Sample Design for the Census 2010 Experimental Program

Michael Bentley
U.S. Census Bureau[1]

Abstract
Since 1970, the U.S. Census Bureau has implemented an experimental program as part of the decennial census to evaluate a variety of alternative methodologies and questionnaire design strategies. For Census 2010, a robust program is planned that includes an ambitious test of race and Hispanic origin question changes, an alternative address collection for improved within-household coverage, and a 2000 Census short form-style questionnaire. In addition, other treatments include a deadline message with and without a compressed mailing schedule, and two alternative confidentiality messages. Finally, a nonresponse followup contact strategy experiment will examine the feasibility of reducing the number of interviewer contact attempts in order to save costs. This paper presents the sample design for each of these experiments, and includes a discussion of the unique challenges in achieving the optimal design.

Key Words: Census 2010, Experiment, Sample Design, Questionnaire

1. Introduction

The 2010 Census Program for Evaluations and Experiments (CPEX) is a comprehensive and multi-faceted program that includes a variety of studies to evaluate numerous decennial census operations, as well as an assortment of experiments designed to test potential improvements to the census questionnaire and other census components. This paper provides an overview and describes the sample design for each of the 2010 experiments, and includes a discussion of the unique challenges in achieving the optimal design.

2. Overview of 2010 Census Experiments

The largest component of the 2010 experimental program is the Alternative Questionnaire Experiment (AQE). The AQE includes a variety of different experimental panels that feature changes to the 2010 census mailout questionnaire. The primary focus is on an ambitious series of changes to the race and Hispanic origin questions, intended to improve the accuracy and reliability of race and origin self-reporting. These changes include:

- Combined race and Hispanic origin question,
- Modified examples in race categories,
- Modified examples for the Other Hispanic write-in,
- Hispanic origin question that allows respondents to mark multiple responses,
- Modified race question stem and format, and
- Modified race response category labels.

[1] *Any views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.*

Two other treatments are also being tested under the AQE umbrella: (1) a person-level alternative address collection for improved within-household coverage and to potentially reduce the workload for the coverage followup operation; and (2) a Census 2000 short form-style questionnaire. The latter evaluates the combined effect of all changes made to the mailout/mailback form since 2000.

In addition, two other mailout/mailback experiments are being conducted. The first tests a deadline message with and without a compressed mailing schedule, and is designed to evaluate the use of a series of messages that provide explicit due dates for respondents on the letters and reminder postcards. The purpose of the deadline messages is to increase mail response and, consequently, reduce the cost and time spent on following up with nonrespondents. The deadline message experiment is crossed with (and without) a compressed mailing schedule for the purpose of reducing the time frame between when the census materials are sent and the April 1 ("Census Day") reference date. Another experiment uses two alternative confidentiality notification messages on the cover letters that accompany the questionnaires. The purpose of this experiment is to notify respondents about how census data is used without jeopardizing cooperation and data quality.

Here is a brief summary of the mailout/mailback experimental panels, which total a mailout size of 800,000 housing units:

- 15 panels will be used to test the effects of various race and Hispanic origin changes (n=30,000 per panel);
- 2 control panels for comparison to the race and Hispanic origin treatments (n=30,000 per panel);
- 1 panel will be used to evaluate the redesigned coverage overcount question (n=30,000);
- 9 panels will be used to test a compressed mailing schedule and four different deadline message treatments (n=20,000 per panel);
- 2 panels will be used to test the modified confidentiality notification messages (n=20,000 per panel);
- 1 panel will be used to test the 2000-style questionnaire content (n=20,000);
- And 1 control panel for comparison to the deadline messaging/compressed schedule, confidentiality, and 2000 content panels (n=20,000).

Also, a nonresponse followup (NRFU) experiment is being conducted in order to determine if a reduction in the number of enumerator contact attempts can maintain the same level of data quality in order to save costs. Two alternative NRFU forms will feature a revised record of contact section in which four or five contact attempts are listed instead of the usual six. The sample size is about 600,000 per alternative.

