

Coverage Improvement for the 2010 Census

Frank Vitrano
U.S. Census Bureau

Race and Ethnicity Advisory Committee

May 2004

Abstract

One of the primary goals of the 2010 Decennial Census is to improve the accuracy of census data, especially the coverage of the population and housing inventory, for all geographic levels and demographic groups. To achieve this goal, the Census Bureau has embarked on a whole series of research and development efforts related to coverage improvement. These efforts were generated by reviewing the results of Census 2000, including Census 2000 experiments and evaluations, and by conducting a series of brainstorming sessions with internal and external experts. In this session, the Data Quality Working Group will review the areas of planned research and development and will be asked to answer the following questions:

1. Do there appear to be obvious areas of research or development that are missing from the Census Bureau's plans?
2. Does the Committee have specific comments on the research and development approaches currently planned?
3. Among the identified areas of research and development, what does the Committee believe are the three or four most important?

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

1. Background

One of the dramatic improvements in Census 2000 was in the area of coverage. The percent net undercount estimate coming out of the Accuracy and Coverage Evaluation, Revision II was – 0.49 percent (or a net overcount of 0.49 percent). Despite this major improvement over the 1990 census (where the net undercount was 1.6%), we are continuing our efforts to improve overall coverage and differential coverage in the census. One of the primary goals of the 2010 Decennial Census is to improve the accuracy of census data, especially the coverage of the population and housing inventory, for all geographic levels and demographic groups. To achieve this goal, the Census Bureau has embarked on a whole series of research and development efforts related to coverage improvement.

Coverage improvement is not simply accomplished by creating new field operations and knocking on doors more times. Coverage problems can occur in just about every aspect of conducting a census, so coverage improvement also permeates the whole census process. From the development of the content of the questionnaire through processing the captured data, the Census Bureau has made coverage improvement a major goal of the 2010 Census.

In this paper, we describe many of the research and development efforts currently underway that we believe will have a positive impact on coverage in the 2010 Census. Many of these efforts came about as a result of a series of brainstorming sessions in which some members of this advisory committee participated.

In the spring of 2002, we began our coverage improvement research and development efforts by looking at several Census 2000 programs. The Accuracy and Coverage Evaluation, the evaluation of the Coverage Edit Followup operation, Demographic Analysis, and the evaluation of the Be Counted Program gave us a fair amount of data that allowed us to characterize persons missed or counted in error. In particular, we identified persons missed as:

- Children, especially young children,
- Adult black males,
- Persons in large households,
- Complex households,
- Recent immigrants, and
- Persons in group quarters.

We identified the following characteristics of persons counted in error:

- College students,
- Persons in group quarters, and
- Persons with multiple residences.

We then used brainstorming to identify why we missed or counted these persons in error. This “root cause” analysis then became the basis for subsequent brainstorming sessions, focused on finding potential solutions to these root causes of coverage error. Headquarters staff, Regional Office staff, and members of this and the other advisory committees participated in this effort. In the fall of 2002, we requested additional input from all members of the advisory committees.

This paper summarizes all of the coverage improvement research and development efforts leading to the 2010 Census. This includes efforts that were generated from the brainstorming sessions described above, Census 2000 evaluation recommendations, literature review, staff experience, and planning group discussion. In some cases, we will also mention ideas that were suggested in the brainstorming sessions but are not being pursued in this decennial cycle.

The rest of this paper is organized by the census operations or activities that are relevant to coverage improvement:

- Communication
- Content
- Language
- Residence Rules and Coverage Probes
- Address List Development and the Types of Enumeration Areas
- Group Quarters Frame Development and Enumeration
- Self-Response Options
- Field Activities
- Processing

In these sections, we do not describe all of the research and development that is taking place in these areas. We focus on those activities that are related specifically to coverage improvement.

2. Communication

The goal of the research and development effort associated with communication is to create an integrated communication plan that will maximize self-response in order to minimize non-response followup workload, and to create an atmosphere of cooperation with enumerators who conduct personal visit interviews to complete the census. The overarching program objectives include increasing self-response, sustaining high levels of awareness, and ensuring relevant and consistent messaging across all programs and operations. Several areas of research and development could have an impact on coverage. These include:

- Understanding how we can work more effectively with communities to identify those most likely to be missed. In doing this, we can better understand the reasons they were missed and convey why it is important for them to be counted.
- Using the Partnership Program to identify and reduce barriers to participation (geography, language, privacy/confidentiality, cultural differences, lack of technology,

disenfranchised from the mainstream, etc.) especially among those groups more likely not to be counted using traditional enumeration methods.

