

Updating the U.S. Census Bureau's Master Address File via Administrative Records

Christine Gibson Tomaszewski¹

US Census Bureau

4600 Silver Hill Road, Washington, DC 20323

Abstract

The U.S. Census Bureau's Master Address File (MAF) is a national address list used for the decennial census. As the first major field operation of the 2010 Census, the 2010 Address Canvassing Operation was important to provide an accurate address inventory for the census enumeration operations, which reduces census costs. It was one of the largest dependent address listing operations in the world, costing \$845 million (Holland, 2012), and procedures need to be developed to avoid similarly expensive field operations for future address list updates and maintenance. Administrative Records (AR) offer a heretofore untapped but highly cost effective approach to updating and maintaining the MAF between the Decennial Censuses. Using AR to update the MAF may reduce the scope of or need for future field operations to update the MAF. For this research, the Census Bureau matched 225 million addresses from the MAF to over 150 million addresses in the 2007 and 2008 Statistical AR System.

Key Words: Administrative Records, StARS, Administrative Records Concordance, Decennial Census, Census 2010, 2020 Census, Census Coverage, Address List Maintenance, Statistical Workload Reduction, Cost Avoidance, Address Listing, Address Matching, Targeting, Targeted Address Canvassing, TAC

1. Introduction

In the Census Bureau's Introduction to the 2020 Census (Bureau of the Census, 2009), one of the potential design innovations was "supplementing the address frame with Federal or commercial administrative records." In the Census Bureau's Strategic Plan for the 2020 Census (Bureau of the Census, 2010), the strategies for "Goal 3: A Efficient 2020 Census" included the "Use administrative records where feasible and publicly supported."

The research presented here used the results of the 2010 Address Canvassing (AC) operation and supplemental field operations to evaluate the use of Administrative Records (AR) from the Statistical Administrative Records System (StARS) for updating the Master Address File (MAF)/Topologically Integrated Geographic Encoding and Referencing (TIGER²) database (MTdb). This manuscript focuses on the address list component of the MTdb commonly referred to as the MAF.

¹Christine Gibson Tomaszewski is a mathematical statistician in the Decennial Statistical Studies Division of the U.S. Census Bureau. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical and methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

² TIGER is the mapping/spatial portion of the database, which contains street and other feature details, as well as mapspots, or geographic coordinates, corresponding to the location of MAF addresses.

The MAF is a central component of the Decennial census as well as important to a number of other census programs. It was the address frame for the 2000 and 2010 Census. It provides the address frame for all of the demographic surveys including the Current Population Survey, Survey of Income Program and Participation, as well as the Decennial Directorate's American Community Survey (ACS). Its centrality to these major census operations mandates that its accuracy be maintained between the Decennial Censuses as well as prior to each census. The following operations contributed towards maintaining the MAF between the 2000 and 2010 Decennial Censuses:

- Biannual Delivery Sequence File updates from the United States Postal Service;
- Demographic Area Address Listing (DAAL); and
- ACS.

The most substantial updating of the MAF, however, occurred prior to the census in the

- Block Canvassing and Address Listing operations (in Census 2000), and the
- 2010 Address Canvassing (AC) operation (in the 2010 Census).

The 2010 AC operation, with a total cost of approximately 845 million dollars including indirect costs (Holland, 2012), offers significant potential for cost savings research.

The research reported here is part of the larger 2010 Census Program for Evaluations and Experiments (CPEX) and a subset of the 2010 Address List Maintenance Using Supplemental Data Sources (ALMSS) Evaluation (Tomaszewski, 2013). The focus of this paper is on examining the potential for AR to reduce the cost of maintaining and updating the MAF with a special focus on how their use may have been able to reduce the costs of the 2010 AC operation. The driving research question is *'How can we use AR to update the MAF?'*

2. AR and MTdb Universes Eligible for Matching

To answer this research question, we selected the 2007 and 2008 StARS data as the source of the input AR data. The 2007 and 2008 StARS data predate the 2010 AC operation, which occurred in 2009, and therefore could have been used to assist in that operation. StARS is composed of AR data collected from other federal agencies, including the Internal Revenue Service (IRS), Centers for Medicare and Medicaid Services, Department of Housing and Urban Development, Indian Health Service, and Selective Service System.

