

2020 Census Program Management Review

Frame Research

Targeted Address Canvassing and Master Address
File Modeling and Testing

Robin A. Pennington
Robert Colosi

March 28, 2014

Outline

- Targeted Address Canvassing Overview
- MAF Error Model
- MAF Model Validation Test
- Frame Schedule Review

Targeted Address Canvassing Objective

- Define the Targeted Address Canvassing (TAC) workloads and geographic distribution
 - Federal Land Use layer
 - Operational layer
 - Aerial Imagery layer
 - Statistical Model layer
- Cost and quality metrics produce set of alternatives

2010 Address Canvassing Assessment

Table 11.7
The 2010 Census Address Canvassing Operation:
Results compared to the Census 2000 Block Canvassing operation

Final Address Actions	2010 Census Address Canvassing		Census 2000 Block Canvassing	
	Count*	Percent of total [†]	Count*	Percent of total [†]
Total	156,705,156	100.00	97,894,639	100.00
Add	10,776,894	6.88	6,389,271	6.53
New	6,624,155	4.23	4,536,234	4.63
Matches to Existing Record	4,152,739	2.65	1,853,037	1.89
Change	19,608,785	12.51	2,295,168	2.34
Move	5,450,563	3.48	2,948,414	3.01
Verify	97,635,517	62.31	81,115,466	82.86
Negative Actions	21,143,737	13.49	4,972,041	5.08
Does Not Exist (Double Delete only)	15,819,921	10.10	4,452,888	4.55
Duplicate	4,085,556	2.61	154,869	0.16
Nonresidential	1,238,260	0.79	364,284	0.37
Uninhabitable	551,566	0.35	174,279	0.18
Rejected Records	1,536,094	0.98		

*Counts and percentages are unweighted.

[†]Percentages may not sum to 100 due to rounding.

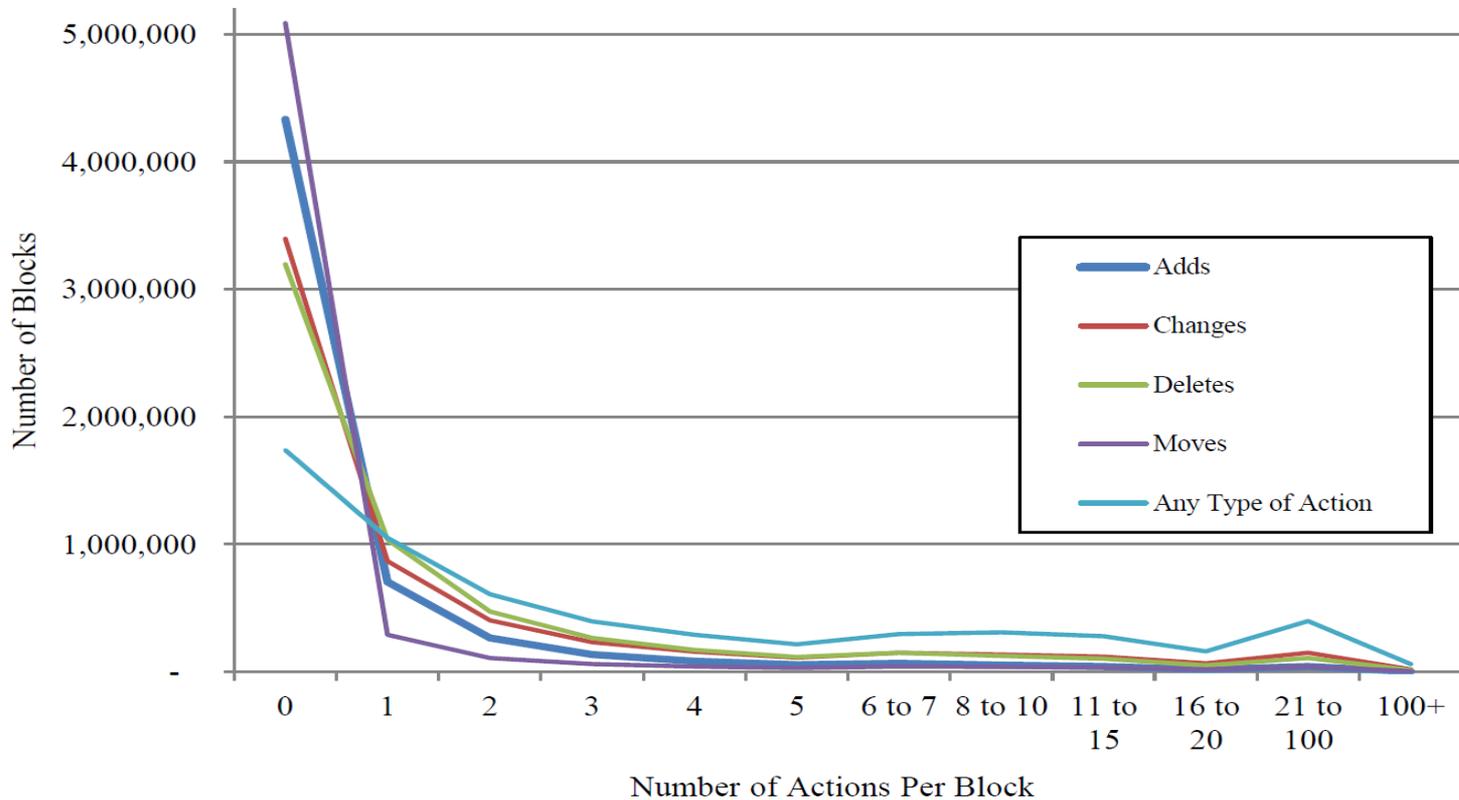
The Census 2000 Address Listing operation, an independent listing not depicted above, added 23,271,819 new Stateside and Puerto Rico records to the MTdb. Adds from Address Listing combined with Block Canvassing represent 25 percent of the total actions to update records on the MTdb.

