lower item nonresponse, we tested the effectiveness of each individual treatment against the combination of the treatments to further study the effect of each treatment, as shown in Table 4.

The results concur with our previous findings. Removing the word "cash" from the renter response options, even in the presence of the loan instruction, results in significantly less item nonresponse to the tenure question.

Table 4. Tenure Item Nonresponse Rate Differences (percentage points) and Standard Errors for Mail Returns at the National and Stratum Level

Panel	National	Stratum	
		High Non-White or Hispanic Concentration	Low Non-White or Hispanic Concentration
"drop cash" - combined drop "cash" and include loan instruction	-0.1 (0.1)	-0.5 (0.3)	0.0 (0.2)
include loan instruction – combined drop "cash" and include loan instruction.	**0.4**** (0.2)	0.5 (0.3)	0.3 (0.2)

Source: 2005 NCT housing unit analysis file

* Indicates difference is statistically significant with family-wise error rate at α=0.1 (using a two-tailed test).
** Indicates difference is statistically significant with family-wise error rate at α=0.05 (using a two-tailed test).
*** Indicates difference is statistically significant with family-wise error rate at α=0.01 (using a two-tailed test).

.

4.2.2 Response Distributions

Next we looked at whether any of these tenure experimental treatments affected how people reported tenure. Table 5 shows the response distribution for tenure for the control and experimental panels. Missing data were excluded in the distributions, as were cases where the respondent marked multiple tenure response categories (about 0.2 to 0.4 percent of returns). Similar tables at the stratum level can be found in appendix B.

Table 5. Tenure Response Distributions (percents) at the National Level for Mail Returns

Panel	Owned with a mortgage or loan	Owned Free and Clear	Rented for cash rent	Occupied without payment of cash rent
Control	49.6	26.1	22.7	1.6
Drop "cash"	50.1	25.1	23.2	1.6
Include loan instruction	50.1	24.9	23.4	1.6
Drop "cash" and include loan instruction	50.5	24.9	23.1	1.6

Source: 2005 NCT housing unit analysis file

We compared the frequencies for each category between the control and experimental panels in Table 6 below. We found no differences in the tenure response distributions between the control and "drop cash" panel, both nationally and by stratum (see appendix B).

The "include loan instruction" and the "drop cash and include loan instruction" panels, however, had significantly fewer housing units that marked "owned free and clear" compared to the control, which is a favorable effect of adding the home equity loan instruction. Given this decrease in the "owned free and clear" response category, we expected to see a significant increase in the "owned with a mortgage or loan" category. We were, however, unable to detect any significant increase in households choosing the "owned with a mortgage or loan" category.

Table 6. Tenure Response Differences (percentage points) and Standard Errors at the National Level for Mail Returns

Panel	Owned with a mortgage or loan	Owned Free and Clear	Rented for cash rent	Occupied without payment of cash rent
Control-Drop "cash"	-0.5	1.0	-0.5	0.0
	(0.6)	(0.5)	(0.5)	(0.2)
Control-Include Loan Instruction	-0.5	**1.3***	-0.8	0.0
	(0.6)	**(0.6)**	(0.5)	(0.2)
Control-Drop "cash" and Include Loan Instruction	-0.9	**1.3****	-0.4	0.0
	(0.6)	**(0.5)**	(0.5)	(0.2)

Source: 2005 NCT housing unit analysis file

* Indicates difference is statistically significant with family-wise error rate at α=0.1 (using a two-tailed test).

** Indicates difference is statistically significant with family-wise error rate at α=0.05 (using a two-tailed test).

*** Indicates difference is statistically significant with family-wise error rate at α=0.01 (using a two-tailed test).

No differences in response distributions were found at the stratum level (see appendix B for the response distributions and differences by stratum).

4.3 Relationship Analysis

Recall that the relationship analysis tested a variety of changes to the response options. The result of changing "Natural-born son/daughter" to "Biological son or daughter" was expected to show slightly higher distributional totals for this response category in the relationship treatment panel compared to the control panel. Similarly, changing "Foster child" to "Foster child or foster adult" was expected to show slightly higher distributional totals for this response category in the treatment panel compared to the control.

