Experimental Treatment Results for the Age, Relationship, and Tenure Items from the 2005 National Census Test

Joan M. Hill, Jennifer G. Tancreto, and Cynthia A. Rothhaas
U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20233

1. Background[1]

The U.S. Census Bureau conducted the 2005 National Census Test (NCT) in the fall of 2005 as part of the mid-decade testing program for the 2010 census. Testing for the decennial census traditionally includes several large-scale tests involving a national mailout of experimental questionnaires, which allows quantitative results to be generalized to a large portion of the United States.[2] The 2005 NCT was the last national test in the 2010 census testing cycle. This test was designed to examine questionnaire content and design, as well as methodology to improve the efficiency of data collection procedures and response to the census.

One of five main objectives of the 2005 NCT involved important questions on the census questionnaire. Specifically, this objective aimed to improve reporting for tenure, relationship, age, date of birth, race, and Hispanic origin. Subject matter experts proposed variations in question wording and response categories to address this objective. This paper focuses on the testing of alternative age, relationship, and tenure items.

2. Survey Design

For the 2005 NCT, approximately 420,000 housing units were selected across 20 panels. Approximately 30,000 housing units were selected for each of the three tenure panels, the age/relationship panel, and the control panel. The three control panel questions were similar to those used in Census 2000. The experimental version of the age item and the relationship test item appeared in the same questionnaire panel as an experimental version of the race and Hispanic Origin series to increase survey design efficiency and contain costs. The tenure item appeared in different panels from the age and relationship experimental versions; tenure panels also contained an

experimental version of the race and Hispanic Origin series. For a more detailed description of the 2005 NCT design, refer to Tancreto (2006).

The sampling frame was partitioned into two strata by population characteristics known to be correlated with differences in response. The two strata were based on race and Hispanic Origin population proportions within blocks using Census 2000 data. The resulting "High Nonwhite or Hispanic Concentration" (High NWHC) Stratum included blocks with a relatively high concentration of residents who were nonwhite, or a relatively high concentration of residents of Hispanic Origin. Approximately 29 percent of the mail area blocks were in the High NWHC Stratum. The remaining 71 percent of the mail area blocks made up the "Low Nonwhite or Hispanic Concentration" (Low NWHC) Stratum. Refer to Bentley (2005) for more detailed information on the formation of the 2005 NCT strata. Note that the sample for each treatment panel was equally allocated across the two strata.

3. Survey Implementation

The implementation strategy for the 2005 NCT included a series of household contacts which were intended to increase unit response, potentially reducing nonresponse bias. The reference date (Census Day) for the Test was September 15, 2005. Respondents were instructed to use this date to determine household composition and the age of household members.

The first of four survey contacts was an advance letter. Each housing unit in sample was sent an advance letter on August 22, 2005, which informed them of the coming request to complete a test questionnaire.

One week later, sample housing units were mailed an initial questionnaire package with a postage paid return envelope and a cover letter. A cover letter from the Census Bureau's Director invited recipients to respond by Internet (an additional Internet invitation appeared on the first page of the questionnaire). For the age, relationship, and tenure portion of the test, the Internet instrument was identical across panels since this objective focused on the paper self-administered mode. Since respondents to the Internet were not directly exposed to the experimental treatments for these items, results from the Internet data collection are not included in this analysis.

---

2 The mailout address sampling frame for the 2005 NCT comprised about 80 percent of the U.S. housing units; these units had city-style addresses based on the Census 2000 housing unit frame, not including Puerto Rico. Survey results can be generalized to responding households in mailout areas, that is, those who returned their questionnaires by mail.

A week after the initial questionnaire packages were mailed, reminder postcards were mailed to all sample housing units to encourage household members to respond if they had not already done so (and included an additional Internet invitation).

About two and a half weeks after the reminder was mailed, a replacement questionnaire was mailed to each non-responding housing unit. The replacement questionnaire was identical to the initial questionnaire in terms of experimental treatments and appearance.

There was no field followup component for the 2005 NCT non-responding housing units since this was deemed too costly for a census test that was implemented nationally.

