

2020 Census Program Management Review

Address Validation Test Results

Evan Moffett, Decennial Census Management Division
Pat Cantwell, Decennial Statistical Studies Division
Mike Ratcliffe, Geography Division

July 10, 2015

V1.0 Final

Outline

- General Background for the Address Validation Test (AVT)
- Results of statistical modeling in the MAF Model Validation Test (MMVT)
- Results of the Partial-Block Canvassing Test (PBC)

Purpose for the Address Validation Test (AVT)

- To evaluate our methods for a reengineered address canvassing
- To test how well in-office procedures can replace in-field procedures
- To assess our ability to ensure an accurate Master Address File (MAF)

Background for the AVT

Two components to the Address Validation Test (AVT)

- MAF Model Validation Test (MMVT)
 - September – December 2014
 - To assess statistical models
 - Nationally representative sample, full block canvassing
- Partial-Block Canvassing Test (PBC)
 - December 2014 – February 2015
 - To test ability to canvass partial blocks
 - Blocks identified through imagery as ideal for PBC

MAF Model Validation Test (MMVT): Background

- National sample of 10,100 blocks
 - 10,000 blocks with at least one address, representative of universe
 - 100 blocks selected from those with no addresses
- \approx 1.04 million addresses
- Full-block dependent canvassing: verify, update, add, or delete addresses on the dependent list

MMVT: Research Questions

- Can we use statistical models to determine specific blocks that need additional action, such as in-office or in-field canvassing?
- Can we use statistical models to predict national totals of coverage errors on the MAF?

MMVT: Summary

- Summary of results: The statistical models we applied were not effective at ...

MMVT: Summary

- Summary of results: The statistical models we applied were not effective at ...
 - identifying specific blocks with many Adds or Deletes

MMVT: Summary

- Summary of results: The statistical models we applied were not effective at ...
 - identifying specific blocks with many Adds or Deletes
 - predicting national totals of MAF coverage errors

MMVT: Address Results for 20% Address Canvassing

	Estimate (×1,000)	Statistical Model (See slide at the end for model descriptions)			
		Model 1	Model 2	Model 3	Model 4
% of Addresses in Canvass	136,309	20%	20%	20%	20%
% of Blocks in Canvass	6,294	47.3%	36.5%	4.4%	26.2%
		Rate of Capture			
Adds	5,688	47.2%	47.0%	27.2%	47.1%
Deletes	7,592	34.5%	53.5%	38.6%	46.3%

MMVT: Block Results for 20% Address Canvassing - Adds

Type of Block	Blocks (×1,000)	Model 1	Model 2	Model 3	Model 4
Total Blocks in Frame	6,294	47.3%	36.5%	4.4%	26.2%
Rate of Capture of Specified Blocks					
Blocks w/ 5+	188	43.2%	45.1%	33.8%	47.7%
Blocks w/ 1 - 4	1,185	56.0%	42.3%	10.1%	37.8%
Blocks w/ 1+	1,373	54.2%	42.7%	13.3%	39.2%
Rate of Blocks Erroneously Canvassed					
Blocks w/ 5+	*	97.3%	96.3%	76.8%	94.6%
Blocks w/ 1+	*	75.0%	74.5%	33.2%	67.4%