No change to item nonresponse or response distribution was expected due to removing the write-in for "Other relative"; removing the spanner "If NOT RELATED to Person 1:"; or replacing the slash "/" and comma "," with the word "or."

Note that since the relationship question is only asked for persons two through six in the household, the tables in this section are computed for persons two through six only.

4.3.1 Item Nonresponse Rates

We start this analysis with a look at the item nonresponse rates. Table 7 compares item nonresponse rates for the experimental relationship item to the control. We found no differences between the control and the experimental relationship panel, which concurs with the hypothesis.

Table 7. Relationship Item Nonresponse Rates (percents), differences (percentage points) and Standard Errors for Mail Returns Only by Stratum

Panel	National	Stratum	
		High Non-White or Hispanic Concentration	Low Non-White or Hispanic Concentration
Control	0.6	0.9	0.5
Relationship Changes	0.7	1.1	0.5
Difference	0.0 (0.1)	-0.2 (0.2)	0.0 0.1

Source: 2005 NCT housing unit analysis file
* Indicates difference is statistically significant at the α=0.1 level (using a two-tailed test).
** Indicates difference is statistically significant at the α=0.05 level (using a two-tailed test).
*** Indicates difference is statistically significant at the α=0.01 level (using a two-tailed test).

4.3.2 Response Distributions

Next we looked at whether any of these relationship experimental treatments affected how people reported relationship. Table 8 compares the response distribution for the relationship categories of the control panel to the relationship experimental panel (see appendix C for relationship response distributions and differences by stratum). Missing data were excluded from the distributions.

The multiple-response category was indicated if two or more relationship categories were provided (through checkboxes or write-in). For example, if the "Grandchild" checkbox was marked and a write-in response was provided and coded to "Biological son or daughter," the relationship response was considered a multiple response. If the write-in response corresponded to the respondent's checkbox selection (e.g. the "Grandchild" checkbox was marked and the write-in was "Grandchild"), the case would not have been categorized as a multiple response.

Table 8. Response Distributions (percents), Differences (percentage points) and Standard Errors for Mail Returns Only[5]

Relationship	National		
	Control	Relationship	Difference
Husband or wife	38.2	38.0	0.2 (0.5)
Biological son or daughter	42.7	43.6	-0.9 (0.5)
Adopted son or daughter	1.3	1.1	0.1 (0.1)
Stepson or stepdaughter	1.9	1.8	0.1 (0.2)
Brother or sister	1.4	1.2	0.2 (0.1)
Father or mother	1.5	1.5	0.0 (0.1)
Grandchild	3.3	3.3	0.0 (0.2)
Parent-in-law	0.6	0.5	0.1 (0.1)
Son-in-law or daughter-in-law	0.5	0.6	0.0 (0.1)
Other relative	1.0	1.1	-0.1 (0.1)
Roomer or boarder	0.8	0.8	0.0 (0.1)
Housemate or roommate	2.0	1.6	0.4 (0.2)
Unmarried partner	3.2	3.4	-0.2 (0.2)
Foster child or foster Adult	0.2	0.1	0.1 (0.1)
Other nonrelative	1.2	1.2	0.0 (0.2)
Multiple Responses	0.3	0.1	**0.2** (0.1)**

Source: 2005 NCT housing unit analysis file
* Indicates difference is statistically significant with family-wise error rate at the α=0.1 level.
** Indicates difference is statistically significant with family-wise error rate at the α=0.05 level.
*** Indicates difference is statistically significant with family-wise error rate at the α=0.01 level.

5 A two-tailed test was used for all the relationship categories except for "Biological son or daughter" and "Foster child or foster adult," where a one-tailed test was used because of the specific directional hypothesis.

As stated earlier, we cannot definitively isolate the reason for any differences between the relationship treatment panel and the control panel, because all the changes occur in the same panel. We can, however, look at the differences in the specific response categories that are most likely affected by these changes.