4. Variance Estimation

The standard errors for the estimates in this report were computed using a stratified jackknife replication procedure with random groups. This computation method accounts for the stratification in the sample, which we expect to lower the standard errors compared to a simple random sample. This method also accounts for the clustering of people within a household when computing errors for person level estimates, since people within households were clustered together in the same replicate. Pairwise comparisons were made to test for differences in the analytical variables for the different treatment groups. Computed differences were compared to critical values using one-sided or two-sided tests, depending on the hypotheses. Critical values were adjusted, where appropriate, to maintain a family-wise error rate of $\alpha=0.10$ within treatment groups; Dunn's and Dunnett's procedures were used to adjust for multiple comparisons. Dunn's procedure was used for analyses in which one panel was compared to the control. Dunnett's procedure was used to gain additional power when comparing each of several experimental panels to the control. The comparisons were driven by the relevant hypotheses and test objectives; not all possible comparisons were made.

5. Housing Unit Self-Response Rates

In order to compare content measures such as item nonresponse rates and response distributions, the responding populations should be comparable. To test this, we compared the housing unit self-response rate[3] of

---

3 The self-response rate denominator is the number of sample housing units minus those cases identified by the United States Postal Service (USPS) as "undeliverable as addressed." The numerator is the number of sample households for which we received a nonblank return. We selected a primary return when multiple responses were received for a given housing unit.

the control panel to each of the three tenure panel self-response rates. In addition, we compared the control panel self-response rate to that of the age/relationship panel. We found no significant differences at the national level between the control panel's self-response rate (61.2 percent) and any of the tenure or age/relationship treatment panels, which ranged from 60.4 percent to 61.1 percent. At the stratum level, one of the tenure treatment panels had a significantly lower self-response rate than the control panel (by 1.2 percent). However, this tenure panel also contained substantial changes to the race and Hispanic Origin items which may have affected the self-response rate for that panel. The relatively small difference in response is not expected to affect the stratum level tenure treatment results.

6. Limitations

The most effective way to assess the effects of each proposed change would be to assign each treatment to a separate panel to isolate the results. However, due to efficiency and cost considerations, all of the changes to the relationship response categories were combined together within one panel.

In addition, each experimental panel in this analysis tested treatments for multiple items. In the creation of the experimental design, we avoided combining treatments that shared a common objective or where there was a hypothesized interaction. However, we cannot be certain that confounding due to the combination of treatments in a panel is nonexistent.

In general, the content treatments for the 2005 NCT were intended to improve clarity and/or response for each of the items. The 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 each treatment 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. Thus, we cannot assess the accuracy or reliability of responses for any of the items tested. Better measures of response error could emerge from a well-designed survey re-interview; however, this method was eliminated from the 2005 NCT design, in part, because of mode comparability and cost considerations.

7. Age Item Alternatives

In recent decennial censuses, questionnaires contained a request for each household member's age in years, as well as each household member's date of birth. Prior research shows that, for the age question, respondents have a tendency to report their children under one year old in months, weeks, or days as opposed to zero years. The 1996 Content Test results showed a substantial number of discrepancies between the age respondents reported compared to age calculated from birth date for children under one year old (Spencer and Perkins, 1998). In that Test, about three-quarters of the babies under one year had an age response other than zero years. Based on previous problems with zero age reporting, the 2005 NCT tested a new instruction that directed respondents to report their children under one year old as age zero.

7.1 Experimental Treatment for Age Item

The age item in the treatment panel for the 2005 NCT is shown below. The only difference in the age items between the control and treatment panels was that the latter contained the addition of the instruction "*Please report babies as age 0 when the child is less than 1 year old*."

Age Instruction Treatment (for Person 1 as example)

What is Person 1's age and what is Person 1's date of birth?
Please report babies as age 0 when the child is less than 1 year old."
Print numbers in boxes.
Age on September 15, 2005 Month   Day   Year of birth

The age instruction was expected to show a decrease in age item non-response and a distributional shift. We expected a higher proportion of persons reported as age zero and a corresponding lower proportion of persons with reported ages 1 through 11. This shift occurs when those who would have erroneously reported babies' ages in months (which would have been processed as years) instead report zero years.

7.2 Age Analysis Results

Item nonresponse rates[4] were computed as an indicator of data quality for the age item. We compared the item nonresponse rate for age from the panel that provides the instruction for baby's age to the control at the national and stratum levels. We found that, at the national level, item nonresponse rates were not statistically significantly different; they were about two percent. Similarly, in the High NWHC Stratum, the rates were not significantly different (about three percent). However, in the Low NWHC Stratum, item nonresponse for age in the instruction treatment panel was significantly (marginally) lower than that of the control panel (1.5 versus 1.8 percent, respectively).

In order to determine if the baby's age instruction was

---

[4] The analysis of item nonresponse was restricted to nonblank, primary paper returns. For the age and relationship analyses, item nonresponse rates are computed at the person level, whereas tenure rates are computed at the mail return level (i.e., housing unit level).

affecting response behavior in the expected way, we isolated the target population for this treatment, that is, babies less than one year old, based on date of birth.

