

USING ADMINISTRATIVE RECORD PERSONS IN THE 1996 COMMUNITY CENSUS

Elizabeth M. Sweet, U.S. Bureau of the Census
U.S. Bureau of the Census, Washington, DC 20233

Key Words: Administrative Records, Coverage Measurement

I. Introduction

Information from administrative records are a possible source of people and housing units missed in the Census. This paper documents the evaluation of the administrative records person level data used in the 1996 Community Census. A coverage measurement personal interview was conducted in all housing units in the three test sites (seven tracts in Chicago, IL, the Fort Hall Reservation, ID, and the Pueblo of Acoma, NM) of the 1996 Community Census. This personal interview was conducted via a Computer Assisted Personal Interview (CAPI), and will be thus forth referred to as the Integrated Coverage Measurement ICM CAPI. Prior to the ICM CAPI interview, administrative records people and people enumerated during the initial phase of the census were unduplicated and loaded into the CAPI instrument for the appropriate housing unit. During the ICM CAPI, it was possible for an interviewer to ask questions about people thought to be living in that housing unit, which could include asking about administrative records people.

Documented in this paper are: (1) the amount of duplication between the administrative records and people enumerated during the initial phase of the census; (2) the quality of the administrative records people by source; and (3) available information on the sensitivity surrounding use of administrative records person information in the 1996 Community Census.

II. Methodology

Although an entire paper could be devoted to the methodology of obtaining and then using administrative records in the field, this section briefly highlights and summarizes the steps accomplished prior to asking about an administrative record person during the 1996 Community Census ICM CAPI interview. For the 1996 Community Census, several steps were taken in order to incorporate administrative records into a census count. Steps 1 through 7 are similar to those used in the 1995 Census Test Phase II analysis (Wurdeman and Pistiner, 1996). The remaining steps are unique to the 1996 Community Census.

1. Federal, state, and local level files were acquired. See Administrative Records Research Memorandum #4 for detailed discussion of the file acquisition.
- For all sites, files included the 1994 tax year

Individual Master Tax Return IRS file, the 1996 Medicare Enrollment Database from Health Care Financing Administration (HCFA), and the Registration File from Selective Service.

- In addition to the files above, for Chicago only, a 1996 Medicaid file for Illinois was obtained. A Student Information System file from Chicago Public Schools was obtained. The Tenant Rental Assistance Certification System (TRACS) and Multifamily Tenant Characteristics Systems (MTCS) were obtained from Housing and Urban Development.

- For the Ft. Hall and Acoma sites, we also obtained an Indian Health Service file from the Department of Health and Human Service, Public Health Service. Unfortunately the files did not include all ZIP Code areas of the sites. Leggieri, 1996 discusses these limitations.

- The Social Security Administrations NUMIDENT file was only used to validate the social security number and to gain missing demographic characteristics for people found on other files.

2. For each individual file, name and address information were standardized. Census geography was added to each file where applicable.

3. Files were merged to form one file of person records. Records were dropped if the address was located outside of the site boundary or if the record lacked a name.

4. Based on a predefined order of selection, duplicate person records within the same housing unit ID were merged into one person record. The file sources creating that person record were retained.

5. The unduplicated administrative person records with the same Social Security number were merged together. The file sources were retained for each merged person record.

6. Person records missing an address and those identified as deceased were dropped.

7. All person records with the same address (housing unit) ID were then combined to form a household.

8. This file was delivered to the Decennial Management Division where it was matched via Winkler's matching algorithm (1994, 1995) by address to the Census housing unit file of addresses for the three sites in the 1996 Community Census Test. Approximately 56 percent (3,969) of the housing unit addresses on the administrative record list in Chicago, 54 percent (405) in Ft. Hall, and 8 percent (132) in Acoma matched to the addresses on the census files. Acoma had a particularly low match rate since most of

the administrative record addresses were non-city style addresses and therefore could not be geocoded to a block. (The matching was accomplished within a block.)

9. For the matched addresses, administrative record people were unduplicated via Winkler's matching algorithm (1994, 1995) with the people enumerated during one of the initial phases of the census (e.g., mail-return and enumerator-return). This process created an unduplicated list of people to load into the ICM CAPI instrument. During this unduplication, if an administrative record person was a duplicate of an initial phase person, the initial phase person was loaded into the instrument. All other unique (i.e., unduplicate) persons, both initial phase and administrative record people were loaded. A flag identifies each type of person: initial phase only, administrative person record only, or an initial phase/administrative record person.