- Natural-born son/daughter

  The results showed no significant difference at the national or stratum levels between the control and the experimental panel for this response category. Although the hypothesis is not supported, the results show us that we can make this change without any negative impact.

- "Other relative" write-ins

  Before doing any analysis of the relationship question, we redistributed write-ins from the control panel to their appropriate response category. Looking specifically at the remaining responses for the "Other Relative" category, no difference was found at the national level or for the low non-White or Hispanic stratum between the control and experimental panels. These results concur with the hypothesis.

  For the "Other Relative" category in the high non-White or Hispanic stratum, however, the relationship panel with no write-in option contained a significantly higher number of responses than the control panel (2.4 percent to 1.6 percent).

  Approximately 2.6 percent of the relationship responses in the control panel had an entry in the write-in space. Of these, about one-third (32.7) were provided as intended by the instructions, that is, the respondent wrote in a specific type of "Other relative" (e.g., cousin) and checked the "Other relative" box. The remaining two-thirds of the write-ins indicated that the respondent experienced some confusion in providing the write-in response. These response patterns are shown below (percent is out of all write-in responses):

  - 30.8 percent of the write-ins were coded to another checkbox category that the respondent marked,
  - 17.8 percent had no checkbox marked, and
  - 18.8 percent resulted in conflicting relationship responses.

- Foster child

  We broke the foster child response category into two groups: those whose reported age was under 18 and those with a reported age of 18 and over. The under-18 population showed no significant difference at the national level between the control and relationship panels. We were unable to determine estimates for those 18 and over in this response category because the cell sizes were too small.

Looking at this response category as a whole, we found no significant difference at the national level or stratum levels between the control and the experimental panel. Although the hypothesis is not supported, the results show that we can make this change without any negative impact.

- The spanner "If NOT RELATED to Person 1:"

  The only difference found nationally between the two panels was in the multiple-response category. There were significantly more multiple responses in the control panel (0.3 percent) than in the experimental panel (0.1 percent). A difference was also observed in the high non-White or Hispanic stratum. It's possible that this difference is related to the removal of the spanner, "If NOT RELATED to Person 1:", above the "nonrelative" categories.

4.3.2.1 Natural-born Son/Daughter Category by Hispanic Origin

As previously mentioned, the terminology "Natural-born son/daughter" translates as "born out of wedlock" in colloquial Spanish. Therefore we studied whether this change in terminology for the experimental panel had an impact on reporting of "son/daughter" when Person 1 (the assumed respondent) is Hispanic. Table 9 below compares the distribution of the control panel for the son/daughter category to the relationship panel at the national level and by Hispanic origin of Person 1.

The Hispanic Origin question between the control and the relationship panel was different; however, the experimental relationship question was also tested in one other panel whose primary focus was a space-saving design. This space-saving panel had the same Hispanic Origin question as the control and the same relationship question as the treatment thus, for this part of the analysis, we used that space-saving panel as our relationship panel.

Table 9. Response Distribution (percents) to Son or Daughter Category by Hispanic Origin OF PERSON 1 for Mail Returns Only

Panel	National	Hispanic Origin of Person 1	
		Hispanic	Not Hispanic
Control	42.7	50.3	41.8
Relationship	42.9	49.7	42.1
Difference	-0.2	0.6	-0.3
	(0.5)	(1.4)	(0.6)

Source: 2005 NCT housing unit analysis file

* Indicates difference is statistically significant at the α=0.1 level (using a one-tailed test).

** Indicates difference is statistically significant at the α=0.05 level (using a one-tailed test).

*** Indicates difference is statistically significant at the α=0.01 level (using a one-tailed test).

No differences were found between the panels at the national level, the stratum level, or by Hispanic origin of Person 1 at the national level. Note that item nonresponse to the Hispanic origin question for Person 1 in the space-saving/relationship panel is significantly higher at the

national level, at the high non-White or Hispanic stratum level, and at the low non-White or Hispanic stratum level. This difference could have an impact on the results in Table 9 above.