* Varies according to the model

MMVT: Matrix of Predictions by Outcomes - Adds

Predicted Number of Adds	Estimated # of Blocks	Observed Number of Adds								
		0	1	2	3 to 4	5 to 9	10 to 19	20 to 49	50 to 99	≥100
0 to <0.25	794,800	92.9%	5.8%	0.5%	0.3%	0.5%	0.0%	0.0%	0.0%	0.0%
0.25 to <0.5	1,347,200	89.8%	7.3%	1.6%	0.7%	0.3%	0.2%	0.0%	0.0%	0.0%
0.5 to <1	1,785,800	85.2%	11.0%	1.7%	1.0%	0.7%	0.1%	0.2%	0.0%	0.1%
1 to <1.5	784,900	73.2%	16.9%	6.2%	2.2%	0.8%	0.6%	0.1%	0.1%	0.0%
1.5 to <2	416,400	69.9%	15.3%	6.7%	4.6%	2.3%	0.9%	0.2%	0.0%	0.1%
2 to <2.5	302,900	60.2%	24.4%	7.5%	4.6%	2.3%	0.8%	0.0%	0.1%	0.0%
2.5 to <3	211,900	55.7%	20.6%	10.1%	7.0%	4.2%	1.1%	0.9%	0.3%	0.2%
3 to <5	378,300	49.1%	21.2%	10.3%	8.5%	6.2%	2.5%	1.5%	0.4%	0.3%
5 to <10	220,300	37.8%	16.8%	12.5%	12.2%	11.2%	4.6%	3.1%	1.1%	0.6%
10 to <20	45,100	33.2%	13.0%	8.5%	8.9%	12.6%	9.6%	7.8%	4.1%	2.3%
20 to <50	5,200	13.6%	18.1%	3.8%	7.7%	11.7%	15.9%	9.4%	15.4%	4.4%
≥50	800	36.4%	0.0%	12.2%	36.4%	0.0%	0.0%	8.1%	0.0%	6.8%

MMVT: Coverage Estimates and Predictions

Type of Address	Sample-Based Estimate (× 1,000)		Model-Based Predictions (x 1,000)
	Estimate	SE	Estimate
Addresses for Canvassing ¹	135,897	0	135,897
Add ²	5,688	592	8,587
Delete	7,592	310	8,707
Net	133,993	654	135,777

¹ A tally of the addresses in the MMVT frame that were eligible for the dependent list, on the MAF in July 2013.

MMVT: Results for Statistical Models

- Determining specific blocks that need additional action:
 - rate of error capture was too low
 - rate of erroneous canvass was too high
- Using statistical models to predict national totals of coverage errors on the MAF:
 - model parameters reflected condition of MAF in 2009
 - now: only halfway through decade, and MAF has improved under Geographic Support System Initiative

PBC: Purpose and Research Questions

Why conduct PBC? Two assumptions:

- PBC would be more efficient in large land area blocks by avoiding the need to traverse the entirety of the block to collect changes;
- In blocks with large numbers of addresses that are in the MAF and can be validated in the office, PBC would save time by focusing effort on the portion of the block in which change has occurred.

Research Questions:

- Can growth or change within a block be accurately identified and listed without also canvassing the portion of the block that has remained stable?
- Can listers efficiently follow instructions to list the block portion of interest?
- Can this be done without compromising address coverage in the MAF?
- How effective are in-office methodologies (specifically, imagery review) at identifying where housing unit changes have occurred?

