Census 2000 Evaluation L.3.a
November 26, 2002

Analysis of the Primary Selection Algorithm

FINAL REPORT

This evaluation reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Stephanie Baumgardner
Decennial Statistical
Studies Division

USCENSUSBUREAU

Helping You Make Informed Decisions

EXECUTIVE SUMMARY

The full report for this evaluation is not available because it contains proprietary information.

There were several ways in which to respond to the Census 2000 including mailing back a questionnaire, completing the form on the internet, using a Be Counted Form, and being enumerated by field operations such as Nonresponse Followup and Coverage Improvement Followup. While these methods, and others, of collecting population data were implemented with the desire of obtaining a more accurate census count, the various methods also presented the possibility of receiving multiple responses for a single Census ID. The Primary Selection Algorithm was the computer program designed to resolve the receipt of multiple responses from housing units.

Major features of the Census 2000 Primary Selection Algorithm design included performing person matching between returns, constructing Primary Selection Algorithm households, selecting the primary Primary Selection Algorithm household, and selecting additional persons for the census household that are not in the primary Primary Selection Algorithm household.

Less than ten percent of all Census IDs on the Decennial Response File were enumerated by more than one return (continuation and supplemental forms were linked to the appropriate form in a process prior to the Primary Selection Algorithm and are not referred to as returns unless they were not linked). More than 95 percent of these were enumerated by only two returns. About 55 percent of the Census IDs enumerated by two returns are the result of two enumerator returns and about 82 percent of these are the result of returns from Nonresponse Followup and Coverage Improvement Followup. About a third of all Census IDs with two returns consist of one mail and one enumerator return; about 96 percent of these are the result of a mailback return and a return from Nonresponse Followup.

The Primary Selection Algorithm defined some returns as ineligible for the Primary Selection Algorithm process. There is a total of 2,656,951 ineligible returns at all Census IDs. More than 67 percent of these returns are ineligible due to being classified as a deleted housing unit record. Taking these ineligible returns out of the universe, we find that 8,960,245 Census IDs (less than eight percent of Census IDs on the Decennial Response File) have more than one eligible return.

A Primary Selection Algorithm household is a set of associated persons at one Census ID. The set may contain no persons (a vacant Primary Selection Algorithm household), or one or more persons. If two or more returns for the same Census ID have at least one person in common (determined by person matching), then these returns form a single Primary Selection Algorithm household. Of Census IDs with more than one eligible return, person matching between person records on different returns was performed on less than 50 percent of cases, mostly because the combination of returns at many Census IDs involved vacant returns.

Over 73 percent of Census IDs with multiple eligible returns have just one Primary Selection Algorithm household. Census IDs with two or more Primary Selection Algorithm households account for just over two percent of all Decennial Response File Census IDs. The primary

Primary Selection Algorithm household is the Primary Selection Algorithm household that is used in further processing. When more than one Primary Selection Algorithm household exists, the primary Primary Selection Algorithm household is selected by sequentially applying criteria to all of the Primary Selection Algorithm households until only one Primary Selection Algorithm household is selected.

Most Primary Selection Algorithm households at Census IDs with multiple returns consist of one or two returns. Two-return Primary Selection Algorithm households are most often formed by two enumerator returns or one mail return combined with one enumerator return. When two enumerator returns form a Primary Selection Algorithm household, over 91 percent are the result of one return from Nonresponse Followup and one return from Coverage Improvement Followup. This is expected due to the design of the Coverage Improvement Followup operation.

Of the 8,716,359 Census IDs with two eligible returns, over 70 percent have a redundant return (a return containing only person records represented on the basic return of a Primary Selection Algorithm household) and almost 57 percent of these redundant returns are not vacant. At about 26 percent of these Census IDs, two Primary Selection Algorithm households are formed; just under half of these Census IDs (1,089,928 Census IDs) have two non-vacant Primary Selection Algorithm households. Just over four percent of these Census IDs have no redundant return but form one Primary Selection Algorithm household meaning there is at least one person in common among the returns.

Almost 85 percent of all redundant returns are enumerator returns. More than 55 percent of redundant enumerator returns result from Nonresponse Followup and nearly 88 percent of these are occupied, most likely due to the receipt of a late mail return. About 43 percent of redundant enumerator returns result from Coverage Improvement Followup and 97 percent of these are vacant as expected.

Of the 2,349,988 Census IDs with two Primary Selection Algorithm households, more than half have an enumerator return as the basic return of both Primary Selection Algorithm households. Nearly 80 percent of these cases result from one return from the Nonresponse Followup and one return from the Coverage Improvement Followup. This is most likely due to a vacant return from the Nonresponse Followup operation and an occupied return from the Coverage Improvement Followup operation.

Recommendations

Based on the results of this evaluation, it is recommended that future research not be devoted to refining rules to handle uncommon cases of multiple enumerations. Most cases of multiple enumerations in Census 2000 were expected by the design of Census operations. The Primary Selection Algorithm was designed to be robust and handle as many unusual cases as possible but the results show that these unusual cases are very few. In fact, many of the multiple responses were the result of field operations enumerating a Census ID that was already enumerated. The Nonresponse Followup operation created multiple responses at a Census ID when a mail return was received after the cut-off for the Nonresponse Followup. It was found that many of these

Nonresponse Followup returns did not supply any additional information for the Census ID. Similarly, the design of the Coverage Improvement Followup operation created many multiple enumerations. If this operation is implemented in the same manner in the future, a processing step done prior to the Primary Selection Algorithm should remove from further processing a Coverage Improvement Followup return that just confirms the status of a Nonresponse Followup return. The Nonresponse Followup return in this case should be flagged to indicate that its status was confirmed.