



MEMORANDUM FOR Distribution

From: Cynthia Clark
 Associate Director for Methodology and Standards

Subject: The Within-Block Search and Primary Selection Algorithm
 Operational Evaluation

I am pleased to present the executive summary of one of the evaluation studies for the Census 2000 Dress Rehearsal. The dress rehearsal was conducted in three sites — Columbia, South Carolina; Menominee County, Wisconsin; and Sacramento, California. The evaluation studies cover detailed aspects of eight broad areas related to the census dress rehearsal — census questionnaire, address list, coverage measurement, coverage improvement, promotion activities, procedures addressing multiple options for census reporting, field operations, and technology.

The executive summary for each evaluation study is also available on the Census Bureau Internet site (<http://www.census.gov/census2000> and click on the link to “Evaluation”). Copies of the complete report may be obtained by contacting Carnelle Sligh at (301) 457-3525 or by e-mail at carnelle.e.sligh@ccmail.census.gov. Please note that the complete copy of the following reports will not be publically released: reports regarding procedures addressing multiple options for census reporting and the Evaluation of Housing Unit Coverage on the Master Address File.

The evaluations are distributed broadly to promote the open and thorough review of census processes and procedures. The primary purpose of the dress rehearsal is to simulate portions of the environment we anticipate for Census 2000, so we can identify and correct potential problems in the processes. Thus, the purpose of the evaluation studies is to provide analysis to support time critical review and possible refinements of Census 2000 operations and procedures.

The analysis and recommendations in the evaluation study reports are those of staff working on specific evaluations and, thus, do not represent the official position of the Census Bureau. They represent the results of an evaluation of a component of the census plan. They will be used to analyze and improve processes and procedures for Census 2000. The individual evaluation recommendations have not all yet been reviewed for incorporation in the official plan for Census 2000. These evaluation study reports will be used as input to the decision making process to refine the plans for Census 2000.

The Census Bureau will issue a report that synthesizes the recommendations from all the evaluation studies and provides the Census Bureau review of the dress rehearsal operation. This report will also indicate the Census Bureau’s official position on the utilization of these results in the Census 2000 operation. This report will be available July 30th.

The Within-Block Search and Primary Selection Algorithm Operational Evaluation

April 1999

Miriam D. Rosenthal
Decennial Statistical Studies Division

EXECUTIVE SUMMARY

The plan for Census 2000 includes making it easier for persons to respond by providing multiple response options. This includes allowing persons to respond on Be Counted Forms, the Internet, and the telephone. In the Census 2000 Dress Rehearsal, a blanket replacement questionnaire, used to improve mail response rates, was another response option. Overlaps between late receipts of mail returns and the identification of nonresponding cases that require personal visit during Nonresponse Followup also result in multiple responses. These and other situations can cause the receipt of more than one census return for an address. A special computer program was designed and implemented in the Census 2000 Dress Rehearsal to control the introduction of errors by resolving situations where more than one form was received from an address. The program consisted of two major steps: the Within-Block Search (WBS) and the Primary Selection Algorithm (PSA).

Census forms fall into one of two categories; each processed differently. Forms that are initiated by the Census Bureau, such as mailout forms, Update/Leave mailback forms and enumerator forms, have a pre-assigned census identification number, or census ID. These forms can be easily checked in and compared to other returns. Forms that are initiated by a respondent, such as a Be Counted Form (BCF) or certain Telephone Questionnaire Assistance (TQA) responses, have no pre-assigned census ID and require that the address provided by the respondent be geocoded and matched against other addresses on the Master Address File to obtain a census ID. The timeliness and accuracy of this address geocoding and matching operation has a critical impact on the WBS and the PSA.

Once data capture activities are complete, the WBS and the PSA are run. During the WBS and the PSA, data capture records are reviewed to identify census IDs with more than one eligible return. The WBS and the PSA compare person records on each return. This person matching application uses software developed within the Census Bureau. Persons found to have been included on two separate returns are flagged as such and their record is ignored in subsequent data processing on all but one of the returns. Other rules are used to make determinations of final household composition.