10. With the people loaded into the instrument, the CAPI interview was conducted. The interviewer initially obtained an independent roster of people living at the housing unit on census day, their demographic characteristics, and census day residence information. Following that, in most cases, a match screen appeared listing the independent roster and the list of people loaded into the instrument for that housing unit. People were matched between the two lists. Additional questions were asked of any unmatched person loaded into the instrument. This could include asking about an administrative record person.

11. Interviewers could not determine the source of the people loaded into the ICM CAPI instrument, thus it was impossible to determine who came from an administrative list.

12. To insure that the addition of administrative records did not corrupt the sample in terms of refusals or computer glitches, a split panel was assigned in the 1996 Community Census so person records from administrative record sources would display in only half the sample (Killion, 1997). The two panels are referred to as the display panel and the do not display panel.

13. Two estimation procedures were tested in the 1996 Community Census. One of the procedures, called CensusPlus, used administrative record people in the estimation process; the other procedure called Duel System Estimation did not use them during estimation. For the CensusPlus estimation procedure a "resolved roster" was created. All residents from the independent roster and all non-matched residents from the list of persons loaded into the instrument (which could include administrative record people) were added to the resolved roster.

III. Results

This section is broken into four questions which attempt to quantify both the effect of adding administrative records to the ICM CAPI interview and their effect on the CensusPlus estimate in the 1996 ICM test sites. Standard error estimates, typically presented within parenthesis, were produced using the jackknife estimation method available in VPLX statistical software (Fay, 1990). T-tests were also performed using VPLX. SAS® was used for chi-square tests of independence.

How much person duplication was there between the administrative records and the census enumeration?

As explained in Step 9 of Section 2, administrative records were unduplicated with initial phase enumerations prior to loading them into the instrument. Although later tables document the results of this step, Table 1 shows how many housing units were affected by the addition of administrative records during this step. Less than half (45.4%) of the housing units in Chicago contained at least one person who was on an administrative list acquired. The percentages in Ft. Hall and Acoma were much smaller because the administrative record files did not include people from all ZIP Code areas of those sites (Leggieri, 1996).

Table 1: Count of occupied housing units with and without administrative record people

Site	Housing units with at least one administrative record or initial phase/ administrative record person loaded		Housing units with no administrative record or initial phase/ administrative record person loaded		Total N
	N	%	N	%	
Chicago	3,222	45.4	3,878	54.6	7,100
Ft. Hall	360	21.9	1,282	78.1	1,642
Acoma	113	18.4	500	81.6	613
Total	3,695	39.5	5,660	60.5	9,355

Table 2 provides the distribution of people loaded into the ICM CAPI instrument. The majority of people were only an initial phase enumeration. There was duplication between administrative records and initial phase enumerations. We find that over all three sites, approximately 44 percent of the administrative record people were a duplicate of an initial phase person. This percentage was relatively consistent across all three sites, even though slightly different administrative record sources were used in the three sites.

Table 2: Distribution of people loaded into the ICM CAPI instrument from occupied housing units with complete or partial interviews

Type of Person	Chicago	Ft. Hall	Acoma
Initial Phase only	64.6%	70.1%	76.0%
Administrative Record only	20.4%	15.4%	11.2%
Initial Phase/ Administrative Record	15.0%	14.5%	12.9%
Total Number of People Loaded into the ICM CAPI	n=18,134	5,940	2,505

What is the quality of the people on administrative records?

If one was to assume that all the administrative record people who were not a duplicate of an initial phase enumeration were missed by the census (i.e., were residents of the housing unit), Table 2 identifies an additional 18 percent that could be added to our estimate. We acknowledge that we cannot automatically add these people without determining their true residency status. Respondent's answers to specific residency questions in the ICM CAPI instrument allowed us to determine whether the administrative records people were residents at the housing unit according to census rules of inclusion.

Because of the split panel design in the 1996 Community Census (see Step 12 of Section 2), residency information was collected for approximately half of the administrative record only people. For our purposes, we'll call these administrative record people "displayable" people. During the ICM CAPI interview, all "displayable" people were considered to be one of three basic categories: a resident, nonresident or unresolved. Table 3 demonstrates that approximately 16 percent (standard error=1.17) of the displayable administrative records people were residents in Chicago. Approximately 15 percent (standard error=2.32) of the displayable administrative records people were residents in Ft. Hall. And approximately 24 percent (standard error=8.37) of the displayable administrative records people were residents in Acoma.