Verify in this table means that the address was found in AC and there were no changes to the address component of the record.

Negative Actions and Uninhabitable in this table is the same as "Delete" category in Burcham, 2002.

Sources: GQV Extract Files, as defined by the matched MAFSRC and ACTION operation variables, GEO AC Listed Records Tally File, Ruhnke, 2002, and Burcham, 2002.

**Figure 1. 2010 CPEX AC Targeting:
Number of Action Codes Per Block - Adds, Changes
Deletes, Moves, and Any Type of Action**



Source: TAC MUMS File.

MAF Error Model Objective

- The objective of the MEM project is to provide statistical models for the MAF that will produce estimates of coverage error at levels of geography down to the block level
 - These models could potentially inform Targeted Address Canvassing decisions

What is the MAF Error Model?

Source: Derek Young, Center for Statistical Research and Methods

- We developed two predictive models at the block level, collectively known as the “MAF Error Model”
 - One each for the number of **adds** and number of **deletes** as functions of identified predictors
- We use zero-inflated (ZI) regression models
- Zero-inflated models can provide a model-based approach to obtaining coverage estimates
 - Provides more granularity at lower levels of geography over other common modeling approaches (e.g., logistic regression)

Files Used to Create 2009 Vintage MAF Error Model

- 2013 American Community Survey Geographic Reference File
- Block Crosswalk File
- Combo File - Selected Variables from:
 - Census 2000
 - Address Canvassing MAF Extract
 - Group Quarters Validation & Enumeration MAF Extract

Some Variables in Current Model

- “Adds” Model (22 Variables Total)
 - Examples: blocks with >1 mobile home, average number of vacants in a block, blocks with >1 small multi-units
- “Deletes” Model (46 Variables Total)
 - Examples: urban/rural area, blocks with >1 small multi-unit, records with nonresidential status

Model-Based Coverage Measures

Dependent Variables

- Used AdCan action codes for 2009 Model
- Undercoverage (“adds”) defined as:
 - “True Adds” + “Reinstated Adds”
- Overcoverage (“deletes”) defined as:
 - “Deletes” + “Single Deletes” + “Duplicates” + “Uninhabitables” + “Nonresidentials”

Formulas for Coverage Measures

1. Undercoverage = Total # Adds / Total # LQs
2. Overcoverage = Total # Deletes / Total # LQs

Note: Coverage measures produced will be point-in-time estimates.