Table 1 below illustrates, for babies with a computed age of zero based on date of birth, the distribution of reported age up to 11 at the national level.

Table 1. National Age Distribution (under 11) by Panel for Babies with Computed Age Zero (weighted percent)

Age Distribution	Panel	
	Control (Unwtd n=288)	Baby's Age Instruction (Unwtd n=339)
0	28.8	**74.8***
1 through 11	49.7	**16.8***
Missing	17.6	**4.9***

*** Indicates differences from the control panel estimate is statistically significant at the α=0.01 level.

Table 1 shows that including the age instruction had the desired effect of increasing the respondents' reporting of age zero (by over 46 percentage points) and decreasing the erroneous reporting of ages 1 through 11 for babies less than a year old (by 33 percentage points). The table also shows that, for babies having a computed age of zero, the proportion of missing data for age is

significantly lower in the presence of the baby's age instruction (by almost 13 percentage points).

For the full age response distribution (not restricted to babies), the increase in reported age zero in the presence of the baby's age instruction remained statistically significant at the national and stratum levels (0.8 percent compared to 0.3 percent for the Control Panel at each level). We did not see a significant decrease in the ages 1 through 11 group in the instruction panel for the full distribution since the number of babies who presumably would have shifted out of this group was small relative to the size of the entire group (i.e., children 1 to 11 years old).

Note that reporting age zero for babies under one year is most important when there is no date of birth present from which age could be calculated. We could not make a reliable comparison between panels for reported age zero with no date of birth because of very small cell sizes for this study.

8. Relationship Item Alternative

The Census Bureau proposed numerous changes to the relationship response categories based on prior research. Changes to the relationship question all occur in the same panel, thus we were not able to definitively isolate the reason for any differences between this panel and the control panel. The five changes included revising wording in two response categories, replacing slashes and commas with "or" throughout the response categories, eliminating the write-in field for "Other relative," and removing the spanner "If NOT RELATED to Person 1:" (these changes are described in detail below).

8.1 Experimental Treatment for Relationship Item

The relationship item in the treatment panel for the 2005 NCT is shown below. Differences between the control and treatment panels are described in this section.

Relationship Treatment Panel

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

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

Based on prior testing, the terminology "Natural-born son/daughter" received unfavorable reaction from adoptive parents. In addition, it translates as "born out of wedlock" in colloquial Spanish (Reiser and Rothhaas, 2005). As a result, one of the objectives of the relationship treatment panel was to compare the response distribution between the control panel showing "Natural-born son/daughter" with the treatment panel showing "Biological son or daughter."

Another test objective was to change "Foster child" to "Foster child or foster adult". The change was intended to address those persons who were reported as both foster children AND greater than or equal to 18 years old. The analysis compared the effect on the relationship distribution and specifically focused on the ages of those reported as a foster child or foster adult.

A third change to the relationship response categories involved the "Other relative" write-in space. The control

panel had a 12-character write-in field under the "Other relative" check box response category. The relationship treatment panel did not have this write-in space (and subsequently excluded the "Print exact relationship." instruction and corresponding arrow). The control panel had the "Other relative" category at the top of the second bank of response options, whereas the treatment had the "Other relative" category under "Son-in-law or daughter-in-law" in the second bank. The removal of the write-in option was proposed because past research showed a relatively large number of write-ins were non-relatives, uncodable data, duplicates of another check box response category, or foreign language equivalents (Reiser and Rothhaas, 2005). During processing and analysis, a response in the control panel's "Other relative" write-in field was assigned a relationship code corresponding to one of the 15 check box response categories (including "Other relative") and included in that group's distribution.

Another change to the relationship category involved the removal of the spanner "If NOT RELATED to Person 1:" which was placed below the "Other relative" write-in space on the control panel and just above the last five check box response categories. This change was intended to discourage respondents from selecting more than one relationship category; discrepant relationship responses required imputation during Census 2000 processing.

To help clarify the relationship question, the slashes and commas were replaced with the word "or" in nine out of 15 response categories in the relationship treatment panel. For example, "Husband/wife" became "Husband or wife" and "Roomer, boarder" became "Roomer or boarder."

8.2 Relationship Analysis Results

For the relationship item analysis, we compared the full response distribution between the control panel and the relationship treatment panel at both the national and stratum levels. This includes persons for whom more than one relationship response was provided (i.e., multiple responses). Although we compared proportions for select response categories, we cannot isolate the cause(s) of distributional differences since all response category changes were tested together in one panel.