4.4 Age Analysis

Our last analysis focuses on age reporting for infants. Recall that the age objective proposed adding an instruction to direct respondents to report babies as age zero when the child is less than one year old. The analysis was expected to show a decrease in age item nonresponse and a shift in distribution (i.e., higher distribution for age zero).

4.4.1 Item Nonresponse Rates

First we looked at the item nonresponse rates to the age question. Table 10 compares the item nonresponse rate between the panel with the age instruction and the panel without the instruction. No difference was found in item nonresponse nationally or in the high non-White or Hispanic concentration stratum. However, in the low non-White or Hispanic concentration stratum, the item nonresponse rate for the age panel was significantly lower than the control panel (by 0.3 percentage points). This finding in the low non-White or Hispanic stratum concurs with the hypothesis.

Table 10. Item Nonresponse Rates (percents) for Reported Age for Mail Returns Only

Panel	National	Stratum	
		High Non-White or Hispanic Concentration	Low Non-White or Hispanic Concentration
Control	2.1	2.8	1.8
Age	1.9	3.0	1.5
Difference	0.2	-0.2	**0.3****
	(0.2)	(0.3)	(0.2)

Source: 2005 NCT housing unit analysis file
* Indicates difference is statistically significant at the α=0.1 level (using a one-tailed test).
** Indicates difference is statistically significant at the α=0.05 level (using a one-tailed test).
*** Indicates difference is statistically significant at the α=0.01 level (using a one-tailed test).

4.4.2 Response Distributions

Next we looked at the reported age distribution. Table 11 compares the reported age response distribution for the control panel to the panel that provides the instruction for baby's age. Missing data were excluded from the distributions.

The percent of persons with a reported age of zero was significantly higher in the age panel than the control panel, both nationally and in each stratum. Distributions for the other age groups did not change significantly nor was there any consistent trend in the direction of the differences in these age groups.

Because of the tendency for parents to report their infants in months instead of years, and because of the increase in age zero reporting, we expected to see a decrease in the one to 24 age group. Since there was no difference in this age group we broke the one to 24 group into smaller age groups (one to 11, 12 to 24, and individual ages from one to 11) to see if the size of the one to 24 age group was suppressing any significant trends. We saw no significant differences between the age panel and the control panel for these age groups. Even though the age zero reporting is significantly higher with the age instruction added, it appears that the numbers in the age group are so small (n=269 with instruction and n=89 without) that this shift in distribution has no impact on the other ages.

Table 11. Distributions (percents) for Reported Age for Mail Returns Only[6]

Reported	National			High Non-White or Hispanic Concentration			Low Non-White or Hispanic Concentration		
Age	Control	Age	Difference	Control	Age	Difference	Control	Age	Difference
0	0.3	0.8	**-0.6*** (0.1)**	0.3	0.8	**-0.5*** (0.1)**	0.3	0.8	**-0.6*** (0.1)**
1-24	27.9	28.0	-0.2 (0.4)	31.3	31.8	-0.5 (0.6)	26.8	26.8	0.0 (0.5)
25 – 44	23.2	23.7	-0.5 (0.4)	25.0	24.7	0.3 (0.6)	22.6	23.4	-0.8 (0.5)
45 – 64	30.6	29.7	0.9 (0.4)	28.2	27.7	0.5 (0.6)	31.4	30.3	1.0 (0.5)
65+	18.1	17.7	0.3 (0.4)	15.3	15.0	0.3 (0.5)	18.9	18.6	0.3 (0.5)

Source: 2005 NCT housing unit analysis file
* Indicates difference is statistically significant with family-wise error rate at the α=0.1 level.
** Indicates difference is statistically significant with family-wise error rate at the α=0.05 level.
*** Indicates difference is statistically significant with family-wise error rate at the α=0.01 level.

Next, we isolated the target population for this treatment (babies age zero) based on date of birth. The table below illustrates the distribution of reported age (under 11) for those with a computed age (based on date of birth) of zero. We did not break this out by stratum because of the small cell sizes.

