

# OUTMOVER TRACING FOR THE CENSUS 2000 DRESS REHEARSAL

David A. Raglin, Susanne L. Bean, United States Bureau of the Census  
David Raglin; Census Bureau; Planning, Research and Evaluation Division; Washington, DC 20233

**Key Words:** Census coverage; Integrated Coverage Measurement (ICM); data collection; nonsampling error; measurement error; person matching

## 1. BACKGROUND

This evaluation provided information to help us determine if outmover tracing needs to be done as part of the Accuracy and Coverage Evaluation (A.C.E.) in Census 2000. Based on the results described here, the decision was made not to conduct outmover tracing in Census 2000. To aid in that determination, this evaluation answered the following questions:

- How many cases did we try to trace and what were the results?
- For households where a traced interview was obtained, how do the proxy and traced data compare?
- What is the person match rate to the census for the proxy data compared to the traced data?
- How are the estimates affected by replacing the outmovers provided by the proxies with the people provided by tracing outmovers?

The Census 2000 Dress Rehearsal was conducted in Columbia, South Carolina and eleven surrounding counties; Menominee County, Wisconsin; and Sacramento, California. Integrated Coverage Measurement (ICM) was the survey that followed the census and was designed to be a quality check survey on the census operations and to adjust census numbers. ICM was conducted independently of the census and collected an independent roster of residents as of census day and the ICM interview day.

## 2. METHODOLOGY

### 2.1 Definition of Movers

In the Census 2000 Dress Rehearsal, census day was April 18, 1998, and ICM data were collected via the computer-assisted personal interview (CAPI) Person

Interview from May to September, 1998. The problem is that people may have moved during that time. The people who have moved out of the housing unit after census day are called "outmovers".

To contrast, the people who have moved into the housing unit after census day are called "inmovers", and people who did not move between census day and the ICM Person Interview date are "nonmovers".

### 2.2 Effect of Movers on Estimation

The Census Bureau used Dual System Estimation (DSE) methodology to get adjusted population estimates based on census and ICM data, using information about both inmovers and outmovers collected during the ICM Person Interview. The DSE calculations are made within poststrata<sup>1</sup>.

Below is the part of the DSE formula that is affected by movers: the term that measures the proportion of ICM people that matched census people:

$$\frac{M}{P} = \frac{M_{NM} + \frac{M_{OM}}{P_{OM}} \times (P_{IM} \times R P_{OM})}{P_{NM} + (P_{IM} \times R P_{OM})}$$

where:

M = weighted estimate of people found in ICM who were matched to census people

P = weighted estimate of people found in ICM

M<sub>NM</sub> = weighted estimate of the number of nonmovers found in ICM who match census persons

M<sub>OM</sub> = weighted estimate of the number of outmovers found in ICM who match census persons

P<sub>OM</sub> = weighted estimate of the number of outmovers found in ICM

---

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

---

<sup>1</sup> The Dress Rehearsal poststratification variables were tenure (owner, renter), race/ethnicity (non-Hispanic White/Other, non-Hispanic Black, non-Hispanic American Indian/Alaska Native, non-Hispanic Native Hawaiian/Pacific Islander, non-Hispanic Asian, and Hispanic), and age/sex (0-17, 18-29 male, 18-29 female, 30-49 male, 30-49 female, 50+ male, and 50+ female).

$P_{IM}$  = weighted estimate of the number of in-movers found in ICM

$RP_{OM}$  = weighted estimate of the proportion of people enumerated as out-movers in ICM that were determined to be residents of the cluster on census day

$P_{NM}$  = weighted estimate of the number of non-movers found in ICM

Why do we use both in-movers and out-movers in the DSE equation? It is relatively easy to determine the number of in-movers from the ICM Person Interview, since we often are talking directly to them. Therefore, the number of movers within the poststrata is estimated based on the number of in-movers.

On the other hand, out-movers are people who lived in the housing unit on census day, so the out-movers are used to estimate the percentage of census and ICM matches.

The DSE methodology helps determine the focus of this evaluation. Out-movers are used in the DSE only to determine the match rate of movers to census people. Census and ICM people are matched using a clerical system, where people do not have to match on every data item to be matched.

### 2.3 Whole Household Out-movers

A whole household out-mover is a household where everyone moved out between census day and the ICM interview day. None of the people who lived there on census day are there when we go to do the ICM interview.

Why do we care specifically about whole household out-movers? Information about out-movers is needed so that they can be matched back to the data provided on the census form. If some of the residents had moved out but some were still there when the ICM Person Interview was conducted, the interviewer collected the information about the out-movers from the non-movers.

However, if everyone who lived there on census day has moved—a whole household out-mover situation—it is not quite as simple.

