

How the U.S. Census Bureau’s Redesign of Household Survey Sampling Impacts Field and Clerical Operations

Alexander Winstead, Rosemary Byrne, Christopher Kuwik, Scott Harclerode

U.S. Census Bureau, 4600 Silver Hill Road, Suitland MD 20746

Abstract

Every 10 years the US Census Bureau uses the results from the Decennial Census to update the sample used for many of the household surveys, including CPS, SIPP, NCVS, NHIS, CE and AHS. Sample Redesigns provide an opportunity to improve operational instructions, improve coverage, and implement new technology. This paper discusses the improvements and changes to the existing processes used by the Regional Offices, Headquarters support staff, and the clerical operations supported by the US Census' National Processing Center.

As a result of the 2010 Address Canvassing, a majority of household addresses in the country now have GPS coordinates. Additionally, the last decade has provided near universal access to technological tools such as online mapping software and personal navigation devices. Combining these advances can radically change how the Regional Offices locate difficult addresses. The introduction of operations to mitigate possible undercoverage (particularly for new construction group quarters) will also be discussed.

Key Words: sample redesign, locating sample units, household surveys, field operations, group quarters, survey sample

1. Introduction

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

The U.S. Census Bureau has historically selected sample for the majority of its demographic household surveys once a decade, based on the results of the latest decennial census. This process, called the Demographic Surveys’ Sample Redesign program (or simply “sample redesign”) allows the U.S. Census Bureau to ensure that an accurate, representative sample is used for its surveys. The current sample was selected in 2002 using data from the 2000 Census, and the systems and operations used to select

and field this sample are called the “2000 sample design”. In order to make use of the more current data from the 2010 Census, new sample will be selected and phased into the current sample beginning in February 2014; the systems and operations used to select and field this new sample are commonly referred to as the “2010 sample design”. In addition to ensuring that the sample provides the most accurate possible representation of the current U.S. population and housing units, the 2010 sample design provides unique opportunities to examine the effectiveness of traditional sampling methodology, systems, and operations.

The 2010 sample design is different from both the current 2000 sample design and historical designs in a number of ways. In historical sample designs, the most current census records were used create the primary sampling frame (called the “Unit” frame). This frame was static throughout the ten-year period for which sample was selected. However, beginning with the 2010 sample design the U.S. Census Bureau will no longer select sample once a decade for its major demographic surveys (U.S. Census Bureau, April 2011, p. 1). Instead, the majority of the sample will be selected annually from a continually updated, nationwide file of addresses called the Master Address File (or “MAF”). This new method of sample selection is commonly referred to as “annual sampling”.

As the primary sampling frame will now be subject to updates on an annual basis, there are a number of implications to current sampling systems and operations. Existing systems and procedures, which largely relied on the presence of a static Unit frame, need to be revised to accommodate the change in sampling methodology. Additionally, technological advances over the past decade, particularly with regard to the collection of Global Positioning System (GPS) coordinates during the 2010 Census and the presence of internet mapping solutions, give us more flexibility than ever before in redesigning our operations. This paper will discuss the operational changes to current systems and procedures for the 2010 sample design, as well as the development of new initiatives that will not only complement the 2010 sample design methodology, but also maintain existing coverage, streamline current operations, and improve operational efficiency for headquarters and field staff. These initiatives were greatly influenced by feedback provided by U.S. Census Bureau regional office staff, as well as ideas generated from discussions between multiple divisions and working groups at the U.S Census Bureau headquarters.

2. Operational Challenges for the 2010 Sample Design

One of the main operational effects of the new annual sampling methodology is that the concept of a “basic street address” will become difficult, if not impossible, to maintain throughout the 2010 sample design. The basic street address (or “BSA”) historically refers to the house number and street name of a sample unit, and was used at the start of the current design’s Unit frame to group records of living quarters in multi-units together. This process allows information about all living quarters at a multi-unit (including non-

sample units) to be provided to field representatives in the form of BSA level listing sheets. The concept of a BSA allows field representatives to know which living quarters at the address are and are not in sample, and is useful when they are unsure which unit to interview. In situations where it is clear that the information at the BSA level is inaccurate, field representatives are able to correct the listing of units at that BSA, apply a predetermined sampling rate to their listing sheets, and determine the appropriate unit to interview (a process generally called “time-of-interview listing”). These listing sheets are then stored at the regional offices to provide the updated information to future interviewers.