MAF Model Validation Test Objectives

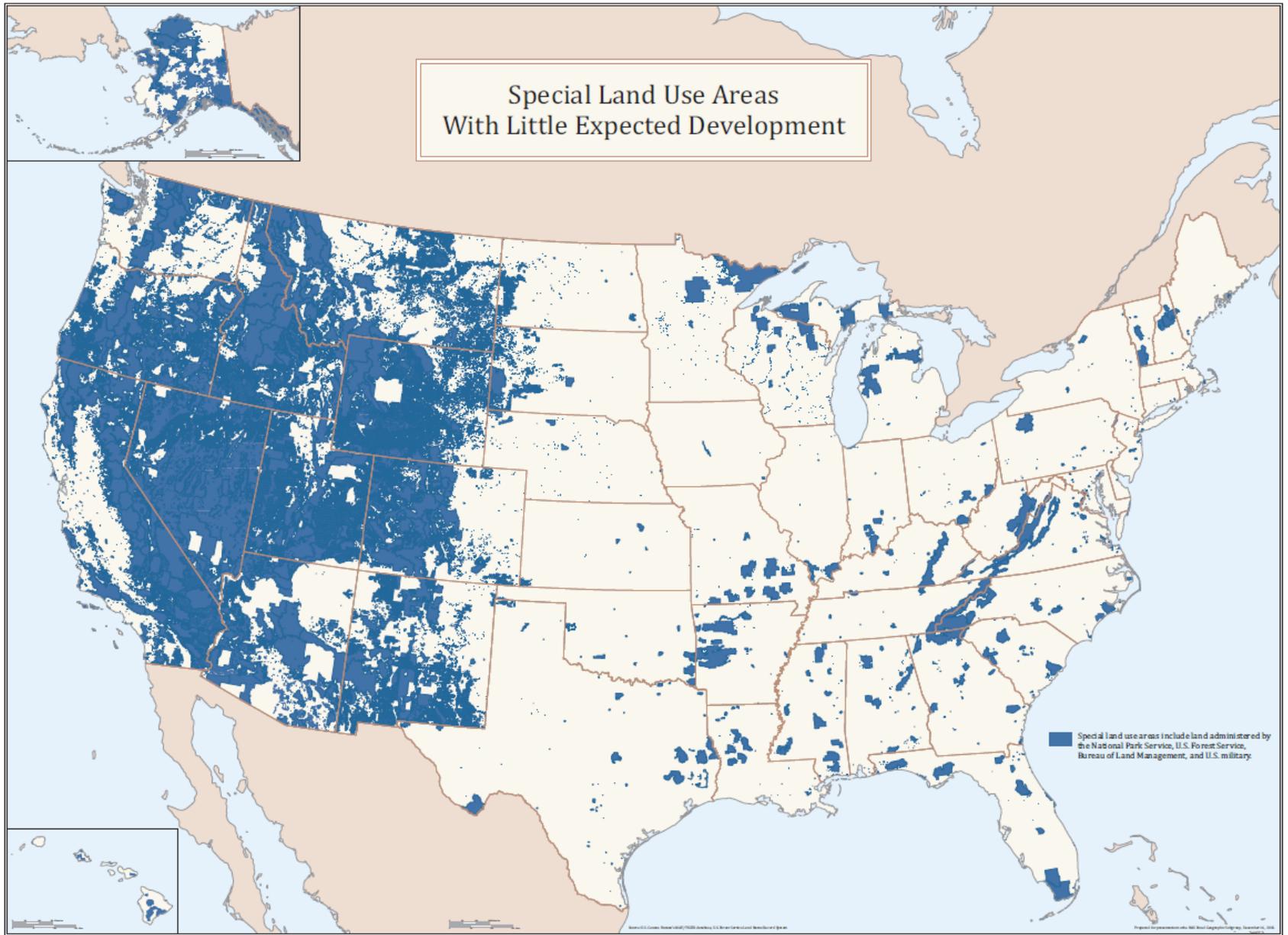
- The purpose of the MAF Model Validation Test (MMVT) is to collect data to inform components of the Targeted Address Canvassing decision-points
 - MAF Error Model
 - Targeted Address Canvassing, Research, Model, and Classification team
 - Models for Zero Living Quarters blocks
- Concept test Micro-Targeting and uses of Aerial Imagery