- Making it easier to respond for residents of agricultural worker dorms by working with their employers to encourage response.
- Continuing to improve the Census in Schools Program, including educating children on where they should be counted in the census.
- Developing effective communication of the various options for responding to the census and for contacting someone for assistance.
- Educating local government leaders that the Be Counted Program should not be a primary method of enumeration.

Early efforts in all of these areas have been spent on conducting focus groups of various stakeholders to better understand what worked in Census 2000 and what might help improve our communication efforts. Almost all of these suggestions are focused on reducing the undercount. The last suggestion is intended to help reduce counting persons in error. One idea that was suggested that we are not pursuing is looking at incentives to increase response.

3. Content

The goal of the research and development effort on content is to maximize data quality while achieving consistent responses across modes (self-response paper, internet, telephone, and enumerator instruments) as well as minimize respondent burden. Several areas of research and development could have an impact on coverage. These include:

- Making the short form questionnaire even more user friendly than in Census 2000 so as to reduce respondent burden, potentially increase response rates, and address literacy issues.
- Collecting sex and age data for all persons in the household on the self-response form to reduce the need for large household followup. Internet, telephone, and enumerator instruments on mobile computing devices allow for the collection of all short form data for all persons. This removes the challenges associated with continuation forms. The self-response paper form will still be limited to collection of all short form data for only six persons. However, in the 2004 Census Test, we are testing the inclusion of sex, age, date of birth, and whether the person was related to Person 1 for up to twelve household members. In the future, we will test adding an instruction to encourage telephone or Internet response for large households and begin testing the followup for large households both by phone and by personal interview.

- Researching and developing how we explain residence rules and what data we should collect to ensure coverage and minimize counting people in error. This topic is discussed in more detail in Section 5, Residence Rules.

Each of these suggestions has the potential to reduce both the undercount and counting persons in error. One idea that was suggested that we are not pursuing is the elimination of names to help us with trust issues. Because one of our coverage goals is trying to reduce duplication in the census, capturing names and birthdates have become a critical component of data collection.

4. Language

The goals and objectives for language include identifying the most effective way to provide the nation's diverse communities the opportunity to respond to the 2010 Census in languages other than English. This includes preparing forms in a broad variety of diverse languages, determining the non-English languages necessary for questionnaires, guides, and assistance, developing procedures for conducting interviews with non-English speaking households, and testing alternative fulfillment and targeting strategies to make it easier and faster for respondents to obtain a questionnaire in-language. Several areas of research and development could have an impact on coverage. These include:

- Identifying areas across the nation with a high concentration of non-English speaking households, making forms and promotional materials available in languages that are prevalent. In the 2006 Census Test, we plan to test identifying census tracts with a high concentration of Spanish speakers and sending them a bilingual form. We may also test including language guides in mailing packages.
- Improving in-language questionnaire and promotional materials translation so that the message of the importance of the census can get through to respondents. The Census Bureau has recently developed a set of best practices associated with testing form translation and we will be using these approaches in our efforts.
- Improving the telephone assistance program to allow respondents to request an in-language questionnaire over the phone.
- Continue our efforts to hire enumerators who are proficient in the languages spoken in the areas they work.

These suggestions have the potential to reduce the undercount. One additional suggestion that is still under consideration is whether we should expand the Census in Schools program by making packets available in languages other than English.