For this research, approximately 184 million records with complete address types and Zone Improvement Program (ZIP) Codes were identified from the MTdb. Up to four unique addresses for each of these 184 million records were created using the pre- and post- 2010 Census AC operation mailing and location addresses. This resulted in approximately 225 million unique addresses (see Table 1) being eligible for matching to the StARS 2007 and StARS 2008 Master Housing Files (MHF). The StARS 2007 and StARS 2008 MHFs contained approximately 152 million and 154 million unique records respectively. This research focuses on how StARS matched the MAF at a unit level.

Valid addresses for this analysis fell into three types: city-style³, rural route, and Post Office (PO) box addresses. Addresses were required to have a non-blank address type and a ZIP Code of length five in order to be included for StARS matching⁴.

These addresses were unduplicated, and then passed to the Data Integration Division (DID) to be matched to the 2007 and 2008 StARS MHFs. The DID produced a one-to-many address match, matching multiple MAF addresses to each StARS record, and also determined the best unit-level MAF address match for each StARS record. These results were unduplicated, then merged by their unique MAF IDentifiers (MAFIDs) to the 2010 Combo file (Ward, 2011) – for 2010 Census AC results - and the 2010 Census Unedited File (CUF) - for final 2010 Census validity.

Table 1 indicates that city-style addresses were the most common type of MAF address eligible for StARS matching, at nearly 97 percent of all unique MAF addresses.

Address Type	Unique MAF Addresses Eligible for StARS Matching	Percent
City-Style	217,927,468	96.9
Multi-Unit	73,590,084	32.7
Single Unit	144,337,384	64.2
Non-City-Style	7,000,246	3.1
Post Office Box	1,131,408	0.5
Rural Route	2,936,609	1.3
Incomplete.....	2,932,229	1.3
Total Eligible for StARS Matching.....	224,927,714	100.0

Source: Pre-AC file and Post-AC file.
Note: Percentages may not sum to 100 due to rounding.

3. Initial Match Results for AR and MTdb Universes

Table 2 shows that non-city-style StARS addresses often did not match to the MAF.

Address Type	2007 StARS	Percent	2008 StARS	Percent
City-Style.....	24,613,478	62	25,010,151	62
Multi-Unit.....	11,308,708	28	11,658,540	29
Single Unit.....	13,304,770	34	13,351,611	33
Non-City-Style	15,081,187	38	15,372,803	38
Post Office Box	11,696,995	29	11,719,022	29
Rural Route.....	1,656,710	4	1,500,424	4
Incomplete.....	1,727,482	4	2,153,357	5
Total StARS Nonmatches.....	39,694,665	100	40,382,954	100

Source: Pre-AC file, Post-AC file, 2010 Combo file, and StARS 2007-8 files.
Note: Percentages may not sum to 100 due to rounding.

³ Addresses with both a house number and a street name are classified as city-style addresses.

⁴ Nearly 3 million MAF records had insufficient address data for the matching process.

Over one-third of the unit level nonmatches and duplicate matches (38 percent) were non-city-style StARS addresses, according to Table 2. From Table 1, there were only one million PO Box MAF addresses eligible for matching. Note that the majority of the StARS non-city-style nonmatches (over 11 million on each of the two StARS files) were PO Box addresses.

Table 3 shows that city-style addresses were the most common type of StARS address that matched at a unit level to the MAF. Few non-city-style StARS addresses matched to the MAF addresses.

Table 3. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Address Types of Unduplicated Unit Level 2007-8 StARS Matches

Address Type	2007 StARS	Percent	2008 StARS	Percent	Both StARS	Percent
City-Style	112,506,196	100	113,957,379	100	119,424,157	100
Multi-Unit	24,311,677	23	24,923,731	22	27,526,580	23
Single Unit	88,194,519	77	89,033,648	78	91,897,577	77
Non-City-Style	2,848	0	2,880	0	3,114	0
Post Office Box	0	0	460	0	334	0
Rural Route	5	0	52	0	39	0
Incomplete	2,843	0	2,368	0	2,741	0
Total Matching Records	112,509,044	100	113,960,259	100	119,427,271	100

Source: Pre-AC file, Post-AC file, 2010 Combo file, and StARS 2007-8 files.
 Note: Percentages may not sum to 100 due to rounding.

Table 3 and Figure 1 show that after unduplication of the StARS/MAF matches there were 112,509,044 StARS 2007 addresses and 113,960,259 StARS 2008 addresses that matched to unique MAF records. Matching between these two years of StARS yielded a total of 119,427,271 unique StARS/MAF matches.