This table shows that including the age instruction increases the reporting of age zero and decreases the erroneous reporting of ages one through 11 for babies less than a year old. The table also shows that, for those with a computed age of zero, item nonresponse for reported age is significantly lower in the presence of the added age instruction. These findings concur with the hypotheses.

6 A one-tailed test was used for the reported age categories of zero through 24 because of the hypothesis. A two-tailed test was used for the remaining age categories.

Table 12. Distributions of Reported Ages (percents) for those with Computed Age of Zero for Mail Returns Only

Reported Age	National		
	Control	Age	Difference
Missing	17.6	4.9	**12.7***** (2.9)
0	28.8	74.8	**-46.1***** (3.7)
1-11	49.7	16.8	**33.0***** (3.5)

Source: 2005 NCT housing unit analysis file
*Indicates difference is statistically significant with family-wise error rate at the α=0.1 level (using a one-tailed test).
** Indicates difference is statistically significant with family-wise error rate at the α=0.05 level (using a one-tailed test).
*** Indicates difference is statistically significant with family-wise error rate at the α=0.01 level (using a one-tailed test).

5. SUMMARY AND RECOMMENDATIONS

5.1 Tenure Conclusions

We found that removing the word "cash" from the two renter response options lowered item nonresponse. This result provides some support for previous findings that the word "cash" in the tenure question was confusing renters who typically pay by check. Based on this finding, we recommend dropping the word "cash" from the two renter response options.

Adding the home equity loan instruction resulted in significantly fewer housing units that reported "owned free and clear." Thus adding the instruction appears to be giving us the desired effect of clarifying for respondents that if they have a home equity loan, they do not own their home free and clear. We recommend adding the instruction to the tenure question to clarify the issue of home equity loans for respondents.

5.2 Relationship Conclusions

A positive effect of the changes to the relationship question was the reduction in multiple responses. Although we cannot isolate the reason for this reduction, this finding was the desired effect for at least one of the changes that were tested.

Other than the reduction in multiple responses, we found no effect on reporting at the national level. There is, however, an increase in the "Other relative" reporting in the high non-White or Hispanic stratum for the experimental panel without a write-in option compared to the control with a write-in option.

Approximately 2.6 percent of the relationship responses in the control panel had an entry in the write-in space. Of these, about one-third were provided as intended by the instructions, that is, the respondent wrote in a specific type of "Other relative" (e.g., cousin) and checked the "Other relative" box. The remaining two-thirds of the write-ins indicated that the respondent experienced some confusion in providing the write-in response.

We recommend making the changes to the relationship question as outlined earlier. The changes show no deleterious effects, and were actually helpful in reducing the selection of multiple relationship categories. Since there was no distributional change at the national level for the "Other relative" category, this may allow us to eliminate the write-in option and thus eliminate relationship coding.

Future research could look at the effect of adding examples to the "Other relative" category, such as aunt, uncle, cousin, niece or nephew, to help clarify what is meant by "other relative."

5.3 Age Conclusions

Results from the age experiment suggest that, although age zero reporting is significantly higher with the age instruction than without, the numbers in this age group are so small that this shift in distribution has no significant impact on the other ages.

Adding the age instruction clearly has the desired effect of informing parents to mark their infants below age one as age zero. Future research could look at ways of allowing parents to report their infant's age in a way that is more natural to them (i.e., in months, weeks, or days) and then converting that to years during the questionnaire editing/processing phase.

We recommend adding the instruction "*Please report babies as age 0 when the child is less than 1 year old"* to the age question since the results suggest the addition of the instruction increases accuracy of age reporting for babies less than a year old.

6. REFERENCES

Bentley, M. (2005) "Sample Specifications for the 2005 National Census Test", U.S. Census Bureau Internal Memorandum from Killion to Longini, March 9.

Hochberg, Y. and Tamhane, A.C. (1987) *Multiple Comparison Procedures*, (Wiley Series in Probability and Mathematical Statistics), John Wiley & Sons, Inc.