Since the relationship question uses "Person 1" as the reference person, that question is not asked of Person 1. All analyses are based on responses (including redistributed responses from the write-in space on the control form) for Persons 2 through 6 only.

For item nonresponse, we found no statistically significant difference between the control and relationship treatment panel at the national level; no differences were found at the stratum level.

The relationship item distributional analyses focused on five experimental changes.

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

  The results showed no significant difference at the national or stratum levels between the control and the experimental panel for this response category (about 43 percent nationally).

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

  We found no significant differences at the national or stratum levels between the control and the relationship treatment panels for proportion of persons in this response category

(about 0.2 percent nationally). The "under age 18" population showed no significant difference at the national level.

We were unable to produce reliable results for those 18 and over in this response category because of small cell sizes.

- Excluding the write-in for "Other relative"

  After redistributing write-ins, we looked specifically at the remaining responses for the "Other Relative" category. No differences were found at the national level or in the Low NWHC Stratum between the control and relationship treatment panels (about 1 percent nationally). However, in the High NWHC Stratum, the proportion in the "Other relative" category in the relationship treatment panel with no write-in option was significantly higher than the control panel (2.4 percent compared to 1.6 percent, respectively). It is important to note that, overall, about 2.6 percent of responses were write-ins. Over two-thirds of the write-ins fit response patterns that indicated confusion, such as marking multiple check boxes, or marking 'grandchild' and providing the same response in the write-in space.

- Deleting the spanner "If NOT RELATED to person 1:" above the "nonrelative" categories to discourage respondents from marking more than one response box

  When looking at the multiple response category, the relationship treatment panel showed a significant decrease compared to the control panel in the number of multiple responses at the national level (0.3 percent and 0.1 percent, respectively) and in the High NWHC Stratum (0.5 percent and 0.2 percent, respectively) No significance difference was found in the Low NWHC Stratum.

- Replacing slashes and commas with "or" for the applicable response categories

  We found no significant differences at the national or stratum levels between the control and the relationship treatment panels for proportion of persons in any of the nine response categories where slashes and commas were replaced with "or."

9. Tenure Item Alternatives

The Census Bureau tested two treatments related to the response categories of the tenure item. One treatment involved adding an instruction to include home equity loans when considering the first owner response category and the other treatment involved dropping the adjective "cash" when referring to rent. The 2005 NCT tested each treatment within its own panel to isolate effects and also tested a third panel which contained the combination of the two treatments to account for any interaction effects.

9.1 Experimental Treatments for Tenure Item

The tenure items for the control panel and combination panel for the 2005 NCT are shown below.

Control Panel

Is this house, apartment, or mobile home -
☐   Owned by you or someone in this household with a mortgage or loan?
☐   Owned by you or someone in this household free and clear (without a mortgage or loan)?
☐   Rented for cash rent?
☐   Occupied without payment of cash rent? *SKIP to question 21*

"Drop cash./include loan instruction" Panel

Is this house, apartment, or mobile home -
☐   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? → *SKIP to question 21*

The "drop cash" panel (not shown) excluded "cash" from the two rent response options; the owner options were the same as the control panel. This treatment tested the impact on the tenure item nonresponse rate of dropping the term "cash" from the two renter response options. The term "cash" has long been criticized by respondents as well as census staff as not accurately depicting how rent is actually paid, for example, by check (Reiser and Rothhaas, 2005).

The "include loan instruction" panel (not shown) contained the added home equity loan instruction in the first owner response option; renter options were the same as the control panel. This treatment tested the effect of adding an instruction to include home equity loans following the first option - *owned with a mortgage or loan*.

We added this instruction because respondents may not recognize that home equity loans are liens against the home (Reiser and Rothhaas, 2005).

The "drop cash/include loan instruction" panel tested the effect of making both changes mentioned above – dropping the term "cash" and adding an instruction for home equity loans.

Table 2. Tenure Item Nonresponse Rates by Panel at the National Level and Stratum Levels (weighted percent)

Domain	Panel			
	Control (Unwtd n=13,786)	Drop "cash" (Unwtd n=13,766)	Include Loan Instruction (Unwtd n=13,839)	Drop "cash" + Loan Instruction (Unwtd. n=13,703)
National	2.0	**1.4*****	1.9	**1.5*****
High NWHC Stratum	3.3	**1.8*****	2.8	**2.3*****
Low NWHC Stratum	1.6	1.3	1.6	1.2

***Indicates differences from the control panel estimate is statistically significant at the $\alpha=0.01$ level.

The tenure item nonresponse rate for the "drop cash" panel was significantly lower than the control panel at the national level (1.4 percent compared to 2.0 percent, respectively) and in the High NWHC Stratum (1.8 percent compared to 3.3 percent, respectively). Similarly, the combined "drop cash/include loan instruction" panel had a significantly lower tenure item nonresponse rate than the control panel at the national level and in the High NWHC Stratum.