While the concept of a BSA currently aids field representatives in problem resolution, it relies greatly on the Unit frame remaining static throughout the sample design. Prior to the implementation of the 2010 sample design, the census records that comprise the Unit frame were grouped together via an automated sorting algorithm that assigned a BSA (and a corresponding identifier) to each record. This combining of records was performed once in the sample design, and remained fixed throughout the decade; changes to the BSA could only come via updates received from field representatives. As the records that comprise the Unit frame will now be changing every year (based on the changes occurring on the MAF), defining a BSA in the 2010 sample design presented a major operational difficulty. BSAs would have to be redefined with each annual update of the sampling frame, and complex procedures would need to be introduced to handle “births” (new construction that did not exist in previous iterations of the MAF), “deaths” (formerly valid units that become invalid on the current iteration of the MAF), and conflicts between our BSA sorting algorithm and updates received from field representatives.

The extreme difficulty in accounting for these births, deaths, and changes in an already complex address combining algorithm posed a problem for the 2010 sample design. While rules could be implemented to account for most of these changes in our systems, the primary concern was the effect that these constant redefinitions would have on field operations. New BSA reference materials would need to be printed with each sample selection, effectively rendering any previous time-of-interview listing at the BSA level useless. Coordinating the printing and distribution of listing sheets on an annual basis would significantly increase the cost of sampling operations, while potentially eliminating any information that was gathered by previous time-of-interview listings. Additional concerns could arise if a sample unit was “detached” from an existing BSA due to MAF updates between interviews, an issue that could potentially cause significant problems in the field.

Given the substantial concerns with the use of the BSA in a non-static Unit frame, the U.S. Census Bureau has decided not to combine living quarters into BSAs for the 2010 sample design. Because most field procedures for multi-units relied on field representatives having information at the BSA level, completely new field procedures have been created for the 2010 sample design (discussed in Section 3 below). By

reducing the reliance on BSA level information in solving field problems, the new procedures will allow field representatives to quickly resolve difficult cases in the field and greatly reduce the amount of time-of-interview listing relative to the current design.

3. Reducing Complexity of Field Procedures

Based on feedback received by regional office staff, reducing the complexity and frequency of time-of-interview listing operations is a key goal of the team that is developing new procedures for field representatives. In the current design, field representatives are often asked to fill out a great deal of information concerning the sample address on paper listing sheets, correct any errors, and apply sampling rates to the corrected units. Once this is done, field representatives are instructed to mail this information back to their regional offices, further increasing field representatives' workloads. Feedback across the regional offices indicated that all levels of field staff (including field representatives, supervisors, and coordinators) dislike these procedures and feel that they are overly complex (U.S. Census Bureau, July 2010, p. 8-9). In contrast, the few 2010 sample design procedures that do require listing use a much simpler listing instrument, in which no sampling rates are applied, that can be delivered electronically to U.S. Census Bureau headquarters as part of the field representative's standard case transmission.

While reducing the complexity of time-of-interview listing procedures is a major factor in the development of new procedures, another important issue involves the number of "listing and coverage" problems (i.e., where field representatives are unsure which unit they are expected to interview) that need to be resolved by highly-trained staff at the U.S. Census Bureau headquarters. A number of current procedures for listing and coverage problem situations instruct field representatives to call their regional office; however, little instruction is provided to supervisors at the regional offices on how to resolve these problems. Often the regional offices will forward these problems to headquarters field liaisons, who will then forward them to analysts who must dedicate a significant portion of their time to answering these requests. Because this process relies on multiple levels of communication, the length of time from the beginning of an interview case to its completion can range from one to three business days; this is often a concern as field representatives frequently operate under significant time constraints.

In order to reduce the amount of time spent resolving listing and coverage problems by analysts at headquarters, listing and coverage field procedures are being redesigned to empower field representatives to resolve almost all problem cases directly in the field. Under these new procedures, virtually all situations that field representatives may encounter can be resolved without guidance from their regional office or headquarters. Most of the remaining listing and coverage situations that cannot be resolved by field representatives will be resolved by performing the time-of-interview listing discussed above and assigning the original sample unit a "Type C" interview outcome code. Data from cases with this outcome code are not included in survey estimates, and the cases

will not be sent out again for later interviewing periods. Quickly resolving these listing and coverage problems should greatly reduce the burden associated with problem resolution for field representatives, regional offices, and headquarters analysts.