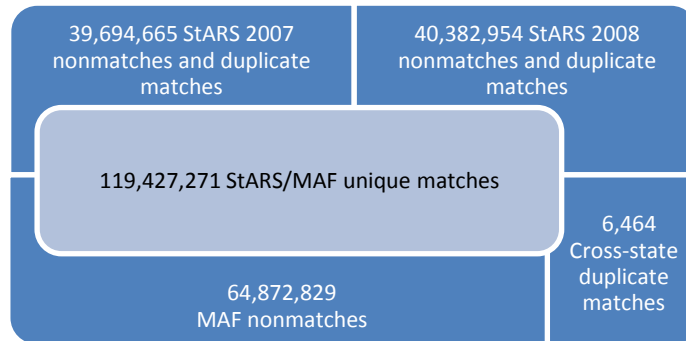


Figure 1. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: Graphical Summary of StARS/MTdb Matches, Nonmatches, and Duplicates
 Source: Pre-AC file, Post-AC file, and StARS 2007-8 files.
 Note that the above figure is not to scale.

4. AR Matches by 2010 Census AC Filter and Actions

The 2010 Census AC filter is a collection of rules used to identify valid MAF addresses for use in the 2010 AC operation. These rules can be recreated every year if desirable. StARS data could also be assembled every year, providing useful updates between censuses. The result of passing the matched data through this filter is shown in Table 4.

This table indicates that nearly 60 percent of the records were valid for the 2010 Census AC filter and matched to StARS.

Table 4. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: MAF/StARS Match Status by 2010 Census AC Filter			
2010 Census AC Filter	StARS Match	StARS Nonmatch	Total
In AC Filter.....	109,031,356 (59.2%)	32,734,928 (17.8%)	141,766,284
Not in AC Filter.....	10,395,915 (5.6%)	32,137,900 (17.4%)	42,533,815
Total.....	119,427,271	64,872,828	184,300,099
Source: Pre-AC file, Post-AC file, 2010 Combo file, and StARS 2007-8 files.			
Note: Percentages may not Sum to 100 due to rounding.			

Table 5 shows the distribution of the 2010 Census AC action codes for the MAF/StARS matches and the MAF/StARS nonmatches. MAF/StARS matches existed for 65 percent of the address list. If records without 2010 Census AC action codes are excluded to approximate the 2010 Census AC workload, then about 79 percent of the 2010 Census AC workload had a StARS address match.

Table 5. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: MAF/StARS Match Status by 2010 Census AC Action Code				
2010 Census AC Action Code	MAF/StARS Match	MAF/StARS Nonmatch	MAF Total	StARS Valid to Invalid Ratio
Adds.....	3,006,304	6,901,473	9,907,777	0.44
New Adds.....	265,163	5,571,084	5,836,247	0.05
Matched Adds.....	2,741,141	1,330,389	4,071,530	2.06
Verify.....	85,909,962	11,601,366	97,511,328	7.41
Change.....	12,639,561	5,805,248	18,444,809	2.18
Move.....	4,813,958	631,709	5,445,667	7.62
Delete (verified).....	3,832,337	10,859,392	14,691,729	0.35
Single Delete.....	371	345	716	1.08
Duplicate.....	1,021,651	2,451,060	3,472,711	0.42
Nonresidential.....	343,450	857,993	1,201,443	0.40
Uninhabitable.....	170,351	352,895	523,246	0.48
Invalid/No Action.....	7,689,326	25,411,347	33,100,673	0.30
Total.....	119,427,271	64,872,828	184,300,099	1.84
Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.				

The last column in Table 5 is the StARS Valid to Invalid Ratio. This is the number of records that were valid in StARS (Match column) divided by the number of records that were invalid in StARS (Nonmatch column) for each 2010 Census AC action code. The positive 2010 Census AC actions were more likely to match to records in StARS, while the negative 2010 Census AC actions were less likely to match to records in StARS.

5. AR Matching Status for 2010 Census AC Adds through the CUF

Overall, 30 percent of the 9,907,777 AC adds existed in 2007 or 2008 StARS before the AC operation. Those StARS-matched 2010 Census AC adds were primarily adds that

matched pre-existing MAF records (67 percent), with a small number of new AC adds (less than five percent). About 98 percent of the 3,006,304 StARS-matched adds were also valid for the 2010 Census, with a ratio of 43 valid records in 2010 Census for every record not valid in the 2010 Census. The balance of 2010 Census AC adds, about 70 percent (6,901,473), did not exist in StARS prior to the 2010 Census AC operation. The majority of those 6,901,473 non-StARS adds were valid 2010 Census records (82 percent), while a smaller number of the non-StARS adds (18 percent) were not in the final 2010 Census. The non-StARS adds had a 2010 Census valid to invalid ratio of 4.4:1.