Finally, another experiment, the Paid Advertising Heavy-Up Experiment, is designed to study the impact of additional paid advertising in certain areas of the country. Pairs of marketing areas will be matched based on their similarity for various demographic and behavioral characteristics. One pair will be treated as a control and the other will receive a planned increase in census advertising. This experiment was a very recent addition to the 2010 experimental program, so the details of the design of the heavy-up experiment are still under development and will not be presented further in this paper.

For purposes of optimal sample selection, a total of four unique designs are used: for the race and Hispanic origin panels (section 4); for the coverage overcount panel (section 5); for all other mailout/mailback experiments (section 6); and for the NRFU contact strategy experiment (section 7).

3. General Information on Mailout/Mailback Design

This section contains general information on the design of the mailout/mailback experiments, that is all of the experiments discussed in this paper except for the NRFU experiment. For mailout/mailback, the questionnaire and other materials are sent to housing units by mail and respondents are asked to mailback the completed questionnaires.

3.1 Sample Universe

The universe for the three sample selections is housing unit addresses in the mailout/mailback areas of the country. This only includes the 50 states and the District of Columbia. The following addresses will be excluded from the sample: addresses not in mailout/mailback areas, such as those in update/leave or update/enumerate areas[2]; addresses that are special places or group quarters; addresses in Puerto Rico and other island territories.

3.2 Targeted Replacement Mailing

All experimental households will be assigned to the targeted replacement mailing operation, in which households that have not responded by the cutoff date will receive a replacement questionnaire. This differs from the regular 2010 Census system, in which the replacement mailing is divided into three groups based on an area's anticipated mail response based on Census 2000 data: no replacement (for high response areas), targeted replacement (for medium response areas), or blanket replacement to all households (for low response areas).

3.3 Use of Experimental Treatment in Initial and Replacement Mailings

Each of the experimental panels, with the exception of the deadline message and compressed mailing schedule panels, will receive the same experimental treatment in both the initial mailing package and the replacement mailing. That is, the questionnaires and other form changes will be identical between the two mailings. Due to the unique time-sensitive nature of the deadline message and compressed mailing schedule panels, these households will simply receive regular 2010 replacement mail materials.

[2] *Update/leave* areas are those in which an enumerator hand-delivers a questionnaire to each housing unit and then updates the relevant map and address information. The mailout/mailback approach is not suitable in some areas where in which many units lack mailing addresses that identify their geographic location, or in locations without mail delivery.
Update/enumerate areas are those in which an enumerator collects the census information on the spot from the household. This includes a variety of enumeration areas that contain special requirements, such as American Indian reservations, areas with a high number of seasonal vacants, and other selected areas.

3.4 Strings of Sample Housing Units

At each sample hit, a string of consecutive housing units (Census IDs) will be selected for panel assignment (e.g., one unit is assigned to one panel, the next unit is assigned to another panel, etc.). This will be a "next-door neighbor" approach. In other recent sampling, such as in the 2003 and 2005 National Census tests, an every-third housing unit approach was used with the experimental panels. This was done to minimize any possible effects of clustering or contamination from one panel to another. However with the 2010 CPEX panels, to the extent possible, it will be advantageous to use strings of housing units. This is because we want to minimize the number of cases where one unit gets a bilingual form and their neighbor gets an experimental form (English-only).

4. Sample Design for the Race and Hispanic Origin Panels

In order to support the objectives of the AQE race and Hispanic origin experiments, we used a complex sample design that is intended to reach the appropriate populations of interest for the proposed treatments, while still providing accurate national estimates. The design concentrates on achieving a significant mailout to minority demographic groups such as Hispanic or Latino people, Asian people, and Black or African American people. The universe consists of census tracts likely to contain relatively high proportions of these subgroups. The final sample of housing units will be selected from within these areas. Three years of American Community Survey (ACS) data from 2005-2007 were used to stratify each area.

In order to optimize the sampling process, we used a hierarchical tract selection process. The hierarchical stratification approach is based purely on the relative size of the population subgroup. That is, we give the smallest subpopulation priority to ensure that an adequate number of tracts are included in the corresponding stratum. The hierarchy is detailed as follows:

1) Tracts with 15 percent or more Asian, Native Hawaiian, or Other Pacific Islander (NHOPI) people
2) Tracts with 25 percent or more Black or African American people
3) Tracts with 40 percent or more Hispanic or Latino people
4) All other tracts

The stratification yields significant differences across a number of different characteristics including mail response rates and the race and Hispanic origin distributions. The majority of housing units are in the "All other" stratum (estimated 89.1 million housing units from three years of ACS data), with the smallest number in the Asian/NHOPI stratum (estimated 8.3 million). Refer to Table 1 for a general overview of the estimated demographics in each of the strata.