5. Residence Rules and Coverage Probes

Research and development related to the residence rules, instructions provided for residence rules, and the introduction of coverage probes on the census form have been a primary focus of coverage improvement research leading up to the 2004 and 2006 Census Tests. In fact, the Census Bureau has asked the National Academy of Sciences to convene a panel to review our plans related to residence rules. The primary goal of this research and development is to improve coverage by making the enumeration task easier and more logical for respondents and to help identify housing units that may need a followup contact to clarify who should be counted in the Census. Several areas of research and development could have an impact on coverage. As was referred to in Section 3, some of these efforts affect the content of the questionnaire. These include:

- Testing providing a reference period when we instruct respondents to count people where they usually live or stay. Currently, we do not provide a reference period to consider when we give them this instruction. In the 2005 Content Test, we plan to test a version of the instructions that gives respondents a specific reference period (the year leading up to Census Day). With this type of instruction, we need to also provide exceptions to account for persons who have actually moved since we are really concerned with where people live or stay as of Census Day. Providing a reference period may improve people's understanding of who should be counted where, but may confuse people who have recently moved. If the result is an overall improvement, this could potentially reduce both the undercount and people counted in error.
- We are not considering changing the rule for counting persons from where they live or stay to where they were on Census Day. This "de facto" approach for counting people would require a change in law that the Census Bureau is not pursuing prior to the 2010 Census. However, we plan to test how collecting data on where people spent Census Night and then counting them where they live most of the time might help us with overall coverage in the census. This is an approach that some other countries use. The idea would be to collect information for who "stayed here on Census Day" and then also collect who usually lives here. For those persons who stayed here on Census Day but do not usually live here, we would try to obtain the address of where they usually live. Through data processing and unduplication, we would then determine if those persons identified as having "stayed here on Census Day" were counted elsewhere or not and then determine where they should be counted. This suggestion could help reduce undercount, however it could also potentially increase the overcounting of persons, depending on our ability to unduplicate.
- Testing changes to the actual residence rules themselves. Throughout this summer, we will be assessing each of the residence rules that are our concrete interpretation of the law requiring us to count people where they live or stay most of the time. We will be looking to see if any rules should be created, deleted, or changed. Making the rules more easily understood or more salient to census enumerators and respondents, as well as having an optimal set of rules, could potentially reduce both the undercount and people counted in error.

- Testing better ways to explain the residence rules. Prior to the 2004 Census Test, we cognitively tested several approaches of explaining the residence rules. The results of that testing lead to the version we are using in the 2004 Census Test. Since then, we have considered more dramatic changes to how we present the residence rules and will be cognitively testing them in the near future, with the hope that some of them will be tested in the 2005 Content Test. One example is to use a principle-based approach. That is, within the instructions, explain that the census must count every person living in the U.S. on Census Day. Then, further explain some specific concepts about our residence rules. A second example provides a step by step accounting of persons living in the household. The first step would be to count persons who live “only in this household.” Then for persons who have more than one place to live, to count them here only if they live here most of the time. As a companion to testing various ways of explaining the rules, we also plan to test various ways of presenting coverage probes for persons who were not counted but maybe should have and for persons who were counted and maybe should not have been. All of these options are intended to reduce both the undercount and the number of persons we count in error.
- Testing collecting sufficient person-level data so that in processing we can define households in order to ensure correct use of the residence rules. This is part of the inspiration behind asking additional coverage probes for members of the household. For example, if a respondent tells us that one of the household members sometimes lives or stays somewhere else to attend college, and we find that person also enumerated at the college, we might create a rule to remove that person from this household since college students are supposed to be enumerated where they are living during college. Also, by asking coverage probes, we can identify households that may have potential undercount or overcount problems. In the 2004 Census Test, we will be testing our ability to clarify individual household rosters by re-contacting households who answered our coverage probes in such a way as to identify potential coverage problems.
- Testing asking households to put all children who are part of the household on the form, whether they are there at the time or not. In the 2004 Census Test, in the instructions for whom to include in the household, our first instruction is to include “Babies and children living here, including foster children. This change in instructions could potentially help us reduce the undercount.
- Testing an additional probe regarding children who might otherwise have been missed, specifically addressing children in joint custody. In the 2004 Census Test, we are testing a coverage question that asks “Are there other people who live or stay at this place part of the time but are not permanent residents, such as live-in employees or children in joint custody?” On the other side of this issue, we are also asking if children in a child custody arrangement might also live and stay somewhere else sometimes. Again, the idea with these questions is to followup with any households who answer this question “yes.” These suggestions could potentially help us to reduce the undercount and counting persons in error. However, this will depend on the extent to which both parents claim their children live with them, even in a followup operation.