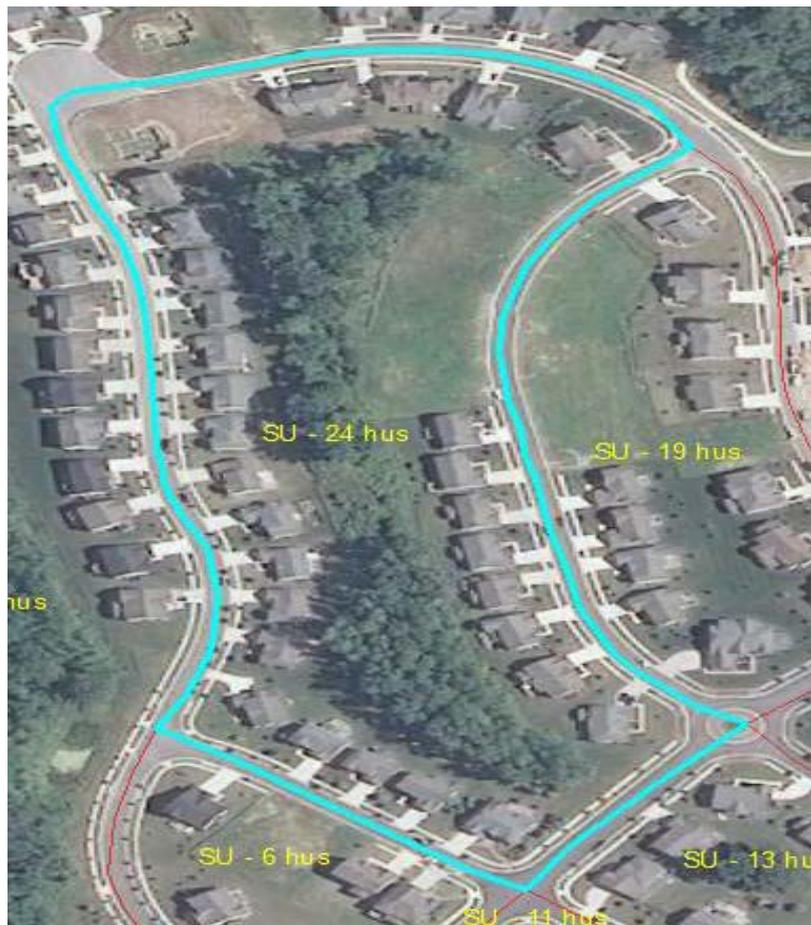
PBC Test Analysis

- Analyze results from fieldwork in comparison to expectations based on in-office review.
 - Did we collect the information we expected?
 - Did we find additional updates in the field?
- For blocks in both the partial-block and full-block canvassing, compare results and assess reasons for differences, if any.
 - Did full-block canvassers find additional updates, especially any that might not be detectable through in-office review?

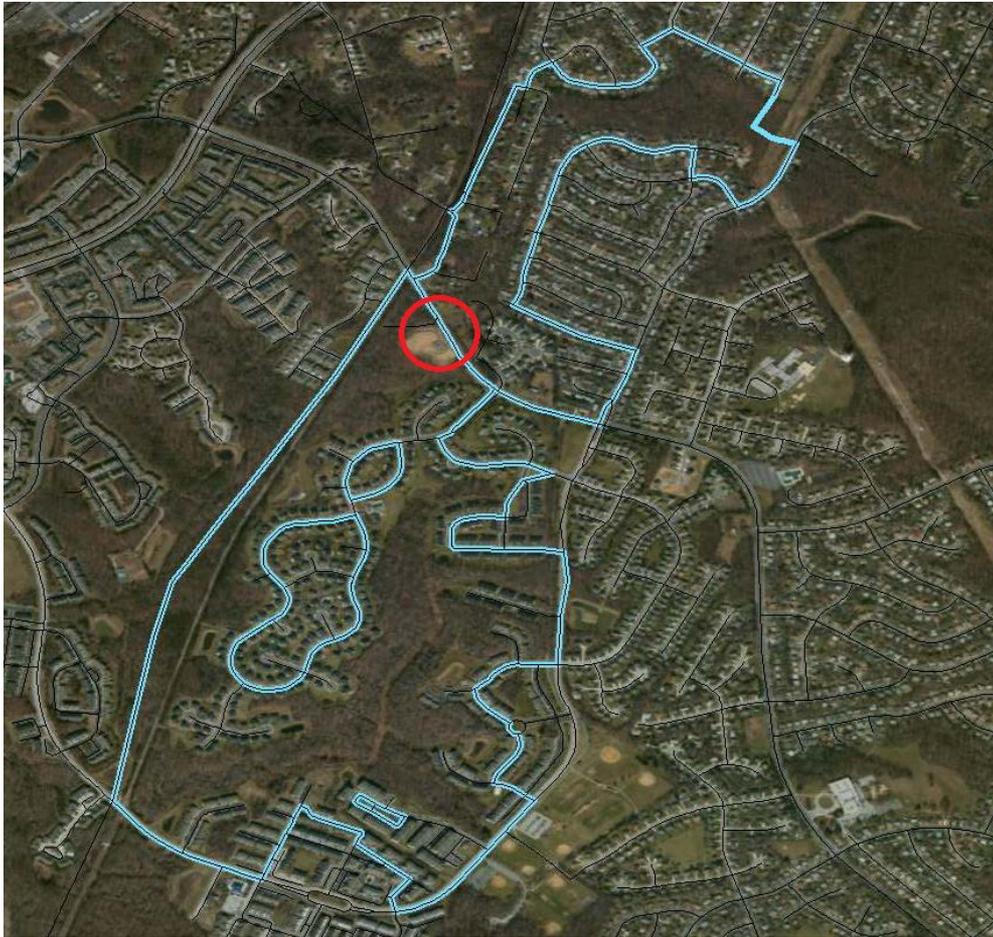
PBC – Field Implementation

- Of the 10,100 MMVT blocks, 615 were selected for PBC listing based on imagery review and comparison to the Master Address File. This is the same imagery-based review that forms a key component of our Reengineered Address Canvassing strategy.
- 37 Assignment Areas, with nationwide distribution (Alaska, Hawaii, and Puerto Rico were out of scope).
 - 705 work areas within 615 blocks.
- 35 professional staff from headquarters and regional offices.
- Listing in the field, December 16, 2014 to January 17, 2015.
 - Quality control listing, December 16, 2014 to February 9, 2015.
- Used Census Bureau's corporate Listing and Mapping Application (LiMA)— first use in a production operation.