We tracked the matching records through the assignment of final 2010 Census validity. Table 6 shows that 4.5 percent, of the 5,836,247 new 2010 Census AC adds existed in the 2007 or 2008 StARS before the 2010 Census AC operation.

Table 6. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC New Adds by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
New Adds.....	Nonmatch Nonmatch	Invalid	1,151,502	19.7	3.84
		Valid	4,419,582	75.7	
	Match Match	Invalid Valid	24,703 240,460	0.4 4.1	9.73
Total			5,836,247	100.0	
Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.					
Note: Percentages may not sum to 100 due to rounding.					

The remaining 95.5 percent of the new 2010 Census AC adds did not exist in StARS before the 2010 Census AC operation. New add records are often new construction, and addresses must be occupied by tax filers before appearing in StARS via an IRS source (99 percent of StARS records that matched had an IRS source for the address). This helps explain the high StARS nonmatch rate for this universe. Interestingly, 19.7 percent of the new 2010 Census AC adds both did not exist in StARS before the 2010 Census AC operation and were not considered valid for the 2010 Census.

Table 7 shows that two-thirds of 2010 Census AC adds that matched to pre-existing MAF records also existed in StARS before the 2010 Census AC operation.

Table 7. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Matched Adds by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Matched Adds	Nonmatch Nonmatch	Invalid	122,330	3.0	9.88
		Valid	1,208,059	29.7	
	Match Match	Invalid Valid	42,977 2,698,164	1.1 66.3	62.78
Total			4,071,530	100.0	
Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.					
Note: Percentages may not sum to 100 due to rounding.					

Nearly all of these 2,741,141 matched adds in StARS (98 percent) were also valid for the 2010 Census, with a ratio of 63 valid records in the 2010 Census for every record not valid in the 2010 Census. The remaining one-third of matched 2010 Census AC adds, which did not exist in StARS, also showed high validity (91 percent) for the 2010 Census

with a ratio of 10 valid records in the 2010 Census for every record not valid in the 2010 Census.

6. AR Matching Status for Positive 2010 Census Actions through the CUF

Table 8 shows that over two-thirds of records with 2010 Census AC change actions matched to 2007 or 2008 StARS records.

Table 8. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Change Actions by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Change	Nonmatch	Invalid	635,501	3.4	8.13
	Nonmatch	Valid	5,169,747	28.0	
	Match	Invalid	256,660	1.4	48.25
	Match	Valid	12,382,901	67.1	
Total			18,444,809	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.
 Note: Percentages may not sum to 100 due to rounding.

The distribution for the 18,444,809 records with 2010 Census AC change actions (address changes) in StARS and in the 2010 CUF was similar to the distribution for matched 2010 Census AC adds. Two-thirds of records with 2010 Census AC change actions existed in StARS prior to the 2010 Census AC operation and were valid records for the 2010 Census. Nearly one-third of records with 2010 Census AC change actions (28 percent) did not exist in StARS prior to the 2010 Census AC operation but were valid records for the 2010 Census.

Table 9 shows that nearly nine-tenths of the 97,511,328 records with 2010 Census AC verify actions in the 2010 Census AC operation matched to 2007 or 2008 StARS records.

Table 9. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Verify Actions by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Verify	Nonmatch	Invalid	1,042,182	1.1	10.13
	Nonmatch	Valid	10,559,184	10.8	
	Match	Invalid	658,897	0.7	129.38
	Match	Valid	85,251,065	87.4	
Total			97,511,328	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.
 Note: Percentages may not sum to 100 due to rounding.

Nearly all of these 2010 Census AC verify records in StARS (99 percent) were valid in the 2010 Census, with a ratio of 129 valid records in the 2010 Census for every record not valid in the 2010 Census. A smaller number, nearly 11 percent, of records with 2010 Census AC verify actions did not exist in StARS prior to the 2010 Census AC operation but were valid records in the 2010 Census, with a ratio of 10 valid records in the 2010 Census for every record not valid in the 2010 Census.

Table 10 reveals that nearly nine-tenths of records with 2010 Census AC move actions (where an address record's census block was changed) matched to StARS. The

distribution for 2010 Census AC move actions in StARS and in the 2010 CUF was similar to the distribution of 2010 Census AC verify records. Nearly all of the records with 2010 Census AC move actions that were in StARS (99 percent) were valid records in the 2010 Census, with a ratio of 111 valid records in the 2010 Census for every record not valid in the 2010 Census. A smaller number of 2010 Census AC move records (11 percent) did not exist in StARS prior to the 2010 Census AC operation but were valid records for the 2010 Census, with a ratio of 11 valid records every record not valid in the 2010 Census.