MMVT Overview

- The field components are scheduled for September – December, 2014
- Areas in the contiguous United States were eligible for the test, with some exceptions, such as Federal Lands
- The model validation test will use existing listing instruments (ALMI) and IT systems, with only some minor modifications to final processing in Geography Division
- The Micro Targeting “proof-of-concept” test hopes to use the LiMA
- We are not testing listing procedures – we’re collecting data for testing the models and proof-of-concept testing for micro targeting and aerial imagery methods.

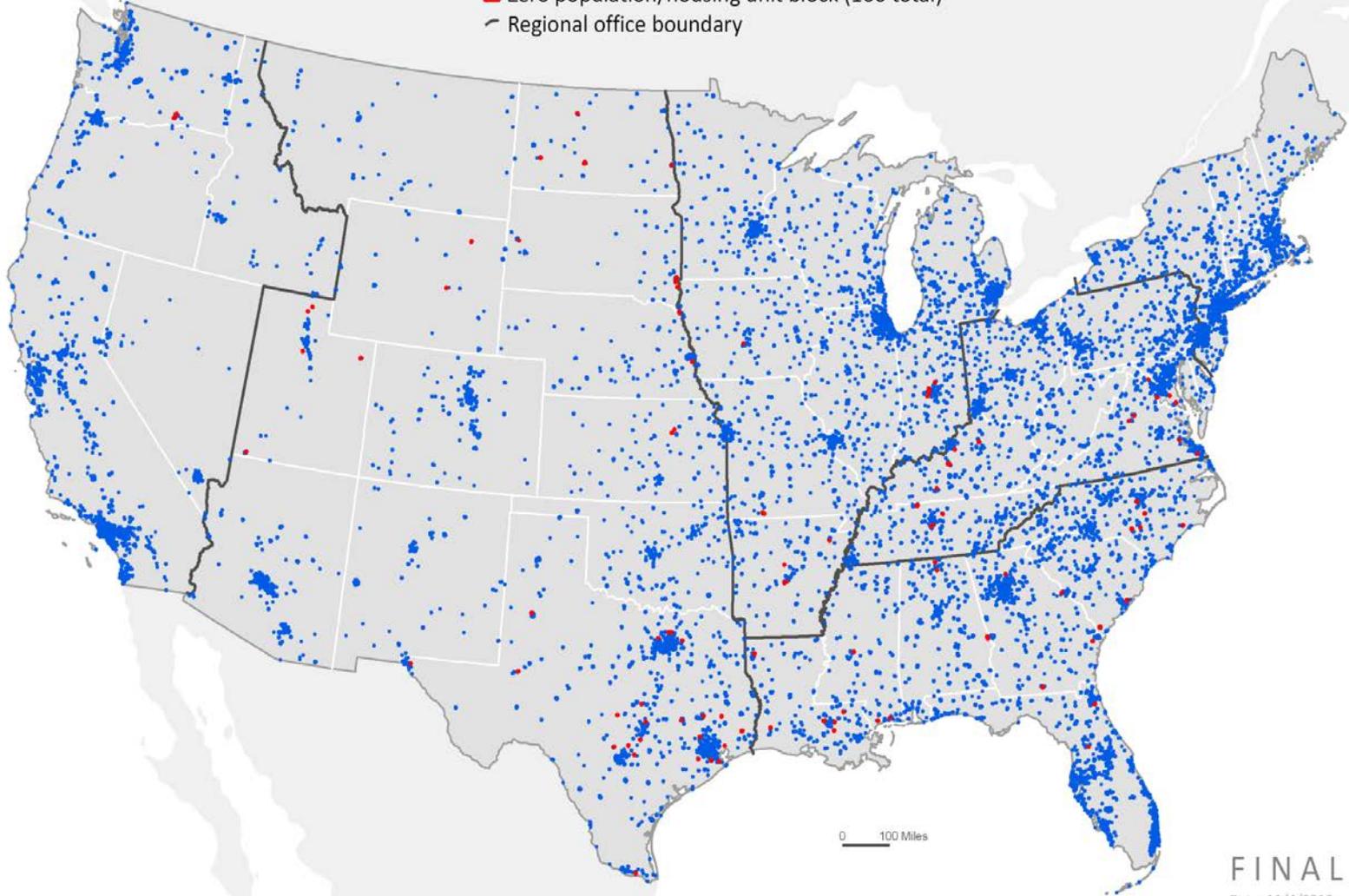
MMVT Design

- A sample size of 10,000 blocks has been determined as the minimum to meet our study needs
 - The sample has been stratified by # of Living Quarters and sorted by state and county for selection.
- An additional selection of 100 blocks with no known living quarters was added to the workload for this test
- Another 600-1000 blocks, some of which overlap with the 10,000 blocks selected by sampling, will be included in a Supplemental MMVT field test to test operationalizing micro-targeting and to inform methods to incorporate aerial imagery review



MAF Model Validation Test

- MMVT selected block (10,000 total)
- Zero population/housing unit block (100 total)
- Regional office boundary



FINAL
Date: 11/4/2013

Next Steps – Model Validation

- Include any updated values of predictors (possibly through proxy variables) in the 2009 and 2013 models
- For each block in the MMVT sample, compute the estimated distribution and a prediction interval (PI); e.g., 90% or 95%
- See how many of the PIs capture the number of each action observed in the MMVT
 - Use a threshold for the proportion of PIs that correctly capture the observed values
- These outputs will be one input into the Geography Division recommendation on Targeted Address Canvassing

Next Steps – Targeted Address Canvassing

- Deliver blocks for identified options (see next slide) to GEO for mapping and to other groups to estimate downstream operational impacts
- Investigate new variables the MAF Error Model Team is adding to current database (*e.g.* AdRecs)
- Develop models for blocks with zero living quarters (commercial, median strips, recently annexed)
- Consider adjustments to blocks targeted based on geographic proximity
- Develop a plan for decision criteria for statistical models, micro targeting, and interactions