Hunter, J., Schwede, L. and Aaker, E. (2003) "Exploratory Research on Person in Custodial Care Relationship Category", U.S. Census Bureau, July 30.

Hunter, J. and DeMaio, T. (2004) "Report on Cognitive Testing of Tenure, Age, and Relationship Questions for the 2005 National Census Test", U.S. Census Bureau, November 22.

Martin, E. (2006) “Analysis of Space Saving Features and Reversal of Age and Date of Birth”, Forthcoming U.S. Census Bureau Internal Memorandum.

Reiser, C. (2005) “Definition of Data Defined Persons for the 2005 National Census Test”, Census Bureau Internal Memorandum for the Record, 2005 Analysis Team, May 31.

Rothhaas, C. (2004) “Duplicate Return Rule for the 2005 National Census Test—revised”, U.S. Census Bureau Internal Memorandum for the Record, 2005 Analysis Team, November 3.

Rothhaas, C. (2005a) “Definition of Undeliverable As Addressed Flag for the 2005 National Census Test”, U.S. Census Bureau Internal Memorandum for the Record, 2005 Analysis Team, March 15.

Rothhaas, C. (2005b) “Definition of Blank Forms for the 2005 National Census Test – revised”, U.S. Census Bureau Internal Memorandum for the Record, 2005 Analysis Team, May 24.

Singer, P. and Ennis, S (2003) “Census 2000 Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview”, Census 2000 Evaluation B.5, September 24.

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

Tancreto, J. and Bouffard, J. (2005) “2005 National Census Test Bilingual Form Analysis Plan”, U.S. Census Bureau Internal Memorandum, June, 8.

U.S. Census Bureau (2002) “Measuring America: The Decennial Census From 1790 to 2000”, POL/02-MA.

Appendix A

Self-Response Rates (percents) by Stratum and Mode

(Asterisked and bolded numbers indicate significance when compared to the control.)

Panel	Stratum					
	High Non-White or Hispanic Concentration			Low Non-White or Hispanic Concentration		
	Total	Paper	Internet	Total	Paper	Internet
Control	44.3	39.4	4.9	68.8	60.4	8.5
Tenure: "drop cash"	44.3	39.8	4.5	**67.7***	59.4	8.3
Tenure: "include loan instruction"	44.6	39.9	4.7	68.6	60.0	8.6
Tenure: "drop cash" and "include loan instruction"	44.2	39.5	4.7	68.4	59.6	8.8
Relationship/Age	44.9	40.3	4.6	67.9	59.8	8.1

Source: 2005 NCT housing unit analysis file

* Indicates difference is statistically significant with family-wise error rate at α=0.1 (using a two-tailed test).

** Indicates difference is statistically significant with family-wise error rate at α=0.05 (using a two-tailed test).

*** Indicates difference is statistically significant with family-wise error rate at α=0.01 (using a two-tailed test).

Appendix B

Tenure Distributions (percents) by Stratum for Mail Returns

Tenure	High Non-White or Hispanic Concentration				Low Non-White or Hispanic Concentration			
	Panels				Panels			
	Control	Drop "cash"	Include Loan Instruction	Drop "cash" and Include Loan Instruction	Control	Drop "cash"	Include Loan Instruction	Drop "cash" and Include Loan Instruction
Owned with a mortgage or loan	42.4	42.4	43.0	43.4	51.7	52.5	52.2	52.6
Owned Free and Clear	20.1	19.7	19.5	18.8	27.9	26.8	26.4	26.7
Rented for cash rent	35.4	36.0	35.3	35.9	19.0	19.3	19.9	19.2
Occupied without payment of cash rent	2.1	1.9	2.1	1.9	1.4	1.5	1.5	1.5

Source: 2005 NCT housing unit analysis file – missing data and multiple tenure responses were excluded.