Table 10. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Move Actions by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Move.....	Nonmatch	Invalid	51,834	1.0	11.19
	Nonmatch	Valid	579,875	10.6	
	Match	Invalid	42,958	0.8	111.06
	Match	Valid	4,771,000	87.6	
Total			5,445,667	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.
 Note: Percentages may not sum to 100 due to rounding.

Overall, records with positive 2010 Census AC actions were likely to exist in StARS, with a low of 68.5 percent of records with 2010 Census AC change actions, to a high of 88.4 percent of records with 2010 Census AC move actions matching to StARS. Similarly, a high percentage of 2010 Census AC adds (67.3 percent) that matched to pre-existing MAF records also existed in StARS. However, the same could not be said for the new 2010 Census AC adds.

7. AR Matching Status for Non-Positive 2010 Census AC Actions through the CUF

Tables 11 through 15 examine records with negative 2010 Census AC actions (delete, duplicate, nonresidential, uninhabitable) or no 2010 Census AC actions.

Table 11. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Delete Actions by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Delete.....	Nonmatch	Invalid	10,489,340	71.4	0.04
	Nonmatch	Valid	370,397	2.5	
	Match	Invalid	3,090,446	21.0	0.24
	Match	Valid	742,262	5.1	
Total			14,692,445	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF .
 Note: Percentages may not sum to 100 due to rounding.

Table 11 shows that just over one-fourth of records with 2010 Census AC delete actions matched to a 2007 or 2008 StARS address. The 2010 Census AC deletes in Table 11 include both verified 2010 Census AC delete actions and single 2010 Census AC delete actions.

About 71.4 percent of deletes did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010 Census, while 2.5 percent of deletes did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census. About one-fifth of deleted records did exist in StARS prior to the 2010 Census AC operation but were invalid for the 2010 Census, while 5.1 percent of deletes existed in StARS prior to the 2010 Census AC operation and were valid for the 2010 Census.

Ninety-seven percent of the deletes that did not match StARS were invalid for the 2010 Census, while 81 percent of deletes that matched to StARS were invalid for the 2010 Census. For the deletes that did not match to StARS, there were 28 invalid 2010 CUF records for each valid 2010 CUF record. For deletes that matched to StARS records, there were 0.24 valid 2010 CUF records for each invalid 2010 CUF record.

Table 12 shows that less than one-third of records with duplicate actions (29.4 percent) matched to 2007 StARS or 2008 StARS records. Over two-thirds of duplicates (68.7 percent) did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010 Census, while 1.9 percent of duplicates did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census. For the duplicates that did not match to StARS, there were 0.03 valid 2010 CUF records for every invalid 2010 CUF address record (or 36 duplicates invalid on the 2010 CUF for each duplicate valid on the 2010 CUF). Over one-fourth of duplicates (28.2 percent) existed in StARS prior to the 2010 Census AC operation but were invalid for the 2010 Census, while 1.2 percent of duplicates existed in StARS prior to the 2010 Census AC operation and were valid for the 2010 Census.

2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Duplicate.....	Nonmatch	Invalid	2,386,256	68.7	0.03
		Valid	64,804	1.9	
	Match	Invalid	979,834	28.2	0.04
		Valid	41,817	1.2	
Total			3,472,711	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF .
 Note: Percentages may not sum to 100 due to rounding.

Overall, the majority (71 percent) of both the 2010 Census AC delete and duplicate addresses were not in StARS immediately before the 2010 Census AC operation, and were considered invalid for the 2010 Census. The distribution for 2010 Census AC records determined nonresidential were similar to the 2010 Census AC records with delete or duplicate codes.

Table 13 shows us that less than one-third of 2010 Census AC records with nonresidential actions (28.6 percent) matched to 2007 or 2008 StARS records. Seven-tenths of 2010 Census AC records with nonresidential action codes (70.2 percent) did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010 Census, while 1.2 percent did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census. About 27.1 percent of 2010 Census AC records with nonresidential action codes did exist in StARS prior to the 2010 Census AC operation but

were invalid for the 2010 Census, while 1.5 percent existed in StARS prior to the 2010 Census AC operation and were valid for the 2010 Census.

Table 13. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Nonresidential Actions by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Nonresidential	Nonmatch	Invalid	843,884	70.2	0.02
	Nonmatch	Valid	14,109	1.2	
	Match	Invalid	325,690	27.1	0.05
	Match	Valid	17,760	1.5	
Total			1,201,443	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.
Note: Percentages may not sum to 100 due to rounding.