Imagery review identifies discrepancy between MAF and imagery; updates are clustered in a portion of the block



Imagery review identifies discrepancy between MAF and imagery; updates are clustered in a small portion of the block



PBC Test Analysis

- Analysis of PBC results indicates that we can successfully implement the PBC methodology in the field.
 - Additional analysis is needed to determine specific contexts in which the PBC methodology will be most effective and efficient, as well as level of expertise and experience necessary for listers.
- All PBC listers successfully navigated to specific work assignments based on written descriptions from in-office imagery reviewers.
 - In some instances, unclear and confusing instructions affected the listers' ability to carry out work effectively and accurately.
 - Work area polygons defined in the office based on new development visible in imagery did not always match the extent of new development on the ground.

PBC Instruction Types

- Type A: Whole Street: 169 work areas
 - Example: “Canvass the entirety of Johnson Court.”
- Type B: Bounded Street Segment: 320 work areas
 - Example: “Canvass Berks Street from 46 Berks St. to 50 Berks St.”
- Type C: Branching Roads: 93 work areas
 - Example: “Canvass starting at the intersection of Freedom Hwy. and Redwood Drive. Canvass Redwood Drive and any roads that branch off of Redwood Drive.”
- Type D: Exclusion: 9 work areas
 - Example: “Canvass the whole block but do not canvass Tealwood Drive or any roads branching off Tealwood Drive.”
- Type E: Targeted Address/Structure: 56 work areas
 - Example: “Check for residential apartments above storefront at 1567 Atwood Avenue.”
- Type F: Other: 58 work areas

PBC Results

Actions	Number of PBC Actions	Percent of Total PBC Actions
Total Actions	17,627	100.00%
Total Adds	10,189	57.80%
True Adds (i.e., new to the MAF)	4,301	
Adds Matched to Ungeocoded Addresses in the MAF	2,931	
All Other Adds Matched to MAF	2,957	
Deletes	387	2.20%
Duplicates	20	0.11%
Changed to Nonresidential	2	0.01%
HU Converted to GQ	1	0.01%
GQ Converted to HU	1	0.01%
Verify	7,027	39.86%

Comparison of MMVT to PBC Actions

Actions in All PBC Blocks	MMVT		PBC Test	
	Number of Actions	Percentage of Total Full-Block Actions	Number of Actions	Percentage of Total PBC Actions
Total Add Actions Compared to All Other Actions in PBC Blocks				
Total Adds	9,344	6.89%	10,189	57.81%
All Other Actions	126,251	93.11%	7,438	42.20%
Total Actions	135,595	100.00%	17,627	100.00%
Comparison of 3 Types of Add Actions in PBC Blocks				
True Add	3,003	2.21%	4,301	24.40%
Adds Matched to Ungeocoded Addresses in the MAF	3,594	2.65%	2,931	16.63%
All Other Adds Matched to MAF	2,747	2.03%	2,957	16.78%

PBC Key Take-Aways

- When PBC did not find an address that was located by MMVT, reasons for the omission tended to be:
 - The area was provided to the PBC lister, but the instruction was poorly-worded or the polygon was poorly-defined, leading to lister confusion.
 - The add represented a situation not detectable by the imagery review step (i.e., changes within existing structures), and therefore was not provided as a work area to the PBC lister.
 - The add represented a situation that was not detected due to imagery quality and/or vintage issues, and therefore was not provided as a work area to the PBC lister.

PBC Key Take-Aways

- Polygons and instructions prepared for PBC listers generally resulted in successful navigation to and within work areas, and facilitated accurate data collection, but improvements need to be made:
 - Ensure instructions match the polygon;
 - Include imagery on the LiMA to aid in understanding the polygon and instruction;
 - Use basic street address information within an instruction; and;
 - Missing and misaligned street features and misaligned block boundaries should be fixed in the office before any block goes to the field.