There are two options to collect information about whole household out-movers. One option is to find a knowledgeable proxy respondent to provide detailed information about the out-movers. Proxy information could be obtained from the in-movers or from neighbors or apartment managers who may have known the

out-movers. This proxy data was used in producing the official estimates for the Census 2000 Dress Rehearsal in Sacramento and Menominee and in measuring the undercount in South Carolina.

Another option is to attempt to trace the whole household out-movers to their new address and interview them about the household on census day. The advantages of this option are clear: theoretically, if the census day resident can be traced, they should know more about the census day household than a proxy would.

On the other hand, tracing out-movers can be difficult, time-consuming, and resource-intensive. In addition, proxies can often provide some of the needed information about the out-movers: name, age, sex, race, ethnicity, and relationship to the first person in the household so the clerical matchers can decide a match occurred. Out-movers are used in the estimation process to obtain the match rate to the census.

This evaluation is designed to determine if tracing whole household out-movers is worth doing by comparing the proxy data (which was used in the official dress rehearsal estimates) and traced data (collected especially for this evaluation).

### 2.4 Operation of Out-mover Tracing

There were two steps to out-mover tracing. First was an operation at the Census Bureau's National Processing Center (NPC) in Jeffersonville, IN. Researchers attempted to obtain the full name, telephone number, and new address of the out-movers (if the proxy had not already provided that information), utilizing all available resources, such as commercial databases, phonedisc and Directory Assistance, to find the necessary information.

Once the researchers got a valid phone number for the movers, they attempted to conduct a CATI interview. The CATI interview asked for the names and characteristics of everyone who lived at the sample address on census day and whether any of the persons had any alternate addresses on census day.

The second step was used if a case had not been found after fourteen days in the CATI unit. It got sent to the appropriate Field Division Regional Office and an interviewer tried to trace the movers to their new address by any means possible, such as knocking on doors and going to the post office. When interviewers successfully traced a mover, they conducted an

interview using a paper-and-pencil interview (PAPI) version of the CATI instrument.

### 3. LIMITATIONS

#### 3.1 Operational Problems

Theoretically in outmover tracing, there never should have been a situation where we found that people never moved. In practice, it can be expected to happen a few times due to measurement error, but not often. However, it happened 15.4 percent of the time in Sacramento and 12.5 percent of the time in South Carolina for completed or resolved traced households. Unfortunately, if in the traced interview the person said they never moved from the Person Interview address, we did not follow up to resolve the discrepancy. An investigation spurred by this finding turned up specific problems that contributed to that and other problems with regard to outmover tracing. This makes the assumption that the Person Interview correctly identified whole household outmovers questionable.

It is important to note that these problems have been corrected for Census 2000.

#### 3.2 General Limitations

The fact the dress rehearsal included only three sites in the country is a limitation in the ability to judge the PAPI operation. In Census 2000, there will be many outmovers in all parts of the country, and during a PAPI tracing operation, the PAPI forms will have to physically be sent around the country as more information is gathered about the outmover. Logistically, that is a problem. In dress rehearsal, we did not send PAPI cases out of the three sites. If a person, say, moved from Sacramento to Detroit but CATI could not find them in Detroit, when the case went back into the field, they only tried to trace them from Sacramento. In Census 2000, the case would be sent to Detroit and traced there.

However, the logistics of moving PAPI cases around the country during the census, which was not an issue during the dress rehearsal, would be very difficult. This issue has been legitimately raised as a reason not to conduct a PAPI operation in a census using outmovers.

In addition, there would be less time to trace outmovers via PAPI in Census 2000 than there was in the Census 2000 Dress Rehearsal. In dress rehearsal, tracing was conducted from June 19 to September 4. In Census

2000, PAPI tracing will take place from late June to late August. Also, because there are relatively few PAPI cases compared to the workload for other A.C.E. operations like Person Interview and Person Followup, and the PAPI cases will be scattered throughout both A.C.E. and non-A.C.E. clusters, it will be difficult for an interviewer to become especially skilled in tracing outmovers

In addition, although these sites were chosen to represent situations found throughout the country, results of this evaluation cannot be generalized to any area beyond the three sites.

### 4. RESULTS

#### 4.1 Amount of Outmover Tracing

How many households were there to trace? How well did we trace them? Table 4.1 answers that question:

**Table 4.1: Results of Outmover Tracing**

	Sacramento	S. Carolina
ICM Housing Units	16,419	17,677
Whole HH Outmov.	918 (6%)	927 (5%)
Traced Households	380 (42%)	482 (52%)

Data for Menominee are not provided since there were only 10 whole household outmovers for the site. The households who reported they never moved are not included in the "Traced Households" line because a large percentage of them should not have gone to outmover tracing, as explained in section 3.1 of the limitations. They made up 7.5 percent of the Sacramento cases and 7.4 percent of the South Carolina cases that went to outmover tracing.

