

## THE FORWARD TRACE STUDY

Mary H. Mulry and Aref Dajani, U.S. Bureau of the Census\*  
Washington, DC 20233

**KEY WORDS:** reverse record check, census coverage error, census evaluation

### 1. Introduction

The Bureau of the Census conducted the Forward Trace Study as part of its research program on census coverage evaluation. The Forward Trace Study tested methods for the tracing necessary in an evaluation of the decennial census with a reverse record check. In a reverse record check, a sample is drawn from the population sometime before the census, traced forward in time to the census, and matched to the census. A sample was drawn from the 1980 census and supplemented by a sample of those missed by the census, a sample of immigrants and a sample of births. The people in the four samples were traced over the years 1980 to 1985.

The reverse record check has been used effectively in the evaluations of the Canadian censuses since 1961. The underlying assumption in a reverse record check is that a person's chance of being interviewed changes with time. For example, some people are very mobile during their late teens and early twenties but are less mobile as children and older adults. They are more likely to be interviewed during these more stable periods of their life. Since the highly mobile twenty-year-olds in 1990 were less mobile ten-year-olds in 1980, 1980 U. S. census is expected to have better coverage of this age group than the 1990 census. Therefore, a sample constructed from these four frames and traced forward to the next census will have better coverage of the population than the census itself.

The Forward Trace Study did not consider the feasibility of the reverse record check for evaluating the census because there was no census in 1985 available for matching. However, the study focused on an essential ingredient, the tracing methods. The three tracing methods considered were:

- Treatment 1, periodic tracing with intermediate personal contact,
- Treatment 2, periodic tracing with one initial contact,
- Treatment 3, periodic tracing without personal contact.

The treatments were compared on the basis of the final trace rates and the cost. Overall, the results indicated that a reverse record check does not appear workable on

the scale that would be required to produce precise estimates of census coverage error. The estimates of the tracing rates obtained at the end of the project in 1985 were not high enough to recommend a reverse record check as a method of evaluating the census in 1990. Although there was some evidence that the tracing rates might be higher with intensive tracing techniques, tracing people over time was difficult to manage and control. Since even experienced interviewers needed a month or more to locate the harder cases, a sufficient number of qualified personnel probably would not be available at the time of the census.

This report contains the major findings of the study. Section 2 describes the various tracing procedures. Section 3 discusses the cost. Section 4 contains the estimates of the tracing rates. Section 5 summarizes the conclusions that can be drawn from the study.

### 2. Description of Operations

The sample for the Forward Trace Study was derived from four sources: (1) the 1980 Census, (2) persons missed in the 1980 census, (3) people immigrating since the 1980 census and, (4) children born since the 1980 census. The sample persons were traced, in time, from location to location. Record was kept of the new locations of movers.

The sources were referred to as the C, M, I and B samples. The C or census sample was a sample of the 1980 Census Post Enumeration Program (PEP) E-Sample. The E Sample was a sample of households enumerated in the 1980 Census.

The M or missing sample was a portion of the PEP population or P sample. The P sample was the April and August 1980 Current Population Survey (CPS) cases that were matched to census questionnaires for the coverage evaluation of the 1980 census. A sample of the persons determined missed in the census by that 1980 Census CPS Match was the Forward Trace M sample.

The I or immigrant sample was a sample of immigrants to the United States from April 1, 1980 to March 31, 1984. The sample was selected for each year by the Immigration and Naturalization Service through seventeen of their control offices.

The B or birth sample was sample of births in the United States from April 1 to December 31, 1980. The B Sample was selected in conjunction with the National

states, the District of Columbia and City of New York vital statistics offices. Seven states were not able to participate in the study because of regulations controlling the disclosure of confidential birth record information. These states were Idaho, Louisiana, Maryland, Michigan, New Jersey, Oklahoma and Rhode Island.