Table 14 indicates that approximately one-third of records with a 2010 Census AC uninhabitable action code matched to 2007 StARS or 2008 StARS records. Slightly more than one-third of records with 2010 Census AC uninhabitable action codes (36.9 percent) did not exist in StARS prior to the 2010 Census AC operation and were invalid for the 2010 Census. About 30.8 percent of uninhabitable records did not exist in StARS prior to the 2010 Census AC operation but were valid for the 2010 Census.

Table 14. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: 2010 Census AC Uninhabitable Actions by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
Uninhabitable	Nonmatch	Invalid	192,847	36.9	0.83
	Nonmatch	Valid	160,048	30.6	
	Match	Invalid	77,139	14.7	1.21
	Match	Valid	93,212	17.8	
Total			523,246	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.
Note: Percentages may not sum to 100 due to rounding.

Table 15 shows that over three-fourths of MAF records with no 2010 Census AC action code did not match to 2007 or 2008 StARS records.

Table 15. 2010 CPEX Address List Maintenance Using Supplemental Data Sources: No 2010 Census AC Action by StARS Match Status and 2010 CUF Validity					
2010 Census AC Action Code	StARS Match Status	CUF Validity	Count	Percent	CUF Valid to Invalid Ratio
No AC Action	Nonmatch	Invalid	24,944,771	75.4	0.02
	Nonmatch	Valid	466,576	1.4	
	Match	Invalid	7,082,630	21.4	0.09
	Match	Valid	606,696	1.8	
Total			33,100,673	100.0	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.
Note: Percentages may not sum to 100 due to rounding.

Overall, records with a negative 2010 Census AC action code (delete, duplicate, nonresidential, uninhabitable) or no 2010 Census AC action code were not likely to match to StARS, with a low of 67.4 percent of records with uninhabitable actions missing from StARS, to a high of 76.8 percent of records with no 2010 Census AC action code absent from StARS. In a way similar to how the presence of an address record in StARS was likely to indicate a positive 2010 Census AC action code (verify, move, change), the absence of an address in StARS was likely to indicate a negative 2010 Census AC action.

8. AR Matching Status and the 2010 Census AC filter through the CUF

Table 16 provides a distribution of 2010 Census AC filter by StARS match status by 2010 Census validity. The majority of records that were both eligible for the 2010 Census AC operation and existed in StARS were valid for the 2010 Census, with a ratio of 19 valid records to one invalid record in the 2010 CUF. Similarly, the majority of records that were both not eligible for the 2010 Census AC operation and did not exist in StARS were invalid for the 2010 Census, with a ratio of one valid address record to four invalid records in the 2010 CUF. Again, there is a strong relationship between 2010 Census AC filter and StARS inclusion with validity in the 2010 Census.

2010 Census AC Filter	StARS Match Status	In 2010 Census	Not In 2010 Census	Total	CUF Valid to Invalid Ratio
In AC	In StARS	103,575,333 (56.2%)	5,456,023 (3.0%)	109,031,356 (59.2%)	18.98
In AC	Not in StARS	17,020,090 (9.2%)	15,714,838 (8.5%)	32,734,928 (17.8%)	1.08
Not in AC.....	In StARS	3,270,004 (1.8%)	7,125,911 (3.9%)	10,395,915 (5.6%)	0.46
Not in AC.....	Not in StARS	5,992,291 (3.3%)	26,145,609 (14.2%)	32,137,900 (17.4%)	0.23
Total.....	129,857,718 (70.5%)	54,442,381 (29.5%)	184,300,099 (100.0%)	

Source: Pre-AC file, Post-AC file, 2010 Combo file, StARS 2007-8 files, and 2010 CUF.
 Note: Percentages may not sum to 100 due to rounding.
 The yellow and green shaded blocks show where the 2010 Census AC filter correctly predicted the 2010 CUF Validity. The blue and green shaded blocks show where the StARS match status correctly predicted the 2010 CUF Validity

This table also demonstrates that the 2010 Census AC filter was 83.5 percent correct⁵ according to the 2010 CUF, while the StARS match status was 80.7 percent correct⁶. The StARS match status and 2010 Census AC filter agreed with the CUF for 70.4 percent of records⁷ and disagreed with the CUF for 6.2 percent of records⁸. The StARS match status

⁵ Correctness of the AC filter was calculated by adding the In AC/In 2010 Census counts of to the Not in AC/Not In 2010 Census counts. In other words, where the AC filter correctly identified records as valid or invalid for the 2010 Census.