Table 3: Distribution of residency status for "displayable" administrative record people from occupied housing units with complete or partial ICM CAPI interviews

Site	Resident/All displayable administrative record people (not a duplicate of an initial phase enumeration)	
Chicago	305/1,900	16.1%
Ft. Hall	72/489	14.7%
Acoma	33/137	24.1%
Total	410/2,526	16.2%

For the CensusPlus estimation procedure a Resolved Roster was created. Basically this roster included all residents of the independent roster and all unique residents who were loaded into the instrument. We say unique since people who matched between the independent roster and the initial phase roster were only counted once. So, in addition to determining the residency status of our administrative records people, we also need to determine how many matched to a person identified during the independent roster process. The nonmatched administrative record people who were residents are added to the Resolved Roster. Table 4 demonstrates that on average 15 percent of the "Residents" from Table 3 were added to the Resolved Roster. Using figures from Table 3 and Table 4, we find that approximately 2.5 percent (16.1 percent*15.7 percent) (standard error=0.56) of the administrative records from Chicago would be added to the split panel estimate; 2.0 percent (standard error=0.78) of the administrative records from Ft. Hall would be added to the split panel estimate and 2.2 percent (standard error=1.22) of the administrative records from Acoma would be added to the split panel estimate.

Table 4: Distribution of match status for all displayable administrative record people who are residents from occupied housing units with complete or partial ICM CAPI interviews

Site	Residents who did not match independent roster during ICM CAPI interview/Total (See 'Resident' in Table 3)	
Chicago	48/305	15.7%
Ft. Hall	10/72	13.9%
Acoma	3/33	9.1%
Total	61/410	14.9%

Can we predict the quality of administrative records people by their source?

Predicting the residency status of the administrative records by their source could help minimize the amount of noise added to the list of persons loaded into the instrument. Tables 5a & 5b provides the frequency distribution of the source(s) by the resident/nonresident status for the three sites.

From Table 5a, we find that roughly 29 percent of the people coming from the School file, 22 percent coming from the Medicaid file, and approximately 42 percent of the IRS & Medicaid and 50 percent of the IRS & Medicare people were residents. There also were many sources or combinations of sources where no residents were identified, such as in the HUD files. For this reason the chi-square test of independence of source by residency status is not appropriate (Cochran, 1954).

Combining categories we find that roughly 94 percent of the administrative records in Chicago, 81.4 percent of the administrative records in Ft. Hall, and 75.9 percent of the administrative records in Acoma came from only one administrative source file. Steps 3-5 of Section 2 note that source files were unduplicated so that only one person record was created. We tested the hypothesis that administrative person records coming from two or more sources are residents at a higher rate than those person records from only one source. In Chicago and Ft. Hall we could not reject this hypothesis using an alpha=0.10 level. In Chicago 15.2 percent (standard error=1.16) of the administrative record people coming from only one source were considered residents as compared to 28.8 percent (standard error=5.21) of the administrative record people who came from more than one source. The t-test statistic of 2.607 was significant, with 1,898 degrees of freedom. In Ft. Hall 13.1 percent (standard error=2.39) of the administrative record people coming from only one source were considered residents as compared to 22.0 percent (standard error=5.08) of the administrative record people who came from more than one source. This is significantly different with 487 degrees of freedom a t-test statistic of 1.715. In Acoma, the difference was not significant with 24.5 percent (standard error=9.7) of the administrative record people coming from only one source were considered residents and 22.6 percent (standard error=11.1) of the administrative record people who came from more than one source.

Table 5a: Frequency of Resident/Total "Displayable" Administrative Records People by Source(s) for Chicago

Source(s)	Resident/ Total	% of Total (standard error)
IRS	158/1165	14% (1.2)
Medicare	23/218	11% (2.2)
Medicaid	36/167	21% (5.1)
School	47/160	29% (4.5)
Selective Service	7/72	10% (4.0)
School & Medicaid	6/39	15% (8.6)
Sel.Service & IRS	11/34	32% (7.9)
IRS & Medicare	12/24	50% (12.6)
IRS & Medicaid	5/12	42% (17.9)
Medicare & Medicaid	0/3	0%
School & IRS	0/2	0%
TRACS & IRS	0/2	0%
Sel.Serv & IRS & Medicaid	0/1	0%
TRACS & Sel.Serv & IRS	0/1	0%
Total	305/1900	16%
Standard Error		1.2

Table 5b: Frequency of Resident/Total "Displayable" Administrative Records People by Source(s) for Ft. Hall and Acoma

Source(s)	Ft. Hall	%	Acoma	%
IRS	4/29	14%	1/7	14%
Medicare	1/18	6%	1/4	25%
Selective Service	0/1	0%	0/2	0%
Sel.Service & IRS	0/1	0%	0/1	0%
IRS & Medicare	0/0	0%	0/0	0%
Indian Health (IH)	47/350	13%	24/93	26%
IH & IRS	15/64	23%	6/23	26%
Sel.Serv & IH	1/15	7%	1/3	33%
Medicare & IH	3/5	60%	0/4	0%
Sel.Serv&IH&IRS	1/4	25%	0/0	0%
Medicare/IH/IRS	0/2	0%	0/0	0%
Total	72/489	15%	33/137	24%
Standard Error		2.3		8.4

Is there sensitivity when we ask about administrative records people?