The C, M, I, and B Samples were divided into three treatments. The treatments are defined by the one tracing procedure particular to itself and different from the others. Treatment 1 cases were defined to be periodically traced, including periodic personal contact tracing. Treatment 2 was periodically traced but the initial contact was the only personal contact. Treatment 3 was defined by only periodic tracings, but with no personal contact. Because of limited fiscal year '85 funds, a subsample of the original C, M, I, and B Samples was selected. The final sample sizes were the following:

| Sample               | Treatment | Households | # of Persons |
|----------------------|-----------|------------|--------------|
| C                    | 1         | 1,373      | 4,119        |
|                      | 2         | 1,341      | 3,977        |
|                      | 3         | 1,310      | 3,854        |
| C<br>Sample<br>Total |           | 4,024      | 11,950       |
| M                    | 1         | 489        | 967          |
|                      | 2         | 504        | 1,071        |
|                      | 3         | 437        | 958          |
| M<br>Sample<br>Total |           | 1,430      | 2,996        |
| I                    | 2         | 1,464      | 1,464        |
|                      | 3         | 885        | 885          |
| I<br>Sample<br>Total |           | 2,349      | 2,349        |
| B<br>Sample<br>Total |           | 945        | 945          |
| Sample<br>Totals     |           | 8,748      | 18,240       |

Seven types of tracing techniques were used during the study. The ones performed on a case depended on its treatment and sample. Because of timing restrictions, the same techniques were not included in each sample.

The tracing techniques were as follows:

- 1) An initial interview of cases assigned to Treatments 1 and 2 was conducted.
- 2) The Post Office was asked to confirm the address we had for a sample person or provide a forwarding address.
- 3) Letters explaining that the sample person had been selected for a research study were mailed to each person in the sample. An address correction was requested on the

envelope, and the clerks recorded the new address when these were returned in addition to mailing a letter to the new address.

- 4) An interim interview of cases assigned to Treatment 1 was conducted.
- 5) An administrative records match (ARM) of Forward Trace records to Internal Revenue Service records was performed to obtain new addresses. The ARM was done for sample people in Treatments 1 and 2 when we had their social security numbers.
- 6) The closeout consisted of mailing out questionnaires to all sample people. If the sample person did not return the questionnaire, an interviewer was sent to the last address we had on file for the person.
- 7) The Super Trace was an intensive field trace of a sample of the people not found in the closeout.

### 3. Cost

One way of comparing the three methods of tracing, the three treatments, was by the cost. Cost records were not kept in a way that permitted determining the cost for individual treatments. However, the cost were kept in such a way that allows for a relative comparison on a case basis.

The estimates of the cost for each treatment and each sample are contained in the tables below. Each table contains an estimate of the cost to select the sample for the Forward Trace Study. The cost of the initial interview was not included for the C and M samples. The initial interview for these cases was the interview for the 1980 PEP.

The assessment of the cost of the C and M sample cases did not include the cost of the selection of the 1980 PEP. The Census Bureau's 1980 PEP cost slightly under \$17 million which included both the E and P sample operations. The sample was selected clerically from boxes of census questionnaires using a list of questionnaire numbers. However, the \$17 million did not include the selection of the P sample because the P sample was the sample for the Current Population Survey.

Since the Super Trace samples were subsamples of those not traced in the closeout, the cost for a person in the Super Trace samples was determined by adding the estimated cost of the Forward Trace and the estimated cost of the Super Trace itself. The Super Trace samples contained 760 of 2890 people not traced in the closeout. The

amount of \$44.50 was the estimated cost of field work and clerical processing for each of the 760 people selected for the Super Trace.

**Table 3.1 Cost Per Person for the C and M Samples by Treatment**

| Tracing<br>Technique | Treatment   |             |             |
|----------------------|-------------|-------------|-------------|
|                      | 1           | 2           | 3           |
| Interview            | -           | -           | -           |
| FTS Selection        | 3.00        | 3.00        | 3.00        |
| PO Check             | 5.00        | 5.00        | 5.00        |
| Letter               | 3.50        | 3.50        | 3.50        |
| ARM                  | 5.10        | 5.10        |             |
| Interview            | 19.00       |             |             |
| PO Check             | 5.00        | 5.00        | 5.00        |
| Closeout             | <u>5.00</u> | <u>5.00</u> | <u>5.00</u> |
| Total                | 54.10       | 35.10       | 30.00       |

**Table 3.2 Cost Per Person for C and M Super Trace Samples by Treatment**

| Tracing<br>Technique | Treatment    |              |              |
|----------------------|--------------|--------------|--------------|
|                      | 1            | 2            | 3            |
| Forward Trace        | 54.10        | 35.10        | 30.00        |
| Super Trace          | <u>44.50</u> | <u>44.50</u> | <u>44.50</u> |
| Total                | 98.60        | 80.70        | 74.50        |