The indication from these results is that about five to six percent of households in Sacramento and South Carolina were whole household outmovers. However, the problems in the limitations section indicate that some of the cases that went to outmover tracing should not have, so the percentage of true whole household outmovers is probably a little less than five percent in those sites.

Among the total tracing workload, 26 percent were traced via the CATI system and 16 percent via PAPI in Sacramento. In South Carolina, 40 percent of the total workload was traced via CATI versus 12 percent by PAPI.

CATI tracing in Sacramento might have been hampered by the large percentage of unlisted phone numbers there.<sup>2</sup> In the debriefing of the CATI interviewers, the interviewers mentioned they could not get addresses for people with unlisted telephone numbers from directory assistance, making it harder to trace those people (Ehni, 1998). However, the number of cases traced in Sacramento indicates that they nonetheless were traceable, just not via a telephone operation.

This indicates that the effectiveness of CATI tracing could vary for different parts of the country, an important finding for planning a census using outmover tracing.

#### 4.2 Comparing Proxy and Traced People

In households we traced, we compared the people found in the ICM proxy interview with the people found in the ICM traced interview. The goal was to determine if we were getting the same people and households in the traced interview as in the proxy one. After all, if we trace and the household list of people contains the same people, tracing is not particularly useful. In 1996, a similar matching operation found that in Chicago, 21.9 percent of the time when we traced a household, we actually got an entirely different list of people.

This matching took into consideration the use of outmover tracing with the DSE methodology in mind. DSE uses outmovers only to compute the match rate between the census and the ICM. Therefore, Census Bureau clerical person matching rules were used in deciding if a proxy and traced person were the same person. A person experienced with the clerical person matching matched the proxy and traced people. Matches were attempted only for data-defined people with a valid name.<sup>3</sup>

Table 4.2 shows the results of the proxy versus traced matching in households where the household was

traced and an interview was obtained that included the listing of people.

**Table 4.2: Comparing Proxy and Traced People**

Person Mentioned In	Sacramento	
	Proxy	Traced
Both proxy & traced ints.	431	431
Only in proxy interview	31	---
Only in traced interview	---	233

Person Mentioned In	South Carolina	
	Proxy	Traced
Both proxy & traced ints	528	528
Only in proxy interview	46	---
Only in traced interview	---	306

If the person was mentioned in the proxy interview, they were almost always found in the traced one too (93 percent in Sacramento (431 / (431 + 31)) and 92 percent in South Carolina). If we assume that the traced interview is better than the proxy interview, since it was supposed to be with a resident, a large percentage of the people that proxies name are really residents.

Table 4.2 also indicates the traced interview found many additional people that the proxy interview did not have. Since the traced interview was supposed to have been done with a resident of the outmover household, this is not a surprise.

In which households were these new traced people found? One theory was that the new people found in the traced interview were in households where we did not collect any proxy people. However, analysis showed that was not true. In Sacramento, 133 of the 233 additional people (57.1 percent) from the traced interview were in households where we got people in the proxy interview who were not in-movers; for South Carolina, it was 144 of 306 (47.1 percent). Therefore, even if we get people in the proxy interview, we get even more outmovers in the traced interview.

The proxy interview seems to yield legitimate people but an incomplete list of the household members (at least according to the traced interview).

<sup>2</sup> Approximately 71.6 percent of the households with telephones in the Sacramento PMSA have unlisted numbers, the largest percentage of any of the biggest 100 metropolitan area in the country (Survey Sampling, Inc, 1999).

<sup>3</sup> To be data-defined, we must have a valid name and one other characteristic. The name must have at least three characters in the first and last name together. The characteristics include relationship, sex, race, Hispanic origin, and age or year of birth or month and day of birth.

### 4.3 Matching to Census People

Remember that under the DSE methodology used, outmovers are included for their person matching rate to the census. The number of movers is estimated by the in-movers. Therefore, the person match rate is a very important indication of the difference in data quality between the proxy and traced data.

Table 4.3 shows the match rate for the data using proxies for outmovers compared to the data using the traced data for outmovers when traced data was available. That means that if a traced interview was conducted and people were collected, the traced interview replaces the proxy interview. If the traced interview indicated the housing unit was vacant or did not exist as a housing unit on census day, the production people were removed.

**Table 4.3: Match Rates to Census People**

ICM Persons	Sacramento	
	Proxy	Traced
In traced households	65.1	65.6
In all of ICM	78.2	78.2

ICM Persons	South Carolina	
	Proxy	Traced
In traced households	79.0	75.5
In all of ICM	78.5	78.5

Look at the first line of the table in Sacramento, the people in households where we collected person data during tracing. Notice that the match rates are almost the same for the proxy people as for the traced people. In the traced interview, we find many more people, but their match rate is similar to the match rate for the people we already had. In South Carolina, we actually had a nominally lower match rate for the people collected in the traced interview.<sup>4</sup>

The last line of the table in each site shows that to one decimal place, the match rate was the same for the dataset that used proxy people versus the traced people. This indicates that even though we find quite a few

<sup>4</sup> The match rates for proxy versus traced people are similar for the poststrata marginal variables, too.

new people in the traced interview, they do not match at a particularly high rate to census people. That is due to the fact that only about two percent of households were whole household outmovers that we were able to trace as well as the similarity of the match rates between proxy and traced people.