A fair amount of skepticism existed concerning the Census Bureau's use of administrative lists and notification of this list sharing as documented by six focus groups composed of individuals who participated in the 1995 Test Census in the Oakland, California area (Bates, 1995). Bryant and Dunn (1995) and Harris and Westin (1994) also document growing public perceptions of privacy invasion by government. The confidentiality issue here might be people's fear of having their private data being mistakenly disclosed to another household. Or perhaps some are simply frustrated by their lack of control over their personal information.

The Census Bureau must weigh any benefits of using administrative records against the potential for mistrust in the system, if the records are used. But first, we must determine how much sensitivity exists. Do respondents become sensitive when interviewers asked about people who the respondent never mentioned during the interview? Previous sections have documented our stringent procedures for assigning administrative record people to housing units based on an address match. These procedures are so stringent that we lose a lot of administrative record people in the process. Even so, a fairly large portion (approximately 40 percent) of the displayable administrative record people were not known by the respondent. We cannot examine the sensitivity those people would have felt if they'd known that we asked a person at another housing unit if they knew them. But we can examine the reaction of the respondents who we interviewed.

We examined four separate items to determine that there was little or no sensitivity associated with asking questions about administrative record people that potentially live at the housing unit in the 1996 Community Census. This is not unlike previous

findings in the South Tucson Arizona Administrative Records Test that concluded, "...asking questions about people that had once been connected in some way to the household seemed reasonable to the respondents interviewed."

A. Response Rates

An ICM interview could be complete, partially complete or a refusal based on the amount of information captured during the interview. We found no significant difference when comparing the ICM completion rate and refusal rates between the two administrative record panels (display administrative records vs. do not display administrative records). This implies that respondents were not ending the interview or refusing to complete the interview when they were asked about administrative records persons

B. Refusal to answer questions about administrative records people

During the interview, if the administrative records person did not match to an ICM person, the interviewer asked the respondent whether or not they knew the person, and if so, did that person live at the housing unit and so forth. The respondent had the ability to refuse to answer that first question, "We show someone named [FILL NAME] living at this address. Do you know this person?" We found that the refusal rate to this question was not significantly higher for administrative records persons.

C. Interviewer Debriefing Question

Interviewer debriefings were conducted at all three sites. Question 12 in the written interviewer debriefing form asked, "Were there any instances when respondents were upset that you already had names in the computer and they hadn't given you those names earlier in the interview? If so, how often did it occur." All five interviewers in Denver, six of the eight Chicago interviewers, and six of the eight Seattle interviewers responded "no" to that question. The four interviewers that said "yes" claimed that it only happened a few times, in one case with a divorce. In comparison to the confusion with the race and ethnicity question and the negative reaction to the "last night" question, any negative reaction to the administrative records people paled in comparison (Hovland, 1997).

D. Help Screens

During the 1996 Community Census, respondents were not automatically notified that our records could possibly include persons from administrative records. Instead, a help screen was available on the screen with the question, "We show someone named [FILL NAME] living at this address. Do you know this

person?" An administrative record person name could have been filled in the question. The help screen associated with this screen provided responses to respondent's inquiries about people we had listed at the address.

Q: Where did this name come from?

A: The name could have come from 2 different sources. It could have come from the original census form completed by someone at this address OR from other government records now being used to supplement census information.

Q: Why are you getting names from other government records?

In an effort to count more people, we are testing the use of lists from some other government agencies to help supplement the number of people we count. These lists look promising as a way to count people we might otherwise miss.

Q: I've never heard of this person. Why are you asking about him/her?

A: Some names from the records are outdated and no longer associated with your address --that's why you may not know the person being asked about.

Q: What government records are you using?

- Medicare records
- Tax return records
- Housing and Urban Development records (HUD)

Q: What about confidentiality?

A: All of the information collected by the Census Bureau, whether it comes from you or from other government agencies, is kept completely confidential under Federal law (Title 13, U.S. Code). This means that no other government agency or court system can see your answers. We will combine your information with others to prepare statistical totals.