**Table 3.3 Cost Per Person for the I Sample by Treatment**

| Treatment     | Years 1, 2 and 3 |              | Year 4       |
|---------------|------------------|--------------|--------------|
|               | 2                | 3            | 3            |
| FTS Selection | 13.00            | 13.00        | 13.00        |
| Interview     | 5.00             | 5.00         |              |
| PO Check      | 19.00            |              |              |
| Letter        | 3.50             | 3.50         | 3.50         |
| Closeout      | <u>13.50</u>     | <u>13.50</u> | <u>13.50</u> |
| Total         | 54.00            | 35.00        | 30.00        |

**Table 3.4 Cost Per Person for I Sample Super Trace Sample by Treatment**

| Treatment     | Years 1, 2 and 3 |              | Year 4       |
|---------------|------------------|--------------|--------------|
|               | 2                | 3            | 3            |
| Forward Trace | 54.00            | 35.00        | 30.00        |
| Super Trace   | <u>44.50</u>     | <u>44.50</u> | <u>44.50</u> |
| Total         | 98.50            | 79.50        | 74.50        |

**Table 3.5 Cost Per Person for the B Sample**

| Tracing<br>Technique | Treatment<br>3 |
|----------------------|----------------|
| FTS Selection        | 11.20          |
| Letter               | 3.50           |
| Closeout             | <u>14.00</u>   |
| Total                | 28.70          |

**Table 3.6 Costs Per Person for B Sample Super Trace Sample**

| Tracing<br>Technique | Treatment<br>3 |
|----------------------|----------------|
| Forward Trace        | 28.70          |
| Super Trace          | <u>44.50</u>   |
| Total                | 73.20          |

Tables 3.1 through 3.6 allow a relative comparison of Treatments 1, 2, and 3 by cost per case. Treatment 1 cost \$54.10 per case without the Super Trace and \$98.60 with it. Treatment 2 cost \$35.10 per case in the C and M samples and \$54.00 per case in the I sample without the Super Trace. With the Super Trace, the cost rose to \$80.50 per case and \$98.50 per case in the C and M samples and in the I sample, respectively. Treatment 3 cost approximately \$30.00 per case in all four samples without the Super Trace. With the Super Trace, the cost rose to approximately \$74.00 per case.

#### 4. Trace Rates

Estimates of the trace rate were made for each sample as a whole and for each treatment within the samples. Estimates of the trace rate also were made for demographic subgroups within each sample. A sample person was considered to have been traced if the person was found during the final closeout operation either by returning the mail questionnaire or by field interview or the person was identified as deceased or emigrated during one of the tracing operations. When the Super Trace results were included, sample persons were considered traced if they were traced during the final closeout or the Super Trace.

##### 4.1 Trace Rates by Treatment

Tables 4.1 through 4.6 contain the trace rates with and without the Super Trace results for the C, M, I and B samples by race and by treatment when appropriate. The I sample did not contain race information, and the B sample people were not divided into treatments. In the regular tracing operation, the estimates of the trace rates were 91.1 percent for the C sample, 83.4 percent for the M sample, 70.6 percent for the I sample, and 73.7 percent for the B sample. The estimated standard errors for these trace rates were 0.6 percent, 1.6 percent, 1.1 percent, and 1.4 percent, respectively.

**Table 4.1 C-Sample Percentage Estimates of Trace Rates and Standard Errors by Race and Treatment**

| Treatment | Race          |               |               |               |
|-----------|---------------|---------------|---------------|---------------|
|           | White         | Black         | Other         | Combined      |
| 1         | 95.4<br>(1.0) | 86.8<br>(1.6) | 90.3<br>(1.4) | 94.1<br>(0.8) |
| 2         | 90.8<br>(1.5) | 81.7<br>(1.9) | 83.3<br>(1.9) | 89.3<br>(1.3) |
| 3         | 90.7<br>(1.2) | 81.7<br>(1.7) | 80.9<br>(3.2) | 89.7<br>(1.1) |
| Combined  | 92.3<br>(0.7) | 83.7<br>(1.0) | 85.3<br>(1.2) | 91.1<br>(0.6) |