Table 1: Statistical Estimates Overview of Race and Hispanic Origin Design by Stratum

Stratum	Total HUs	Percent Mail Response	Percent Asian/NHOPI	Percent Black	Percent Hispanic/Latino	Percent Minority
Asian/NHOPI	8,315,237	50.4	29.9	7.3	19.0	59.2
Black	18,548,256	35.4	1.8	56.1	11.5	70.9
Hispanic	10,311,984	32.8	3.0	5.7	68.0	77.9
All other	89,067,569	53.9	2.6	4.4	7.2	16.5

Source: American Community Survey (ACS) 2005 - 2007

The results are not too surprising. With this design, each of the first three strata has a significantly high percentage of minorities (Hispanic or non-White people), ranging from 59.2 percent to 77.9 percent of the respective populations. In the Asian/NHOPI stratum, there is a moderately high 29.9 percent of Asian or NHOPI people overall, compared to less than 3 percent for the other three strata. In the next stratum (Black) there is a large increase in the Black population (56.1 percent). The Hispanic stratum consists of 68 percent people identified as Hispanic or Latino, and contains nearly half of all Hispanics in the country. Finally, the "All other" stratum is comprised primarily of non-Hispanic White people (just 16.5 percent minority).

The sample allocation for the race and Hispanic origin panels includes a disproportionate oversampling to the Asian/NHOPI, Black, and Hispanic strata in order to measure statistically significant differences between the experimental panels and the control panel. There will be a total of approximately 9,000 hits in the "Asian/NHOPI Stratum;" 9,000 hits in the "Black Stratum;" 9,000 hits in the "Hispanic Stratum;" and "3,000 hits in the "All other Stratum."

Table 2: Targeted Stratum Sizes for Each Race and Hispanic Origin Panel

Stratum	Mailout Size per Panel (# of HUs)
Asian/NHOPI Stratum	9,000
Black Stratum	9,000
Hispanic Stratum	9,000
All other Stratum	3,000
Total	30,000

5. Sample Design for the Coverage Overcount Panel

The sample design for the AQE overcount panel concentrates on reaching areas of the country expected to have significant numbers of households susceptible to coverage overcounts. Coverage overcounts are people that have indicated that they sometimes live or stay somewhere besides the address to which the form was mailed. Given that the focus is on reaching the maximum number of possible overcounts, rather than making any national estimates or comparisons, there is no comparison to a control panel. Instead the primary goal is to compare and evaluate a series of business rules for determining true erroneous enumerations.

The design includes areas with a dense active-duty military population; areas with a large number of older people (i.e., possible nursing home residents); urban areas with higher poverty levels (i.e., potentially more incarcerated people - see Patterson (2006), for

example, for a study showing a correlation between poverty and crime); areas with potentially more college-age students; areas with potentially higher numbers of seasonal residences; and areas more likely to have child custody coverage issues. The goal was not to simply identify areas where the alternative residence is located, but the households *where they might be coming from*. For instance, not areas with college dorms or nursing homes, but areas more likely to have households with students attending college, or households with an elderly relative. Tracts not identified under one or more of these conditions were excluded from the sample selection:

1) The "College Stratum" includes tracts where 30 percent or more of the HUs had a young person between the ages of 6-22, and a college-educated adult (Bachelor's Degree or higher) between the ages of 40-60.

2) The "Child Custody Stratum" includes tracts where 30 percent or more of the HUs had a separated or divorced adult and the presence of a child less than 18 years old.

3) The "Military Stratum" includes tracts where 70 percent or more of the HUs had at least one person in active-duty military since September 2001.

4) The "Nursing Home Stratum" includes tracts where at least 70 percent of the HUs had the presence of someone age 70 years or older.

5) The "Jail Stratum" includes tracts where at least 50 percent of the HUs were located in an urban area and met the poverty definition.