We were able to track if the interviewer accessed this "help screen." Roughly 2 percent of all the interviews accessed any of the available help screens. There were approximately 29 help screens available through out the instrument. Eight of these 29 help screens had sub-help screens. Although this is not a definitive test of whether there is sensitivity with asking about administrative records people, and an interviewer could have memorized some answers to typical questions on this screen, the fact this help screen was accessed in only five interviews suggests that sensitivity surrounding administrative records people rarely occurs. Of the five interviews which accessed the help screen, two of the interviews were in the display panel and the other 3 were in the do not display panel. And of the two interviews in the display panel, only one had actual administrative record people loaded. This suggests that respondents, perhaps those who didn't complete the census form initially, might wonder where we "got our names" regardless of whether they were from an administrative list or not. Thus, we conclude that respondents do not question administrative records people any more than any other person loaded.

IV. Conclusions

Much effort by the Administrative Records Research Staff of the Census Bureau went into creating the administrative records database to use in the 1996 Community Census. As documented the match rate between the administrative records database and the Census housing unit file of addresses was relatively low. Any possible person match rate or within household coverage improvement gain is greatly affected by the ability to match administrative record addresses to the census address mail lists.

In the 1996 Community Census for the addresses that matched, approximately 56 percent of the administrative records people matched to an enumeration during the initial phase of the census. Of the remaining "displayable" administrative record people, 16 percent were potential residents, and about 15 percent of those could be considered potential adds to the Resolved Roster for CensusPlus estimation. This finding is in keeping with previous studies of the quality of administrative records (Hill and Leslie, 1996) and confirms the need for a personal interview with a household member to confirm the residency status of these administrative record people. The addition of administrative records people to the ICM CAPI interview does not appear to adversely affect the interview process.

The Census Bureau's future use of administrative records must weigh the cost-benefit of these procedures. We did find some within household coverage gain due to using administrative record persons, but there are still difficulties in the address match process between administrative list and our census mail lists. We found basically that we could only match approximately 1/2 of the addresses. And then within these matched addresses we found a large amount of noise in the quality of people on the administrative record database, so much so that we would have to do a reinterview of some sort to verify the residency status of nonmatched administrative record people.

V. References

- Bates, Nancy. July 12, 1995. Focus Group Discussion on Data Sharing and Notification of Data Sharing. Center for Survey Methods Research. Statistical Research Division.
- Bryant, B. and W. Dunn. 1995. "The Census and Privacy." *American Demographics*, pp. 48-54, May, 1995.
- Cochran, W.G. 1954. Some methods for strengthening the common chi-square test. *Biometrics* 10: 417-451.
- Documentation on 1996 Administrative Record File Acquisition. Administrative Records Research Memorandum #4.
- Farber, James. October 17, 1996. "Specifications for Panel Assignments For the 1996 Community Census." DSSD 1996 Community Census Memorandum Series No. A-2.
- Fay, R. January 17, 1990. "Draft Documentation of VPLX." Unpublished document. U.S. Bureau of the Census.
- Harris, L. and A. Westin. 1994. "The Equifax Report on Consumers in the Information Age." A national opinion survey conducted by Equifax, Inc., by Louis Harris and Associates.
- Hovland, Claire. 1997. 1996 Person Interview Operation, Enumerator and Crew Leader Debriefing — Fort Hall, Idaho. Integrated Coverage Measurement Field Memorandum Series #IP-FT-16.
- Leggieri, Charlene. December 9, 1996. "ZIP Code Difficulties in the Acquisition of Administrative Records for the 1996 Community Census."
- Hill, Joan and Theresa Leslie. "1995 Coverage Study Results." DSSD 1995 Census Test Memorandum Series #38.
- Singh, Rajendra P. July 10, 1996. "Sample design and estimation team minutes." Internal e-mail message from Rajendra Singh to team members.
- Whitford, David. May 31, 1996. "Specification for Data Input for Computer Assisted Personal Interview for Integrated Coverage Measurement." 1996 Community Census Memorandum Series IP-D-16.
- Whitford, David. March 20, 1997. "Specification for Selecting the ICM Interview for CensusPlus and Dual System Estimation Processing." 1996 Census Test Memorandum Series IP-D-36.
- Winkler, W. E. (1994), "Advanced Methods for Record Linkage, American Statistical Association, *Proceedings of the Section on Survey Research Methods*, pp. 467-472.
- Winkler, W. E. (1995), "Matching and Record Linkage," in B. G. Cox (ed.) *Business Survey Methods*, New York: J. Wiley, 355-384.
- Wurdeman, Kent, Arona Pistiner. March 26, 1997. 1995 Administrative Records Evaluation- Phase II. 1995 Census Test Results Memorandum No. 54.