**Table 4.3 M-Sample Percentage Estimates of Trace Rates and Standard Errors by Race and Treatment**

| Treatment | Race          |               |                | Combined      |
|-----------|---------------|---------------|----------------|---------------|
|           | White         | Black         | Other          |               |
| 1         | 87.8<br>(2.8) | 82.5<br>(3.3) | 89.6<br>(7.2)  | 86.8<br>(2.2) |
| 2         | 84.0<br>(3.7) | 77.4<br>(3.7) | 77.7<br>(9.4)  | 82.4<br>(2.9) |
| 3         | 83.7<br>(3.3) | 80.8<br>(3.5) | 48.8<br>(14.2) | 81.2<br>(2.8) |
| Combined  | 85.0<br>(2.0) | 80.1<br>(2.0) | 73.9<br>(6.9)  | 83.4<br>(1.6) |

**Table 4.5 I-Sample Percentage Estimates of Trace Rates and Standard Errors by Treatment All Four Years Combined**

| Treatment | Trace Rates   |
|-----------|---------------|
| 2         | 72.9<br>(0.9) |
| 3         | 66.1<br>(1.9) |
| Combined  | 70.6<br>(1.1) |

**Table 4.6 B Sample Percentage Estimates of Trace Rates and Standard Errors by Race**

| Race     | Trace Rates   |
|----------|---------------|
| White    | 76.5<br>(1.6) |
| Black    | 62.6<br>(3.2) |
| Other    | 61.6<br>(6.9) |
| Combined | 73.7<br>(1.4) |

In general, tracing with personal contact was more successful than tracing without personal contact. Treatment 1 with intermediate personal contact in the C and M samples was more effective than Treatment 2 with only initial personal contact and Treatment 3 without personal contact. In the I sample, which had only Treatments 2 and 3, Treatment 2 was the more successful. The B sample was not divided into treatments.

In the C sample, the difference in the trace rates between Treatments 1 and 2 was 4.8 percent with an estimated standard error of 1.5 percent. The difference between Treatments 1 and 3 was 4.4 percent with an estimated standard error of 1.3 percent. These differences were significant at the 90 percent level of confidence. The difference between Treatments 2 and 3 was 0.4 percent with an estimated standard error of 1.6 percent. This difference was not significant at the 90 percent level of confidence.

The difference in the trace rates between Treatments 1 and 2 in the M sample was 4.4 percent with an estimated standard error of

3.6 percent. The difference between Treatments 1 and 3 was 5.6 percent with an estimated standard error of 3.6 percent. For Treatments 2 and 3, the difference was 1.2 percent with an estimated standard error of 4.0 percent. These differences were not significant at the 90 percent level of confidence. The addition of the results of the Super Trace bore the same pattern.

In the I sample, the difference between Treatments 2 and 3 was 6.8 percent with a standard error of 2.1 percent. The difference was significant at the 90 percent level of confidence.

The difference between Treatments 2 and 3 observed in the I sample did not appear in the C and M samples. A possible explanation for why the difference did not appear in the C and M samples, but did in the I sample was that the I sample persons were interviewed by Forward Trace interviewers while the C and M samples were interviewed by interviewers for the Post Enumeration Program (PEP). Interviewers for the PEP had a different goal which was the address of the person on census day. A contributing factor for small differences in the trace rates for Treatments 2 and 3 in the C and M samples was that PEP interviewers interviewed people in both treatments, but information obtained in this interview was not used during the Forward Trace Study for Treatment 3 cases. However, the contact might have an effect on the results.

Table 4.7 contains estimates of the trace rates for the Super Trace. All the sample people in the Super Trace were not traced during the final closeout operation. The estimates of the trace rates for the Super Trace alone indicated that it was a successful operation. With the addition of the Super Trace results, the estimated overall trace rate increased to 95.6 percent for the C sample, 92.5 percent for the M sample, 78.5 percent for the I sample, and 86.0 percent for the B sample. These tracing rates are the result of a longer tracing period and more intensive effort by the interviewers.