4. Implementing New Technology and Training Methods

4.1 New Technology

In addition to pursuing innovative ways to streamline existing field procedures and operations, the 2010 sample design offers the U.S. Census Bureau a chance to implement new technologies. GPS coordinates, which were collected during the 2010 Decennial Census Address Canvassing operation, are now available for use on the MAF. Additionally, personal GPS navigation devices were provided to field representatives beginning in October 2010, which allows them to make best use of the GPS coordinates that will be provided as part of their case assignments in the 2010 sample design. Research has established that 94 percent of addresses on the MAF currently have GPS coordinates, which should greatly enhance field representatives' abilities to locate their assigned sample units (U.S. Census Bureau, November 2010, p. 8).

While the use of GPS coordinates in the 2010 sample design will greatly benefit field representatives, they (along with all US. Census Bureau staff) have also been given permission to query for addresses in online geographic search engines such as Google Maps or Mapquest. Previously, field representatives have been required to use an internal geographic mapping program maintained by the U.S. Census Bureau called the ALMI-TOI (Automated Listing and Mapping Instrument – Time-of-Interview). While it can be a very effective tool for locating sample units, the ALMI-TOI has a steep learning curve, is not particularly user-friendly with regard to navigation, and does not provide aerial photography like many commercial geographic mapping sites. The ability for field representatives to quickly identify address locations through an online geographic search engine will be a powerful tool in locating sample units.

Field representatives have always requested that as much information as possible be provided in order to help them locate their sample units. In the 2010 sample design, new technology is being used to provide field representative more information about the administrators and employees of a group quarter (GQ). A GQ is a place where people live or stay, in a group living arrangement, that is owned or managed by an entity or organization providing housing and/or services for the residents (for instance, a college dormitory, group home, or monastery). For sample units selected within a GQ, the professional e-mail addresses of staff and GQ website information will be available to assist field representatives in locating staff to arrange interviews. Providing this information should serve to reduce the amount of time field representatives spend attempting to arrange interviews within GQs, further increasing the efficiency of interviews within this type of living quarters.

As we will be sampling from the MAF on an annual basis, it is critical that information on the MAF be as accurate as possible to assist field representatives in locating sample units. In certain situations where this is not the case (or where it is clear that the MAF contains an invalid unit that is not appropriately marked as such), the time-of-interview listings done by field representatives will be used to update the MAF. Field representatives will only be required to perform a time-of-interview listing when they have exhausted all resources, are unable to determine which unit to interview, and would otherwise be assigning the unit the aforementioned “Type C” outcome code. The listings will be submitted electronically back to headquarters, where staff will review the results of the case and submit the appropriate updates to the MAF.

4.2 New Training Methods

While the newly-revised field procedures for the 2010 sample design are more transparent, efficient, and comprehensive than in any previous sample designs, field representatives must be provided adequate training on these new procedures to make sure that they are implemented correctly. A recent evaluation of the current listing and coverage training program revealed that there were a number of inconsistencies in listing and coverage procedures across surveys, despite the fact that these procedures are the same for each survey (U.S. Census Bureau, February 2010, p 2-3). In order to ensure that each field representative receives the same level of exposure to the material regardless of survey, standardized training for listing and coverage field procedures will be established for the 2010 sample design. Additionally, a refresher training course will be provided every two years to all field representatives to ensure an adequate understanding of the procedures. By providing standardized and frequent training, field representatives should be able to respond to problems in the field quickly and efficiently.

5. Operational Efficiency

5.1 Elimination of Paper Materials

While the 2010 sample design will provide significant improvements to the U.S. Census Bureau’s sampling operations, it will also allow for a number of opportunities to increase the efficiency of our field operations. One major way that this will be done is by eliminating all paper materials relating to listing and coverage for this sample design. Based on the redesigned procedures described above, the current paper materials will no longer be required. The existing paper materials that are still necessary (such as manuals and reference guides) will now be delivered electronically to field representatives’ computers. Items that are currently mailed from the field representatives to the regional offices (such time-of-interview listing information) will now be delivered electronically as well. Not only should this minimize costs relating to the printing, storage, and distribution of paper materials at the headquarters level, it should eliminate mailing costs associated with these items from the regional offices.