⁶ Correctness of the StARS match status was calculated by adding the In StARS/In 2010 Census counts to the Not in StARS/Not In 2010 Census counts. In other words, where the StARS match status correctly identified records as valid or invalid for the 2010 Census.

⁷ The table entries are highlighted in green: the In AC/In StARS/In 2010 Census count plus the Not in AC/Not in StARS/Not In 2010 Census count.

and 2010 Census AC filter disagreed with each other for 23.4 percent of records⁹, with a net difference of 2.8 percent¹⁰ in predicting the correct final 2010 Census status. This indicates a strong similarity in predicting the final 2010 Census status between the StARS match status and 2010 Census AC filter.

Integrating an AR database match status into the MAF may prove useful for developing address filters for future operational activities. For instance, in a Targeted Address Canvassing (TAC) scenario (Boies, Shaw, and Holland, 2012), records that matched to the AR database and were valid for an address filter could be categorized as valid while records that did not match to StARS and were invalid for an address filter could be considered invalid. In both cases, a canvassing operation would not examine these cases (or blocks, tracks, or other geographical areas with a majority of these records). In a TAC scenario, records where the AR database match status and address filter disagreed on validity (or geographical areas with a majority of these records) could be included for canvassing to determine whether to include these records for future operations. In summary, the relationships shown here were strong – the presence of an address in StARS increased the likelihood of a positive listing outcome and positive final census status; and the absence of an address in StARS increased the likelihood of a negative listing outcome and negative final census status.

9. Conclusions & Recommendations

AR data from StARS 2007 and StARS 2008 has strong potential for contributing to MAF maintenance activities. Overall, of the approximately 225 million unique records in the MAF studied, 53 percent matched to StARS. Almost 100 percent of the address matches were city-style addresses.

Of the approximately 9.9 million adds in the 2010 Census AC operation, StARS contained 67 percent of the adds that matched back to the MAF and five percent of the new adds. This confirms that AR data can provide independent evidence of address validity for the universe of records excluded from the 2010 Census AC filter. If valid AR addresses could be identified in an efficient and accurate manner prior to the listing operation, this AR data potential would have represented a possible addition of over two percent of the addresses in the eligible 2010 Census AC operation workload, potentially reducing address undercoverage.

Identification of potential adds that do not meet the current filter rules could assist MAF maintenance efforts by identifying areas containing records that might be targeted for field operations. For instance, if these potentially valid records are ungeocoded, they could be processed through imputation software to see if sufficiently high quality imputed geocodes are available. Alternatively, the Census Bureau could request

⁸ The table entries are highlighted in red: the In AC/In StARS/Not In 2010 Census count plus the Not in AC/Not in StARS/In 2010 Census count.

⁹ The table entries are highlighted in blue and yellow: the AC filter's correct, yellow predictions of In AC/Not in StARS/In 2010 Census and Not in AC/In StARS/Not In 2010 Census; and the StARS correct, blue predictions of Not in AC/In StARS/In 2010 Census and In AC/Not in StARS/Not In 2010 Census.

¹⁰ The difference between the blue entries and the yellow entries listed in the previous footnote.

governments with significant numbers of these records to consider submitting address files with GIS coordinates.

This research examined how many of the records with positive 2010 Census AC operation action codes were contained in StARS. StARS contained: 88 percent of all 2010 Census AC verify records; 88 percent of all move records; and 69 percent of all change records. This confirms that AR data can provide independent evidence of address validity for the universe of records included from the 2010 Census AC filter. The 2010 Census AC operation cost nearly 845 million dollars, including field and contract overhead costs (Holland, 2012), and therefore offers high cost reduction potential. This AR data potential, if valid AR addresses could be identified in an efficient and accurate manner prior to the listing operation, would have represented a possible reduction of over 72 percent to the 143,356,106 addresses in the eligible 2010 Census AC operation workload.

Further, we examined the number of records with negative 2010 Census AC operation action codes that were in StARS 2007 and 2008. Of all records that received negative action codes (delete, duplicate, nonresidential, uninhabitable), 73 percent were absent from both StARS 2007 and 2008. In particular, 74 percent of records with delete actions were not present in both StARS 2007 and 2008. This AR data potential could have represented a possible reduction of over 10 percent to the 143,356,106 addresses in the eligible 2010 Census AC operation workload.