PBC Key Take-Aways

- Based on the results of the PBC Test, we recommend:
 - Testing PBC in the 2016 Address Canvassing Test with traditional listers.
 - Overlap with full-block canvassing for a sample of blocks to compare results.
 - Improve clarity of written instructions as well as training to minimize lister confusion in the field.
 - Conduct additional analysis at the individual address level to fully understand differences between MMVT and PBC listing results and imagery review results. This analysis is critical for informing implementation of in-office canvassing methodologies.

Comparison of Imagery Review to MMVT Adds in the 10,100 MMVT Blocks

Category of MMVT Add Actions in Comparison to Imagery Review Results	Number of MMVT Adds Detected by Imagery Review (10,100 blocks)	Percentage of Total MMVT Adds
Total MMVT Adds Actions (“good adds”)	18,367	100.0%
Detected by Imagery Review	12,984	70.7%
Detectable by imagery, but missed due to imagery quality, review process issues, reviewer errors	1,168	6.4%
Detectable by imagery, but missed due to imagery vintage	996	5.4%
Undetectable by imagery (within structure changes)	2,682	14.6%
Unable to assess (location description only; unable to determine location of add)	537	2.9%

Comparison of Imagery Review Results to Observed Number of MMVT Adds

Detected from Imagery Review	Observed Number of Adds in Block							Total Blocks (×1,000)
	0	1	2	3 to 4	5 to 10	11 to 20	≥21	
-100 to -2	62.1	13.4	10.8	0.4	10.9	0.0	2.4	24
-1	56.8	17.8	13.5	6.9	3.0	0.0	2.1	66
0	82.1	11.1	3.0	1.9	1.3	0.3	0.3	5,696
1	42.7	28.7	14.5	8.6	3.9	1.1	0.5	355
2 to 10	27.9	22.3	10.2	13.9	14.5	5.4	5.8	127
11 to 20	11.1	13.4	2.6	0.9	15.0	37.9	19.1	13
≥21	12.6	4.9	0.3	0.3	1.6	14.3	65.8	14
Total Blocks (×1,000)	4,921	779	247	159	113	37	38	6,294

Analysis of Imagery Review Results to Inform In-Office Canvassing

- Based on weighted results of imagery review for the 10,100 MMVT blocks:
 - 84% of blocks with at least one address are stable.
 - These blocks encompass an estimated 85% of all housing units.
 - These blocks would be placed in a “passive” category, with ongoing monitoring for change, but not requiring active processing to acquire updates.
 - 15% of housing units are located in “active” blocks, with updates acquired through the USPS’ Delivery Sequence File, local government partner files, other administrative or commercial address lists, or fieldwork.

Summary of the Address Validation Test

- Objective was to evaluate our methods for a reengineered address canvassing:
 - Statistical models we applied were not effective at (a) identifying specific blocks with many Adds or Deletes, or (b) predicting national totals of MAF coverage errors
 - Showed that Partial Block Canvassing methodology offers the potential to implement a more efficient approach to canvassing
- Objective was to test how well in-office procedures can replace in-field procedures:
 - Demonstrated the utility of imagery review to guide decision-making and operational planning for address canvassing
 - Demonstrated the value of fieldwork to gather information for use in assessing the effectiveness of in-office methods

Summary of the Address Validation Test (Continued)

- Objective was to assess our ability to ensure an accurate Master Address File (MAF):
 - Statistical models were ineffective at measuring MAF coverage error
 - Ongoing research will focus on collecting metrics via the MAF Coverage Study

Extra Slides

MMVT: Key Milestones

Activity	Begin	End
Field Data Collection	September 2, 2015	December 18, 2015
Receive MMVT Data for Analysis	January 30, 2015	January 30, 2015
Analyze MMVT Data	February 2, 2015	May 1, 2015
Release Draft Report for Comment	May 1, 2015	May 15, 2015
Issue Final AVT Report		June 30, 2015

Statistical Models in the Tables

Model 1: Zero-inflated negative binomial distributional model for number of Adds in block

Model 2: Zero-inflated negative binomial distributional model for number of Deletes in block