5.2 Elimination of Incomplete Address Operation

Another potential source of operational efficiency is the elimination of the Incomplete Address operation that exists in the current sample design. Designed to provide assistance to field representatives who are given “incomplete” (i.e., lacking a house number and/or street name, making them difficult to locate) addresses, the operation forwards incomplete addresses to the U.S. Census Bureau’s National Processing Center (NPC) for clerical review. Analysts at NPC research these addresses and provide any additional information that they discover to the regional offices in order to aid field representatives in locating the sample unit. Incomplete addresses represent approximately 0.6 percent of the current Unit frame sample.

Though the Incomplete Address operation helps to increase coverage in the current sample design, internal research suggests that it will not be necessary for the 2010 sample design. As GPS coordinates will now be provided to field representatives for each assigned sample unit (when available), it is possible that addresses lacking a house number and/or street name may now be easier to locate. Additionally, more precise “geocodes” (i.e., the state, county, tract, and block location of sample units) exist now as a result of the 2010 Decennial Census Address Canvassing operation.

The increased amount of available address information for the 2010 sample design has prompted a revision of the concept of the “incomplete” address. Based on various combinations of variables on the MAF (including geocodes and GPS coordinates), sample units for the demographic surveys will now be classified as “identifiable” (i.e., where sufficient information exists to determine the exact location of the unit) or “imprecise” (i.e., where insufficient information exists to determine the exact location of the unit). Imprecise addresses comprise .005 percent of addresses on the MAF as of July 2010, a much lower number than are currently classified as incomplete (U.S. Census Bureau, March 2011, p. 1). This difference indicates that the number of problems field representatives currently have in locating sample addresses should be markedly reduced by the 2010 sample design procedures.

5.3 Time-of-Interview Problem Resolution

In addition to reducing the number of problem cases that return to headquarters (via the new procedures described above in Section 3), operational efficiency can be achieved by providing procedural guidance directly to field representatives at time-of-interview. In previous sample designs, this guidance was only provided through the listing and coverage manual that was available to field representatives. In the 2010 sample design, however, new and interactive software is being developed for the purpose of automated problem resolution.

The Address Companion Workflow is an instrument designed to integrate seamlessly into the field representatives’ existing laptop case management systems. Available for all cases, the instrument will guide the field representatives through a series of specific questions related to their problem situation. Initial questions will be general, and will

then increase in precision as the instrument determines the correct procedure to display. The specific procedure that the instrument provides will be transmitted back to headquarters; this will allow analysts to compare the instruction provided to the action that field representatives take. By acting as an immediate resource for field problems, the instrument should reduce the time spent by analysts at headquarters and staff at regional offices for problem resolution.

5.4 Extra and Additional Units

Though all of the above initiatives serve to increase the efficiency of field operations, the most significant change involves eliminating the concept of “extra” and “additional” units for the 2010 sample design. Extra and additional units are living quarters discovered in the field that field representatives believe may not have been included in our Unit frame; they are an attempt to reduce undercoverage from the current sample design. Extra units are created when one sample unit actually contains two or more unique living quarters (for example, a household address was sampled, and the residence also contains a separate basement apartment). Additional units are created when field representatives notice units within multi-units that are not on their BSA level listing sheets; this serves to reduce undercoverage and capture potential new growth within multi-units.

While the presence of extra and additional units does slightly reduce coverage error and increase sample quality, it requires significant resources to maintain systems that support them. Unique operations and procedures need to be introduced at the headquarters, regional office, and field representative level to capture these units. Internal estimates have determined that it costs roughly one million dollars per year to support the systems and staff that work on these units. Further, this investment in infrastructure has resulted in very few of these units being discovered; in the current sample design, extra and additional units represent approximately .42 percent of the total sample (U.S. Census Bureau, June 2011, p. 11,12,17). This suggests that, while extra and additional units may serve some small role in reducing coverage error, their costs are far outweighed by the benefits that they provide to sample quality. Additionally, it is possible that many of these units will now be added to our frames via our “new-growth” (i.e., post 2010 Census) processing operations.

6. Coverage of Group Quarters

In both the current and 2010 sample designs, there is a concern about how to best improve coverage of “newly-established” GQs (i.e., those established after the most recent decennial census). In the 2000 design, coverage of these living quarters was improved by listing blocks thought to contain them (such as blocks containing 66 percent or more college dormitories). As our block listing activities will be reduced for the 2010 sample design due to cost concerns, this is no longer a viable solution. Additionally, our primary source of intercensal updates to the MAF, the U.S. Postal Service’s Delivery Sequence File, does not distinguish between newly-established housing units and newly-

established group quarters. In order to maintain sufficient coverage of these newly-established group quarters, new and innovative operations will be introduced.