6) The "Seasonal Stratum" includes tracts where at least 50 percent or more of the HUs had a person between the ages of 50-70 and a household income of at least 100,000 dollars.

The design includes an allocation of HUs in each of the six overcount strata that takes into consideration the different selection criteria thresholds (see Table 3). There will be a total of approximately 6,000 hits in the "College Stratum" and "Child Custody Stratum," 4,000 hits in the "Military Stratum" and "Nursing Home Stratum," and 5,000 hits in the "Jail Stratum" and "Seasonal Stratum".

Table 3: Targeted Stratum Sizes for Overcount Panel

Stratum	Mailout Size per Panel (# of HUs)
College Stratum	6,000
Child Custody Stratum	6,000
Military Stratum	4,000
Nursing Home Stratum	4,000
Jail Stratum	5,000
Seasonal Stratum	5,000
Total	30,000

6. Sample Design for All Other Mailout/Mailback Experimental Panels

The sample design for the remaining three experiments was combined. This includes the deadline messaging and compressed mailing schedule experiment, confidentiality notification experiment, and the AQE 2000-content panel. The design focuses on stratifying based on an area's response propensity. This method allows for estimates by high, medium, or low response areas in the analysis, if necessary.

To achieve this, we use the areas as delineated by the Census 2010 replacement mailing strategy in which: high response areas will not receive a replacement mail form; medium response areas will receive a targeted replacement mailing to nonrespondents by a certain date; and low response areas will receive a blanket replacement mailing in which all housing units will receive it, regardless of their response status. See (Letourneau and Zajac, 2008) for further details on the identification of the replacement mailing housing units.

There will be a total of approximately 5,000 hits in the "High Response Stratum;" 5,000 hits in the "Medium Response Stratum;" and 10,000 hits in the "Low Response Stratum." The sample allocation for these panels utilizes a substantial oversampling because there will be greater benefit in improved response for harder-to-count areas and we want to be certain that we can effectively measure any such improvements for that stratum.

Table 4: Targeted Stratum Sizes for Each Deadline Message, Compressed Schedule, Confidentiality, and 2000 Content Panel

Stratum	Mailout Size per Panel (# of HUs)
High Response Stratum	5,000
Medium Response Stratum	5,000
Low Response Stratum	10,000
Total	20,000

7. Sample Design for the Nonresponse Followup Contact Strategy Experiment

For several decades, the decennial census NRFU questionnaire has instructed enumerators to make up to six contact attempts with households. After Census 2000, the Office of Management and Budget was interested in whether the Census Bureau could feasibly use fewer than six contact attempts to collect data during nonresponse followup and other field operations in an effort to reduce costs.

Researchers used the Master Trace Sample (MTS) Database to study this issue and found that "the rate of successfully conducting interviews decreases with each consecutive contact attempt up to the fifth contact" (Tancreto and Bentley, 2004; Tancreto and Bentley, 2005). At the sixth contact attempt, though, there was a large increase, which was most likely a result of last chance efforts to interview proxy respondents. There were several limitations with the study and without a proper experimental design, it was difficult to estimate what the response and data quality would have been if the number of contacts were fewer than six.

As part of the 2010 CPEX, we will conduct an experiment to compare the results from two experimental enumerator forms. For a random sample of NRFU enumerator forms,

the record of contact section will have space to record a maximum of either five contact attempts or four contact attempts. The standard production form (the control group for purposes of analysis) will have a maximum of six contact attempts. It is important to conduct this long-awaited study in 2010 to understand the effects of changing the number of NRFU contacts in a census environment. For example, the potential impact of advertising and higher anticipated response rates would not be reflected in the results if this study were conducted as part of a mid-decade site test. This study has the potential to provide large cost savings through a reduction in NRFU contacts.

Two solutions were proposed to identify the optimal sample design for yielding meaningful results for the NRFU Contact Strategy Experiment. One option was a case-level sample design whereby experimental forms are randomly placed among 2010 Census enumerator questionnaires during assignment preparation in all 494 Local Census Offices (LCOs). This option would require a single training approach for all LCOs.