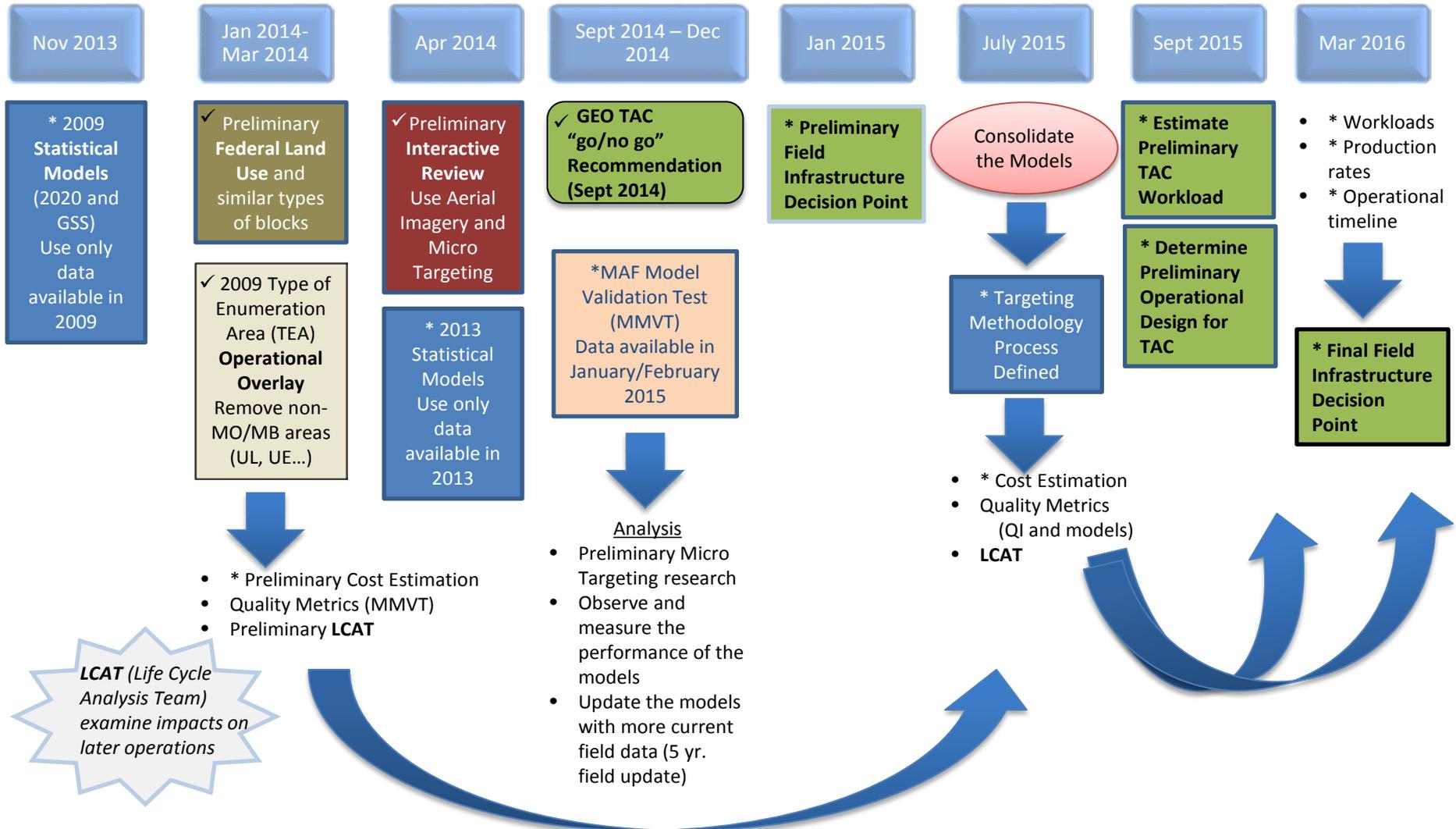
Cutpoint Scenarios

- 2020 RPO and GEO management have defined different methods for determining blocks that are “in” or “out” for a TAC operation
 - **Quality-driven** example: 0.5% undercoverage
 - **Data-driven** example: Let MEM team members interpret results for a suggested cutpoint
 - **Cost-driven** example: Canvass 10% of Housing Units

Frame Schedule

- Interaction with other programs like GEO/GSS
 - Incorporated into the 2020 schedule
- Statistical Model evolution (vintages and consolidation plan)
- Interaction with other 2020 Teams like cost estimation and field infrastructure
- Defined content of decision points

Frame Schedule



Incremental Savings from Targeted Address Canvassing (TAC) Strategies

Options in First Column Range from Most Conservative (Least TAC, at Top) to Most Liberal (Most TAC, at Bottom)

Blocks Included in Canvassing	Workload Outcomes				\$\$\$	Coverage Outcomes			
	# of Blocks Included*	%All Blocks Included	# of HUs Included	% AC HUs Included	Range of Potential Cost Reduction Estimates**	# of Adds Found	% Adds Found	# of Deletes Found	% Deletes Found
1. All Blocks Included		100%		100%			100%		100%
2. Remove Federal Lands Areas (e.g., National Parks, government installations)									
3. Apply 2., then keep only Mailout/Mailback areas (2009 TEA definitions)									
4. Apply 2. and 3., then apply Interactive GEO Review	Using imagery and other data to classify blocks. Can be used as input to the statistical models, as a method to determine canvassing status, or both.								
Statistical Model Outcomes. Each of the scenarios below is applied to the blocks remaining after the first three rows of the table are completed. Each of three models can be used: TRMAC statistical model, MAF Error Model statistical model, or the Interactive GEO Review.									
5. All methods agree to remove a block with 1+ adds									
6. MAF Error Model identifies blocks to remove based on 1+ adds or 1+ deletes									
7. 10% Targeting Range based on the statistical models				10%					

*Values are all in millions

**The cost reductions are based on the decennial cost model. The workload, productivity, and miles per case were varied for each scenario here.