One new operation for the 2010 sample design involves the modification of questions within the Group Quarters Automated Instrument for Listing (GAIL). The GAIL is currently the main software used by field representatives to list all GQs that have been selected into sample. Currently, field representatives are only asked to obtain information about their assigned cases in the instrument. In contrast to the current procedures, for the 2010 sample design field representatives will be instructed to ask their contact at the assigned GQs about the presence of any newly-established group quarters within the block (or within the college/university for college dormitories). If the contact person indicates that one exists, field representatives will be able to record basic information about it and create a new assignment in the field. They will then be instructed to begin this assignment and collect basic information about the newly-established GQ. This information will be returned to headquarters analysts, who will ensure that the GQ is valid (i.e., not a duplicate of a GQ already on the MAF). Once they have done this, staff will update the MAF with this new record.

Another key operation that will be introduced in the 2010 sample design to reduce undercoverage of newly-established GQs is the new telephone college housing survey. Internal research from the 2000 sample design indicates that 44 percent of newly-established group quarters discovered in our Area frame (the frame in which block listings are conducted) are college dormitories. Based on these data, it is clear that there is a need for operations specifically designed to collect information about newly-constructed college dormitories; the telephone college housing survey will be one of these operations for the 2010 sample design. Administered at headquarters, this telephone-based operation will be conducted in colleges that are not currently selected for interviewing. Once information about a GQ is discovered, it will be verified by headquarters analysts (to ensure that it is not a duplicate of an existing MAF record) and then added to the MAF. The operation will likely be conducted intermittently, as funds and resources allow. The introduction of this operation and the new questions in the GAIL should reduce the undercoverage of newly-established GQs during the 2010 sample design.

7. Conclusion

Sampling from a continually updated frame in the 2010 sample design offers a number of challenges with regard to field operations; however, it also provides an ideal environment to brainstorm innovative solutions. The elimination of the BSA definition necessitates entirely new and streamlined field procedures, which rely more on the interviewer (and less on headquarters staff) to make informed decisions. The introduction of GPS coordinates will greatly aid field representatives in locating sample units, reducing costs associated with traveling to them and searching for the correct sample unit. The Address Companion Workflow will act as a guide for field representatives in resolving problem

cases, reducing the number sent to the regional offices and headquarters and improving operational efficiency. By re-evaluating our existing operations, implementing new technology, and determining areas for increased efficiency, the U.S. Census Bureau will continue to field quality, representative samples for its survey sponsors in the 2010 sample design.

Acknowledgements

The work discussed in this paper is the result of many hours of research and discussion generated from two working groups of the 2010 sample design; WG 5.3 (the Locating Sample Units Team) and WG 3.4 (the Group Quarters Sampling Team).

Special thanks to T. Trang Nguyen and Jeffrey P. Hayes, statistical and content reviewers of this paper.

References

- U.S. Census Bureau. (April 2011). *Decision to Implement Annual Sampling for the 2010 Sample Redesign* (Doc.#2010-3.8-D-4). Suitland, MD: Demographic Statistical Methods Division
- U.S. Census Bureau. (February 2010). *Recommendation on Duration and Format of Time-of-Interview Listing and Coverage Training* (Doc # 2010-5.12-R-1). Suitland, MD: Demographic Statistical Methods Division
- U.S. Census Bureau. (July 2010). *2010 Demographic Surveys Sample Redesign – Summary of the Demographic Statistical Methods Division’s Regional Office Visits* (Doc. #2010-1.2-G-10). Suitland, MD: Demographic Statistical Methods Division
- U.S. Census Bureau. (June 2011). *Recommendation on Handling Multi-Unit Addresses and Extra Units for the 2010 Design* (Doc.#2010-5.3-R-1). Suitland, MD: Demographic Statistical Methods Division
- U.S. Census Bureau. (March 2011). *Summary of Recommendation on Locating Sample Addresses for the 2010 Design* (Doc # 2010-5.3-R-2). Suitland, MD: Demographic Statistical Methods Division
- U.S. Census Bureau. (November 2010). *Summary Report for Evaluating MAF Content Quality After the 2010 Decennial Address Canvassing* (Doc. #2010-4.0-G-16). Suitland, MD: Demographic Statistical Methods Division