- Testing asking if there are other places where someone could be counted. One of the versions of questionnaires we would like to test in the 2005 Content Test asks for the addresses of people who “sometimes live or stay somewhere else.” Also, in the 2004 Census Test, our Coverage Research Followup questionnaire, which will be used for households who answer yes to our coverage probes, goes into great depth about other places where people could be counted. By identifying other places where a person could be counted, we have a better chance of unduplicating them and thus reducing the number of persons we count in error.

6. Address List Development and Types of Enumeration Areas

The primary goals for research and development on address list development and defining the types of enumeration areas are to ensure the most up-to-date accurate Master Address File (MAF) is being used in the 2010 Census and that we identify the best method for enumeration in each block of the country. Several areas of research and development could have an impact on coverage. These include:

- Keeping the MAF up-to-date during the decade by applying ongoing updates by the U.S. Postal Service’s Delivery Sequence File (DSF). The Census Bureau has been using the DSF to update the MAF twice a year since Census 2000. This is a requirement for the ongoing American Community Survey. As we approach the 2010 Census, we will most likely apply more frequent updates that are timed to coincide with stages of conducting the census (Address Canvassing, Mailout, Field enumeration, etc.).
- Obtaining up-to-date information regarding E-911 address conversions. In particular, just because some areas are obtaining house number/street name addresses for emergency services, they may not start using these addresses for mail delivery. In these situations, we will continue to need to use Update/Leave as our enumeration methodology because the U.S. Postal Service would not recognize those house number/street name addresses for mail delivery.
- Improving our ability to eliminate all duplicate addresses before mailing out census forms. In the 2004 Census Test, we are evaluating the use of probabilistic matching to see if it can improve our address unduplication operations. We are also testing several rules for identifying duplicates that have not been considered valid up to this point.
- Revising the Local Update of Census Addresses Program to take full advantage of a realigned TIGER database. A realigned TIGER database will facilitate broader use of a web-based approach to providing updates to the address list and road features and hopefully, increase participation.
- Focusing more on information from the U.S. Postal Service when determining the areas of the country where we will attempt enumeration by using mailout/ mailback. Some work has already been accomplished in this area. The Geography Division has created a new block level code on the MAF that is called an Address Characteristic Code. This

code characterizes each block by its mix of city-style/noncity-style addresses and their representation on the U.S. Postal Service's DSF. The 2005 American Community Survey will be the first program to make use of this code, but we will also be using it in the 2006 Census Test.

- Reviewing the full set of criteria for determining the types of enumeration areas we will use in each block. This effort will begin with the planning of the 2006 Census Test.

All of these suggestions have the potential to reduce both the undercount and the counting of persons in error.

7. Group Quarters Frame Development and Enumeration

The primary goal for the group quarters research and development effort is to review, develop, and test ways to improve the identification, enumeration, and processing of group quarters in the 2010 Census. This includes determining which types of living quarters should be in the group quarters universe versus the housing unit universe, appropriate classification of group quarters types, and how to improve the group quarters frame development. Several areas of research could have an impact on coverage. These include:

- Having a consistent frame development plan for housing units and group quarters. Starting with the 2004 Census Test, we are including all group quarters on the listings that are sent out for the Address Canvassing operation. In that operation, canvassers are asked to identify housing units and "other living quarters." A subsequent visit, in the Group Quarters Validation operation, will be made to all "other living quarters" to determine if they are, in fact, housing units or group quarters and to classify the types of group quarters. This should help reduce the duplication of addresses between the housing unit frame and the group quarters frame.
- Conducting one inclusive Local Update of Census Addresses program so that local officials can help us clean up our address list by making sure we have addresses in the correct geography and by helping us remove duplication among housing units and group quarters. As we continue to develop the LUCA program for the 2010 Census, we are planning to have a single LUCA program so that local governments do not need to respond to separate efforts as they did in Census 2000. LUCA for group quarters, LUCA for city-style housing unit address areas, and LUCA for noncity-style housing unit address areas will all be combined into a single request for local governments, giving them more flexibility in how they wish to participate.
- Conducting an ongoing unduplication operation, as returns are data captured. Starting with the 2006 Census Test, we will be attempting to identify duplicate persons that we believe were counted at a group quarters facility, but also in a housing unit. We will also evaluate the extent to which we need to conduct followup operations to confirm the duplication and to determine the correct location for the person(s) to be counted.