**Table 4.7 Percentage Estimates of Trace Rates for the Super Trace by Sample**

| Sample | Trace Rates |
|--------|-------------|
| C      | 45.6        |
| M      | 55.0        |
| I      | 26.8        |
| B      | 46.9        |

#### **4.2 Trace Rates by Race**

As shown in Tables 4.1, 4.3, and 4.6, the Forward Trace Study was more successful in tracing whites than blacks and other races in the C, M and B samples. The race of the I sample people was not available.

The difference in the trace rates for whites and blacks was significant at the 90 percent level of confidence in the C sample, but not in the M and B samples. In the C sample, the difference in the trace rates for whites and blacks was 8.6 percent with an estimated standard error of 1.2 percent. In the M sample, the difference in the trace rates for whites and blacks is 11.4 percent with an estimated standard error of 7 percent. In the B sample, the difference was 13.9 percent with an estimated standard error of 12.8 percent.

The difference in the trace rates for whites and others was significant at the 90 percent level of confidence in the C and B sample, but not in the M sample. In the C sample, the difference in the trace rates for whites and others was 7.0 percent with an estimated standard error of 1.4 percent. In the M sample, the same difference was 11.1 percent with an estimated standard error of 7.1 percent. In the B sample the difference was 14.9 percent with an estimated standard error of 7.1 percent.

Treatment 1 with intermediate personal contact was the most successful across the racial categories in the C and M samples, which had all tracing treatments and all racial categories. The pattern between the tracing treatments was the same in the racial categories as when they were combined for the overall trace rates for the samples. Although the differences between Treatments 2 and 3 for the racial categories were small, Treatment 3 had slightly higher trace rates than Treatment 2 in several instances. In the M sample, Treatment 2 had a higher trace rate for blacks. However, no strong conclusions can be drawn. Again, the small differences are probably attributable to little variation in the application of the treatments themselves.

### **4.3 Comparisons With Other Studies**

Trace rates in Table 4.8 for reverse record checks in the United States in 1960 and in Canada in 1976 and 1981 provide a basis of comparison for the Forward Trace Study. (Gosselin, 1980).

The samples for the reverse record check in the U. S. in 1960 were drawn from four sources at the time of the 1960 census. The sources were the 1950 Census, the 1950 Post Enumeration Survey, state birth records and alien registration records. Each sample person was traced retrospectively from the address available from the source of selection. This method differed from the Forward Trace Study which traced sample people forward in time. The trace rate of 70.6 percent for the I sample in the Forward Trace Study cannot be compared legitimately with

the 1960 results. The trace rate for immigrants in 1960 was 100 percent because at that time legal aliens were required to register their address with the government every year, and the sample was drawn from this list. The estimated trace rates for the C and M samples differ only by 0.1 percent and 0.3 percent, respectively, from the trace rates from the 1960 results. When the Super Trace results are included, the trace rates from the Forward Trace Study are higher.

**Table 4.8 Percentage Trace Rates in Reverse Record Checks in the U.S. and Canada**

|            | <u>U.S.</u> | <u>Canada</u> |             |
|------------|-------------|---------------|-------------|
|            | <u>1960</u> | <u>1976</u>   | <u>1981</u> |
| Census     | 91.0        | 96.0          | 97.1        |
| Missed     | 83.2        | 91.4          | 96.1        |
| Births     | 85.6        | 92.4          | 92.3        |
| Immigrants | 100.0       | 89.4          | 96.1        |
| Combined   | 87.8        | 95.2          | 96.6        |

The Canadians achieved higher trace rates than the 1960 U. S. study or the Forward Trace Study, with the explicable exception of the immigrant sample in 1960. However, the Canadians had an advantage over the Forward Trace Study in that they were able to start with a sample that had addresses and names for most all the sample people. The percentage of sample persons in each sample that did not have sufficient information to attempt tracing is shown in Table 4.9. For five percent of the sample persons selected for the I sample and over six percent of those selected for the B sample, no attempts at tracing were made. With the I sample the problem was usually that the address was not complete. The main problem with the B sample was that confidentiality laws in some states, particularly California, permitted parents to request that all information not be released. With these cases, not even a name was received. The sample persons in the C sample without sufficient information for tracing were often enumerations without any names.