The second option was an area-level sample design, in which Crew Leader Districts (CLDs) are randomly selected within 24 randomly selected LCOs. For the CLDs within sample, to deploy a four-contact or five-contact questionnaire, specialized training would be developed and conducted. The CLD approach would require three sets of training materials. Because of the risk of not knowing the actual effect of experimental treatments, there was some concern that there may be a clustering of poorer data quality. In addition, an area-level design would be more difficult to successfully implement and control operationally, though it would allow for better estimates of cost savings.

Ultimately, we chose to employ the national case-level sample design for implementing the experiment. Even though the CLD-level design provides more control in terms of enumerators' actually implementing the experimental contact number limits, the case-level sample design still enables analysis on the impact on data quality and cost savings while minimizing operational logistical issues in the field, as much as possible, and removing the risk associated with perceived poorer data quality. Next, is a summary of the benefits of the case-level design:

- Relatively easy to accommodate the sampling because the LCOs can insert the experimental questionnaires into the boxes of blank forms over a period of time between when the boxes arrive in the office until the time that training begins.

- Easy to implement in the field because this design would still enable the use of a single training approach and a single set of training materials for all enumerators.

- Any negative impact of fewer contacts on data quality will be dispersed across all NRFU cases, thereby eliminating the possibility and/or perception of clusters (e.g., entire geographic or political regions) of data quality loss.

- The experimental treatment will be administered to the enumerators consistently over the duration of the operation; the design remains intact regardless of replacement training of new enumerators or shifting of enumerators across management levels to accommodate language needs or other enumeration challenges.

- Cost savings estimates using this approach would be based on savings realized on a per-case basis due to the fewer attempts.

In order to implement the experiment, a sample of the experimental enumerator forms will be systematically inserted into the boxes of blank questionnaires in all 494 LCOs, prior to assignment preparation activities. The insertion will be random, with no oversampling to any specific geographic areas or other known characteristics. We estimate that this design would result in a sample size of approximately 1.2 million cases, which will be split evenly between the two experimental treatments (600,000 4-contact cases and 600,000 5-contact cases). This sample size will be large enough to ensure that enumerators, on average, will receive approximately one experimental questionnaire in each assignment area assigned to them.

The enumerator training has been revised to include reference to the experimental forms and instructs the enumerators to use the questionnaire to determine the correct maximum number of visits for each particular housing unit. That is, if the questionnaire for a case contains six contact fields, they should make six attempts before taking a proxy response; if the questionnaire contains five contact fields, they should make five attempts; if the questionnaire contains four contact fields, they should make four attempts.

8. Summary of Sample Designs

Table 5 provides a brief summary of each sample design for the experiments.

Table 5: Summary of Sample Designs for Census 2010 Experiments

Design	Experiments Included	Synopsis	Sample Size (# of HUs)
1	Race and Hispanic Origin	Oversampling of Asian/NHOPI, Black, and Hispanic tracts	30,000 per panel (510,000 total)
2	Coverage Overcount	Oversampling of certain tracts expected to be more susceptible to overcounting	30,000 total
3	Deadline Message, Compressed Schedule, Confidentiality, and 2000 Content Panel	Stratification based on high/medium/low response propensities	20,000 per panel (260,000 total)
4	NRFU Contact Strategy	Random systematic sample during assignment preparation	600,000 per questionnaire (1,200,000 total)

References

Letourneau, E., and Zajac, K., (2008), "Specification to Identify Replacement Mailing Housing Units in the 2010 Census," DSSD 2010 Decennial Census Memorandum Series #G-04, Internal U.S. Census Bureau memorandum.

Patterson, E.B. (2006), "Poverty, Income Inequality, and Community Crime Rates," Criminology, Vol. 29:4, p. 755-776.

Tancreto, J., and Bentley, M., (2004), "Enumerator Contact Study 2000", Planning, Research, and Evaluation Division TXE-2010 Memorandum Series: #EO-MTS-F-01, Internal U.S. Census Bureau memorandum.

Tancreto, J., and Bentley, M., (2005), "Determining the Effectiveness of Multiple Nonresponse Followup Contact Attempts on Response and Data Quality," American Statistical Association, Proceedings of the Section on Survey Research Methods, p. 3626-3632.