### 4.4 DSEs Using Proxy and Traced People

This is really the most important section of this paper. We can say the match codes do not seem to differ between the proxy and traced people, but if there are significant differences in the DSEs, we can say the differences in the match rates were actually significantly large.

Table 4.4 shows the DSEs calculated from the production data (excluding people in groups quarters and in the service-based enumeration), the DSEs calculated from the data using the traced outmovers in place of the proxy outmovers in households we were able to trace, the differences, and whether or not those differences are significantly different than zero.

**Table 4.4: DSEs Using Proxy and Traced People<sup>5</sup>**

	Sacramento	S. Carolina
Est. w/ proxy people	395,005	693,724
Est w/ traced people	395,025	693,579
Difference (st error)	20 (279)	-145 (522)
Significant ( $\alpha = .10$ )	No	No

We also did this comparison for each poststrata marginal variable using the Dunn method of controlling for multiple comparisons.<sup>6</sup>

There were not significant differences in the DSEs calculated using proxy and traced outmover people in either site for any of the poststrata marginal variables, with  $\alpha = .10$ . In fact, the p-values are not close to being significant most of the time. Outmover tracing

<sup>5</sup> The numbers here are the estimates for the whole site minus people collected from group quarters and the service-based enumeration operation.

<sup>6</sup> In the Dunn method, the alpha level was divided by the number of comparisons to be made: one for the total, two for tenure, seven for race/ethnicity, and six for age/sex, to come up with the significance level used in the tests. See Toothaker (1993).

provided a nominal increase of 20 people in Sacramento (0.005 percent of the production estimate) and a nominal loss of 145 people in South Carolina (0.021 percent of the production estimate).

There is no reason to believe that with the current DSE methodology, the lack of tracing caused any significant difference in the production estimates provided in the Census 2000 Dress Rehearsal.

## 5. ANALYSIS AND CONCLUSIONS

The big question is whether or not the outmover tracing operation should be included as part of the A.C.E. in Census 2000. Of course, this evaluation is based on data from two sites -- Sacramento and the parts of South Carolina that were included in the Census 2000 Dress Rehearsal. It could be possible that outmover tracing could have a significant effect on estimates in parts of the country that differ from these two sites.

It is understandable why outmover tracing might not have a significant effect on the estimates. Remember that outmovers are used to produce an estimate of the match rate between the census and the ICM. Look back to Table 4.2, the table that showed the results of matching the proxy and traced people to each other. We found that almost all of the time if a person was mentioned by a proxy and they knew enough about the person that we could consider them data-defined, the person was later mentioned in the traced interview.

To contrast, the people mentioned by the proxy but not the traced respondent might logically be assumed to not be residents, and in fact they did not match well to the census: 70 percent did not match in Sacramento and 43 percent did not match in South Carolina. However, there are so few of them that their effect on the match rate is relatively small.

Therefore, for the people in the traced interview to have a much higher match rate than the proxy interview people, the new people mentioned in the traced interview but not the proxy one would have to match to the census at a higher rate than the people mentioned in both interviews. There is no reason to think that to be true, and in fact it isn't.

From the people in the traced interview, 32 percent of the people also mentioned in the proxy interview did not match anyone in the census in Sacramento, while 39 percent of the people mentioned in the traced interview but not mentioned in the proxy interview did not match census people. In South Carolina, the

figures were 20 percent versus 30 percent.

While it would be nice to have all of the people in the household (as tracing would help us do), we really don't care about the number, just the match rate. The proxies seem to be giving us good enough data for matching purposes.

We therefore recommended that outmover tracing not be conducted as part of the Census 2000 Accuracy and Coverage Evaluation. Based on this analysis, this recommendation was accepted, and there will be no tracing of outmovers in Census 2000.

---

## References

Ehni, Sandy (1998), "Debriefing Summary for the Census 2000 Dress Rehearsal of the Integrated Coverage Measurement Outmover Tracing Operation", Internal Census Bureau memorandum.

Survey Sampling, Inc (1999), "Unlisted Rates", [http://www.ssisamples.com/ssi.x2o\\$ssi\\_gen.product?id=71](http://www.ssisamples.com/ssi.x2o$ssi_gen.product?id=71).

Toothaker, Larry (1993), "Multiple Comparison Procedures" (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-089), Sage University Press, Newbury Park, CA.