Each of these suggestions has the potential to both reduce the undercount and reduce counting persons in error.

8. Self-Response Options

The primary goal of the self-response options research and development effort is to conduct research, testing and development of self-response options and strategies for marketing these options. These self-response options include mail, Internet, telephone and other potential options. Response strategies refer to the marketing and communications aspect of encouraging the public to respond to the census. The ultimate outcome of this research is the identification of appropriate options and marketing strategies designed to ensure maximum self-response in the 2010 Census. Several areas of research and development could have an impact on coverage. These include:

- Taking advantage of Internet and telephone response options by not limiting the number of persons that can be enumerated in a household. This limit will only exist with the paper version of the census form. Expanding the number of persons that can be counted by internet and telephone is a requirement already built into our plans for the 2010 Census. This suggestion could help us reduce the undercount.
- Improving the telephone questionnaire assistance system to minimize the amount of time it takes to get through to have questions answered and to request paper forms. As we build our telephone questionnaire assistance system, improving the time it takes to get through to the Census Bureau has already been identified as a requirement. This suggestion could help us reduce both undercount and counting persons in error.
- Improving the Questionnaire Assistance Centers, telephone questionnaire assistance, and Internet questionnaire assistance, giving respondents more tools for a successful self-enumeration. This suggestion could help us reduce both undercount and counting persons in error.
- Changing the timing for the Be Counted Program. We plan to test implementing the program after the Nonresponse Followup operation instead of implementing the program before Nonresponse Followup. However, we would need to allow for persons contacting us before Nonresponse Followup if they didn't obtain a questionnaire through mailout or update/leave. The reason for this testing is that a large number of Be Counted forms ended up being enumerations of households that were also enumerated during Nonresponse Followup. At the same time, we do not want to revert back entirely to the 1990 approach. In 1990, if a person contacted the Census Bureau after mailout to say that they did not receive a form, we did not have a mechanism for collecting the response until after the Nonresponse Followup operation. Leaving the respondent hanging for a few months is not considered a valid option in the 2010 Census. This suggestion could help reduce counting persons in error.

9. Field Activities

The primary goal of field activities research and development efforts is to develop and test improved field enumeration processes and infrastructure components, particularly as they may be impacted by using mobile computing devices for data collection activities. Several areas of research and development could have an impact on coverage. These include:

- Reducing duplication by using GPS technology on the mobile computing devices. We are testing the procedures and usability of collecting GPS coordinates in the 2004 Census Test. Starting with the 2006 Census Test, we plan to collect GPS coordinates in advance of the enumeration (during Address Canvassing) so that we can assess whether having GPS coordinates helps reduce confusion in the field (adding units already on the list and knocking on the wrong door during Nonresponse Followup). This suggestion could potentially help us reduce counting persons in error.
- Reducing tabulation geography errors by using GPS technology on the mobile computing devices. An additional advantage to collecting GPS coordinates for housing units is to improve the accuracy of the location of housing units. With a TIGER system positionally realigned, we can be much more precise in geocoding housing units to the correct tabulation blocks. This suggestion could potentially help us reduce both undercount and counting persons in error.
- Allowing address canvassers to identify duplicate addresses by linking census IDs. This is a functionality we have built into the 2004 Census Test where address canvassing was still done on paper. This is also a stated requirement for automated listing operations, which we will begin testing with the 2006 Census Test. This suggestion could potentially reduce counting persons in error.
- Continuing to look for ways to improve training of field staff in the areas of identifying and reducing housing unit duplication and working the correct times of day to maximize contact with household members to reduce proxy responses. This suggestion could potentially help us reduce both undercount and counting persons in error.
- Improving address list development in advance of enumeration. Starting with the 2004 Test, we continue to work towards improved address canvassing procedures. We also believe that the introduction of the Group Quarters Validation operation to clarify “other living quarters” will provide major improvement. This suggestion could potentially help us reduce both undercount and counting persons in error.
- Attempting to reduce proxy information which can cause both undercoverage and overcoverage problems. Starting with the 2004 Census Test, we have added a monitoring feature in the field so that managers will be able to see if too many cases are coming back with incomplete data such as no housing unit count. We hope that this quality assurance improvement will yield fewer cases with missing data and fewer proxy responses.