The Within-Block Search

The WBS was implemented for the first time in the Census 2000 Dress Rehearsal. It is a person-based search operation that occurs prior to the PSA and is designed to screen out certain person records on respondent-initiated returns before the PSA is applied. Those records found to match persons who were enumerated on another census return are flagged during this operation and are not eligible for selection during PSA processing. We want to know if the additional searching required by this operation is effective in identifying persons who are enumerated more than once. We also want to know if the rules that were used to identify eligible persons for the WBS were appropriate. The purpose of this evaluation is to quantify how many persons were eligible for the WBS and of the persons searched, how many were found to have been enumerated on another census return.

The WBS had a noticeable effect in update/leave areas of Columbia and a minimal to nonexistent effect elsewhere. About 9 percent of the persons in the WBS workload for update/leave areas of Columbia were matched to persons in the expanded search area. Rates in the other sites were 1 percent or less. Additional analysis is needed to see if address matching complexities for rural-style addresses in Columbia could have caused the higher rates. If address matching limitations are found to exist and can be improved, the additional searching seems unnecessary. The value of the WBS should be reassessed in light of the software and processing complexities inherent in this operation, the resources required to implement it, and the apparent low payoff in the identification of person matches.

The Primary Selection Algorithm

The PSA is used to determine the person records and housing data that will represent each census ID. PSA processing is performed on all eligible records after the WBS has been run. The purpose of this evaluation of the PSA is to quantify how often multiple returns were recognized during dress rehearsal processing and to determine the major reasons for these multiple returns. In addition, we want to better understand how the PSA accounts for these types of returns.

Most of the multiple returns in the Census 2000 Dress Rehearsal occurred when:

- respondents completed both an initial and a replacement questionnaire;
- a household was enumerated during Nonresponse Followup and also on a late mail return;
- a household was enumerated twice during Nonresponse Followup; or
- a household completed either a BCF or a TQA interview and was also enumerated during Nonresponse Followup.

With the exception of update/leave areas in Columbia, all sites recognized more than one return for about 12.5 percent of the census IDs. The rate of census IDs with more than one return in Columbia update/leave areas was about 6 percent. At all sites, fewer than one-half of 1 percent of the IDs had more than two returns.

A review of IDs with two returns identified which response options generated the returns. A blanket replacement mailing in mailout/mailback areas intentionally created multiple contacts. This was the major reason for multiple returns in Sacramento and in mailout/mailback areas of Columbia. Other options inadvertently overlapped with Nonresponse Followup. These included the receipt of late mail returns, BCFs, and TQA interviews. The overlap of late mail returns with Nonresponse Followup is one area that should be studied. Dress rehearsal results indicate that between 3 and 4 percent of all housing units were enumerated on both a mail return and a Nonresponse Followup return. In addition, responses on BCFs and from TQA required address geocoding and matching to obtain a census ID. This was not completed prior to the identification of the workload for Nonresponse Followup. Many of these households were enumerated again during Nonresponse Followup. Finally, there is evidence that specific Nonresponse Followup cases were assigned to more than one enumerator, resulting in two Nonresponse Followup returns being generated for the same ID.

The dress rehearsal data also inform us that most of the returns at two-return IDs were similar or

identical in content. This finding is of value when assessing the potential of multiple returns to introduce coverage error.

Reducing the incidence of multiple returns is desirable. The Census Bureau decided against the use of a replacement mailing; this will reduce the incidence of multiple returns. The dress rehearsal data provide feedback in several areas suggesting that we look at the timing of the identification of the Nonresponse Followup workload. Prematurely identifying this followup universe will result in increased Nonresponse Followup workloads and multiple returns due to receipt of late mail responses. The additional time required to obtain a census ID for responses on BCFs and from TQA make it very hard for this category of responses to be kept out of the Nonresponse Followup workload. The data also suggest that additional controls be considered to avoid multiple Nonresponse Followup visits to the same housing unit.

Additional analysis of two-return IDs is needed in situations where each return involves a unique household. We need to understand the causes for the differences and the best rules to use in selecting those persons that should be included in the final census counts. Other dress rehearsal evaluations will provide critical data for Census 2000 on how to revise the rules that were used for the PSA.