We found no significant differences in the tenure item nonresponse rate between the control panel and the "include loan instruction" panel, either nationally or by stratum.

Since this difference seems to be related to the "drop cash" feature, we tested (not shown) the effectiveness of each individual treatment against the combination of treatments (i.e., comparing item nonresponse for "drop cash" versus combined panel, as well as "include loan instruction" versus combined panel). The results showed that removing the word "cash" from the renter response options, even in the presence of the loan instruction, resulted in significantly less item nonresponse to the tenure question (a 0.4 percent difference nationally).

Table 3 shows national response distributions for each tenure panel including the control panel.

Table 3. Tenure Response Distributions by Panel at the National Level (weighted percent)

Tenure Response Options	Panel			
	Control (Unwtd n=13,447)	Drop "cash" (Unwtd n=13,514)	Include Loan Instruction (Unwtd n=13,525)	Drop "cash" + Loan Instruction (Unwtd. n=13,437)
Owned with a mortgage or loan	49.6	50.1	50.1	50.5
Owned free/clear	26.1	25.1	**24.9***	**24.9****
Rented for cash rent	22.7	23.2	23.4	23.1
Occupied without payment of cash rent	1.6	1.6	1.6	1.6

*Indicates differences from the control panel estimate is statistically significant at the α=0.1 level.
**Indicates differences from the control panel estimate is statistically significant at the α=0.05 level.

The combined "drop cash and include loan instruction" panel and the "include loan instruction" panel 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. We found no difference in tenure response distributions between the control and "drop cash" panels, either nationally or by stratum.

10. Conclusions

Results from the 2005 NCT age treatment panel showed that, for babies less than one year old, including the baby's age instruction had the desired effect of increasing the respondents' reporting of age zero and decreasing the erroneous reporting of ages 1 through 11. In addition, for babies less than one year, the proportion of missing data for age is significantly lower in the presence of the baby's age instruction. Finally, the item nonresponse rate for the age item with the baby's age instruction was generally not

different from that of the control (although lower than the control in one stratum). Thus, we recommend adding the baby's age instruction to help respondents report age zero for babies under one year.

Results for the relationship item show a beneficial reduction in the respondents' selection of multiple responses, which presumably resulted from removing the "If NOT RELATED to Person 1:" above the nonrelative categories. In addition, removing the "Other relative" write-in option generally had no effect on the proportion of those reported in the "Other relative" category after redistributing write-in responses to the existing check box categories (overall, about two-thirds of the write-in responses were coded back to an existing check box category).[5] This is a positive result in that the proportion

---

5 Results showed a higher proportion of persons in the "Other relative" category without the write-in option in the High NWHC Stratum. This suggests future research on the accuracy of reporting within the group.

of persons in the "Other relative" category would likely not be substantially affected by not having a write-in option yet the burden/expense of relationship coding would be eliminated. Overall, the other changes to the relationship categories, such as "Biological" rather than "Natural-born" son or daughter, including "foster adult" with the "Foster child" category, and replacing slashes and commas with "or" had no significant effect. Finally, we found no significant differences in item nonresponse between the relationship treatment alternative and the control. Thus results support implementing the changes tested in the relationship item alternative.

The combined "drop cash/include loan instruction" panel generally had a significantly lower tenure item nonresponse rate than the control panel. This difference is driven by the "drop cash" feature, since the combined treatment panel exhibited a lower tenure item nonresponse rate than the "loan instruction" treatment by itself. In addition, the combination of treatments (as well as the loan instruction treatment by itself) 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. Thus, we recommend dropping "cash" when describing rent, as well as including the home equity loan instruction to help respondents select the appropriate owner category.

11. References

Bentley, M., and Tancreto, J. (2006) "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. (2006) "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.

Reiser, C., and Rothhaas, C., (2005) Internal memorandum from Killion to Gore entitled, "2005 National Census Test Age, Relationship, and Tenure Analysis Plan," June 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.

Tancreto, J. (2006) "An Overview of the 2005 National Census Test", 2006 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.