- Providing a list of mailout/mailback nonresponses sooner than in 2000. With the use of mobile computing devices, we are attempting to get the Nonresponse Followup operation in the field sooner, but we need to see what impact that has on workloads and the size of the required workforce. Perhaps more importantly, we will be able to update enumerator workloads by removing cases that were late mail returns. This approach could help us reduce counting persons in error.
- Verifying and accepting field changes in a more timely fashion. Starting with the 2006 Census Test, for both Address Canvassing and Nonresponse Followup, we will be attempting to build verification operations into the basic operation itself. In the past, after Address Canvassing, we had a separate operation verify cases that were deleted. Also, following Nonresponse Followup, we had a separate operation verify any cases that were deleted or coded as vacant. In the 2006 Census Test, with automation, we believe it may be feasible to recycle these cases during the operation and have them verified while the operation is still in the field. By reducing the time between the original action and the verification action, we reduce the confusion caused by real calendar time passing between attempts. This suggestion could help us reduce both the undercount and counting persons in error.
- Introducing a followup operation to verify cases we believe are duplicates. This is something we are beginning to evaluate in the 2004 Census Test.
- Expanding the collection of data for large households beyond telephone. We plan to start testing both a telephone and personal visit component to the large household followup starting in the 2006 Census Test.

10. Processing

The goal of research and development related to post-data collection processing is to improve and evaluate all components of processing. This includes such things as controlling cases received, editing and imputing for missing and inconsistent data, and conducting the Primary Selection Algorithm when we receive more than one return for the same housing unit. Several topics of research and development in the processing area could have an impact on coverage. These include:

- Introducing a computerized unduplication operation as part of the post-data collection processing operations. In the 2004 Census Test, we are doing this as an evaluation operation. Our intention is to make improvements based on the 2004 Census Test and then make computerized unduplication part of processing starting with the 2006 Census Test.
- Introducing a match between returned questionnaires and administrative records. The idea behind this research would be to test the feasibility of identifying households that may have potentially missed persons by finding some matched persons in administrative

records, but where the administrative records show additional persons who were not recorded on the census return. These households would be added to our coverage followup operation where more probing questions would determine if there are more persons who should have been enumerated. This does not require using the actual names of persons found in administrative records. The approach is intended to identify households that have a high probability of having missed persons.

- Shortening Census ID numbers so that they will be easier to use. We have already introduced a shorter ID number in our testing. This should facilitate situations where respondents need to provide their Census ID such as during Internet response and telephone response.
- Tracking receipt of questionnaires so that we can turn around and reduce the workload during Nonresponse Followup. We are testing this ability starting with the 2004 Census Test.
- Considering proxy data to be less accurate than household member data when we are determining which duplicate to remove. We have not developed formal rules for whom to keep when we identify duplicates. We hope to be informed by the 2004 Census Test evaluations. We will need to develop rules in preparation for the 2006 Census Test and we will certainly consider self-response over proxy response.
- Evaluating alternative statistical procedures in our imputation methodology. We are currently developing improvements to our existing hot deck approach, pursuing a modeling approach, and pursuing the use of administrative records as possible sources for addressing missing data. All of these approaches will initially be assessed using simulations from Census 2000 data. We plan to recommend a final approach in time for development and use in the 2006 Census Test.
- Improving the imputation methodology so as not to bias imputing older children. This is something we are aware of from Census 2000 and will be including this in our research to improve upon.

11. Summary

All of the research and development initiatives described above represent our plans as of today. Each of them will evolve over time based on the success of each of our tests. By the time we design the 2010 Census, we hope that these efforts will yield continued improvements in coverage by both reducing the undercount and reducing the number of persons counted in error.

1. Do there appear to be obvious areas of research and development that are missing from the U.S. Census Bureau's plans?

2. Does the Committee have specific comments on the development and testing approaches currently planned?

3. Among the identified areas of research and development, what does the Committee believe are the three or four most important?