Model 3: Logistic regression model predicting the probability that the block contains 2 or more Adds + Deletes

Model 4: Logistic regression model predicting the probability that the block contains 1 or more Adds

Data in the Research of Statistical Models from the 2010 Census

- Dependent variables, outcomes from 2009 Address Canvassing
- Covariates, available prior to Address Canvassing
 - Census 2000
 - USPS Delivery Sequence File (DSF)
 - ACS eligible units
 - Land coverage database

MMVT: Time Lag Impacting Analysis

- 2013, July: Drew dependent address list, and ran model predictions
- 2014, Sept – Dec: Conducted field work

Analysis of Adds: Difficult, perhaps impossible, to distinguish addresses that were missing from the MAF in July 2013 from those that are new additions to the housing inventory by late 2014.

MMVT: Block Results for 20% Address Canvassing - Deletes

Type of Block	Blocks (×1,000)	Model 1	Model 2	Model 3	Model 4
Total Blocks in Frame	6,294	47.3%	36.5%	4.4%	26.2%
Rate of Capture of Specified Blocks					
Blocks w/ 5+	305	39.7%	54.3%	38.2%	48.9%
Blocks w/ 1 - 4	1,558	52.3%	48.1%	6.9%	37.3%
Blocks w/ 1+	1,863	50.3%	49.1%	12.0%	39.2%
Rate of Blocks Erroneously Canvassed					
Blocks w/ 5+	*	95.9%	92.8%	57.5%	91.0%
Blocks w/ 1+	*	68.5%	60.1%	18.3%	55.7%

* Varies according to the model

MMVT: Matrix of Predictions by Outcomes - Deletes

Predicted Number of Adds	Estimated # of Blocks	Observed Number of Deletes								
		0	1	2	3 to 4	5 to 9	10 to 19	20 to 49	50 to 99	≥100
0 to <0.25	794,800	87.9%	9.5%	1.6%	0.3%	0.7%	0.0%	0.0%	0.0%	0.0%
0.25 to <0.5	1,347,200	83.7%	11.8%	2.6%	0.9%	0.9%	0.1%	0.0%	0.0%	0.0%
0.5 to <1	1,785,800	78.4%	13.6%	3.9%	2.9%	0.9%	0.1%	0.1%	0.1%	0.0%
1 to <1.5	784,900	68.3%	18.4%	6.5%	4.0%	1.4%	1.0%	0.3%	0.1%	0.0%
1.5 to <2	416,400	61.5%	20.1%	7.3%	6.3%	3.0%	1.4%	0.3%	0.1%	0.0%
2 to <2.5	302,900	47.5%	22.3%	15.8%	6.4%	5.7%	1.4%	0.7%	0.2%	0.0%
2.5 to <3	211,900	40.0%	24.2%	12.6%	12.4%	8.2%	1.2%	1.3%	0.0%	0.1%
3 to <5	378,300	33.5%	20.2%	12.2%	17.7%	10.7%	3.9%	1.1%	0.4%	0.3%
5 to <10	220,300	20.9%	13.6%	9.7%	17.5%	21.4%	9.4%	4.3%	2.6%	0.5%
10 to <20	45,100	20.7%	9.4%	7.5%	9.1%	17.3%	16.0%	10.4%	5.0%	4.6%
20 to <50	5,200	7.9%	25.5%	0.9%	3.1%	12.7%	17.9%	15.3%	5.1%	11.6%
≥50	800	36.4%	0.0%	2.7%	0.0%	36.4%	2.7%	12.2%	0.0%	9.5%

MMVT: Coverage Estimates and Predictions (full table)

Type of Address	Sample-Based Estimate (× 1,000)		Model-Based Predictions (× 1,000)		
			Applied to Sample Blocks (weighted)		Applied to Entire Frame (unweighted)
	Estimate	SE	Estimate	SE	Estimate
Addresses for Canvassing ¹	135,897	0	135,897	0	135,897
Add ²	5,688	592	8,533	96	8,587
Delete	7,592	310	8,546	293	8,707
Net	133,993	654	135,884	285	135,777

¹ A tally of the addresses in the MMVT frame that were eligible for the dependent list, on the MAF in July 2013.

² Recall the lag between the time the dependent list was drawn and the data were collected.

Questions



2020.census.pmr@census.gov