**Table 4.9 Percentage of Sample Persons Without Sufficient Information to Attempt Tracing**

| <u>Sample</u> | <u>Percentage</u> |
|---------------|-------------------|
| C             | 0.5               |
| M             | 0.1               |
| I             | 6.5               |
| B             | 5.2               |

## **5. Summary**

The estimates of the trace rates from the Forward Trace Study did not merit a recommendation that a reverse record check be used to measure census coverage in 1990. More importantly at the time, a reverse record check did not appear workable on a scale that would be required to produce very precise estimates of census coverage error such as would be suitable for census

adjustment. An interesting observation was those people missed by the census in the M sample were also more difficult to trace than the people included in the census in the C sample.

The estimates of the trace rates were comparable to those achieved in the Census Bureau's 1960 reverse record check. The estimates of the tracing rates in the 1960 reverse record check are shown in Table 4.8. However, the Super Trace techniques would be necessary to ensure the highest trace rates possible. Since very precise estimates would require high trace rates, the Super Trace techniques should be used for all personal interviews.

The cost estimates illustrated that tracing was expensive. Treatment 1 which had the highest trace rates also was most expensive because of the periodic personal contact.

A significant finding was that the Forward Trace Study was also a challenge to manage and control. The records for the original sample filled 22 file cabinets. File folders were constantly being pulled and refiled for recording results of the tracing techniques. Controlling records with computers would have been easier had they been available at the time.

Tracing was not a quick operation that could have been accomplished with an inexperienced staff. The final closeout option took four months to complete the clerical and field work. The Super Trace required an additional two months. All the interviewers were experienced, dedicated, and motivated. The letter in the Appendix illustrates how motivated and interested the interviewers were. These time and staff requirements would be difficult to meet in a census environment.

The primary methodological advantage a reverse record check has over a post enumeration survey is that there is no response correlation between the independent record frames and the census being checked. This type of correlation introduces bias in the estimates of census coverage error.

However, which of the two methods achieves better coverage of the population, particularly for minority subgroups, is not obvious. The effective coverage rate can be viewed as the product of the frame coverage rate and the response rate.

$$\begin{array}{l} \text{EFFECTIVE} \\ \text{COVERAGE} \\ \text{RATE} \end{array} = \begin{array}{l} \text{FRAME} \\ \text{COVERAGE} \\ \text{RATE} \end{array} \times \begin{array}{l} \text{RESPONSE} \\ \text{RATE} \end{array}$$

The response rate for a post-enumeration survey will probably be higher than the estimates of trace rates observed for persons with Treatment 1 in the Forward Trace Study. However, how well the sampling frames for

the two methods cover the population and their respective strengths is not clear, especially for subgroups that are hard to enumerate in the census. The estimates of the trace rates for minorities are lower than those for whites. The post-enumeration survey method has also shown weak response rates in these groups.

The Forward Trace Study shows that a reverse record check is more complicated to execute than a post enumeration survey. When very precise estimates of census coverage are required, no evidence was produced to indicate that a reverse record check would be more effective.

### Acknowledgements

The authors wish to acknowledge Howard Hogan for his large contribution in the early stages of the project. David Altwater, Dan Childers, and Laurie Moyer also contributed to the project.

### Appendix

The Chief of Field Division, Stanley Matchett, sent a letter to the interviewers inviting their comments. The portion of the response from Ida Reiter of Smyrna, Georgia, that pertains to the Forward Trace Study follows:

Hello!

9/9/85

I've just finished the Supertrace survey for Linda Leier at 2900 and I'm feeling pretty proud of myself on being able to find 4 of 4 sample persons. I'm a good sleuth but I'd like to call whoever's attention to the fact that a little more time to follow up on it really helped (plus a bit of overtime)--even the passage of time helped locate one person!

At this end what you need is a person like me, experienced interviewer WHO KNOWS THE TERRITORY, knows how to LISTEN and ask questions with a happy enough disposition to disarm people so they'll tell you stuff. Too many people don't want the old girlfriend, the landlord, the ex, the cops, the bank, the WORLD to know where they are that will simply try to disappear and they do a good job of it, too! At "your" end give me please a slightly warmer trail to follow (5 years is too long!) CORRECT information, if you can, please. (Well, I worked on Decennial too, I know how it was.) I hope this feedback is of some value, I may be dead in 1990.

### Reference

Gosselin, J.F. (1980). "Reverse Record Check: Tracing People in Canada, " Survey Methodology, 6, 84-113.

\*This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.