Tenure Response Differences (percentage points) and Standard Errors by Stratum for Mail Returns

Tenure	High Non-White or Hispanic Concentration			Low Non-White or Hispanic Concentration		
	Panels			Panels		
	Control-Drop "cash"	Control-Include Loan Instruction	Control-Drop "cash" and Include Instruction	Control-Drop "cash"	Control-Include Loan Instruction	Control-Drop "cash" and Include Instruction
Owned with a mortgage or loan	0.0	-0.6	-1.0	-0.8	-0.5	-0.9
	(0.9)	(0.9)	(0.8)	(0.7)	(0.7)	(0.7)
Owned Free and Clear	0.4	0.6	1.4	1.1	1.5	1.2
	(0.8)	(0.8)	(0.8)	(0.7)	(0.7)	(0.6)
Rented for cash rent	-0.6	0.0	-0.6	-0.3	-1.0	-0.3
	(0.8)	(0.9)	(0.8)	(0.6)	(0.6)	(0.5)
Occupied without payment of cash rent	0.2	0.0	0.2	0.0	0.0	0.0
	(0.3)	(0.3)	(0.3)	(0.2)	(0.2)	(0.2)

Source: 2005 NCT housing unit analysis file – missing data and multiple tenure responses were excluded.

* Indicates difference is statistically significant with family-wise error rate at α=0.1 (using a two-tailed test).

** Indicates difference is statistically significant with family-wise error rate at α=0.05 (using a two-tailed test).

*** Indicates difference is statistically significant with family-wise error rate at α=0.01 (using a two-tailed test).

Appendix C

Relationship Distributions (percents), Differences (percentage points) and Standard Errors for Mail Returns by Stratum[7]

Relationship	High Non-White or Hispanic Concentration			Low Non-White or Hispanic Concentration		
	Control	Relationship	Diff.	Control	Relationship	Diff.
Husband or wife	28.5	28.5	0.0 (0.6)	41.5	41.3	0.2 (0.6)
Biological son or daughter	45.2	45.2	0.0 (0.8)	41.9	43.1	-1.2 (0.6)
Adopted son or daughter	1.3	1.1	0.2 (0.2)	1.3	1.2	0.1 (0.2)
Stepson or stepdaughter	1.9	1.9	0.1 (0.3)	1.8	1.8	0.1 (0.2)
Brother or sister	2.2	2.2	0.0 (0.2)	1.1	0.9	0.2 (0.1)
Father or mother	2.6	2.4	0.2 (0.3)	1.2	1.2	-0.1 (0.2)
Grandchild	5.9	5.9	0.0 (0.5)	2.4	2.4	0.0 (0.3)
Parent-in-law	0.9	0.8	0.1 (0.1)	0.5	0.4	0.1 (0.1)
Son-in-law or daughter-in-law	0.9	0.9	0.0 (0.2)	0.4	0.4	0.0 (0.1)
Other relative	1.6	2.4	**-0.8** (0.3)**	0.9	0.7	0.2 (0.1)
Roomer or boarder	1.4	1.1	0.3 (0.2)	0.6	0.7	-0.1 (0.1)
Housemate or roommate	2.1	2.0	0.1 (0.3)	2.0	1.5	0.5 (0.2)
Unmarried partner	3.1	3.6	-0.6 (0.3)	3.2	3.3	-0.1 (0.2)
Foster child or foster Adult	0.3	0.2	0.1 (0.1)	0.2	0.1	0.1 (0.1)
Other nonrelative	1.6	1.7	-0.2 (0.3)	1.1	1.0	0.0 (0.2)
Multiple Responses	0.5	0.2	**0.3** (0.1)**	0.3	0.1	0.1 (0.1)

Source: 2005 NCT housing unit analysis file

* Indicates difference is statistically significant with family-wise error rate at the α=0.1 level.

** Indicates difference is statistically significant with family-wise error rate at the α=0.05 level.

*** Indicates difference is statistically significant with family-wise error rate at the α=0.01 level.

7 A two-tailed test was used for all the relationship categories except for "Biological son or daughter" and "Foster child or foster adult", where a one-tailed test was used because of the hypothesis.