The presence of an address in StARS increased the likelihood that the address was valid in the 2010 Census AC operation. Of the approximately 143 million addresses in the 2010 Census AC workload¹¹, about 131 million addresses, or 87 percent, were valid. For the approximately 112 million addresses that matched StARS, about 106 million addresses, or 95 percent, were valid in the 2010 Census AC operation. For this universe of records, a match to the StARS files increased the likelihood of a positive listing outcome. Similarly, the absence of an address in StARS increased the likelihood of a negative 2010 Census AC outcome (delete, duplicate, nonresidential, uninhabitable). Of the 143 million addresses in the 2010 Census AC workload, there were about 20 million addresses with negative listing outcomes. For the 39 million addresses that did not match StARS, about 15 million addresses received negative outcomes. For this universe of records, the absence of a match to the StARS files increased the likelihood of a negative listing outcome by 2.7 times.

The 2010 Census AC filter offered an 83.5 percent correct prediction of the 2010 CUF, while the StARS match status offered an 80.7 percent correct prediction of the 2010 CUF. This net difference of 2.8 percent in predicting the correct final 2010 Census status indicates a strong similarity in predicting the final 2010 Census status between the StARS match status and 2010 Census AC filter.

In summary, the strong relationships demonstrated here offer substantial cost reduction potential - the presence of an address in StARS increases the likelihood of a positive listing outcome and positive final census status, while the absence of an address in StARS increases the likelihood of a negative listing outcome and negative final census

¹¹ The 2010 Census AC workload is all 151,199,426 records that were eligible for StARS matching and that received 2010 Census AC operation actions (including all add actions).

status. The potential value indicated by this research and by earlier research of AR data (Prevost and Leggieri, 1999) to the updating and maintenance of the MAF and related databases indicates that the Census Bureau should construct a single integrated administrative record data repository. The process of acquiring, building, integrating, and conducting analysis on the large data files for this AR research required substantial effort. If the matching and integration activities are done once at an agency level, similar to the integration and maintenance of MAF and the Business Register, all research and production activities would benefit significantly. Additionally, the results presented here suggest that expanding access to administrative records for research throughout the Census Bureau under Title 13, 15, and 26 rules and guidelines would offer substantial benefits. Very specifically, updating the MAF with AR flags could facilitate the intra-census updating of the MAF and thus reduce the cost of pre-2020 MAF updating operations. From these findings, AR data sources, including those used to create StARS, will undoubtedly prove very useful in future data modeling, data mining, microsimulations, and targeting research to reduce costs and maintain accuracy during future censuses and surveys.

Acknowledgements

Many people contributed to the successful completion of this report. My branch (John Boies, Jonathan Holland, Nancy Johnson, Kathleen Kephart, Kevin Shaw, Matthew Virgile, and Justin Ward) offered feedback on the methodology, coding, statistical analysis, and supporting prose. Thanks to Jennifer Reichert for her facilitation and feedback. Many, many thanks to Claude Jackson for his endless and invaluable IT support. Thanks to my family for their support throughout this long project. I dedicate this, my first publication, to my Aunt, Therese Marie Tischler, whom I dearly miss.

References

- Boies, John L. (2012), Kevin Shaw, and Jonathan Holland, "Final Report for the 2010 Census Evaluation of Address Canvassing Targeting and Cost Reduction," DSSD 2010 CPEX Memorandum Series A-09, July 12, 2012.
- Bureau of the Census (2009), "Introduction to the 2020 Census," Washington, D.C., June 18, 2009.
- Bureau of the Census (2010), "Strategic Plan for the 2020 Census," Washington, D.C., Version 2.0, September 30, 2010.
- Holland, Jonathan P (2012), "2010 CPEX: Evaluation of Automation in Field Data Collection in Address Canvassing Report," DSSD 2010 CPEX Memorandum Series A-05, July 24, 2012.
- Prevost, Ron (1999) and Charlene Leggieri, "Expansion of Administrative Records Uses at the Census Bureau: A Long-Range Research Plan." Presented at the November 1999 Meeting of the Federal Committee on Statistical Methodology, November 1999.
- Tomaszewski, Christine Gibson (2013), "2010 CPEX Report: Evaluation of Address List Maintenance Using Supplemental Data Sources," DSSD 2010 CPEX Memorandum Series A-06, February 20, 2013.
- Ward, Justin (2011), "Documentation for the 2010 Census Address Frame Combination File, Version #1," DSSD 2010 Decennial Census Memorandum Series D-19, August 24, 2011.
- Reports in the 2010 Census Program for Evaluations and Experiments (CPEX) are located at <http://www.census.gov/2010census/about/cpex.php>.