

2020 Census Program Management Review

Administrative Records Modeling Update Thomas Mule Team Leader, Administrative Records Modeling Team

June 21, 2013

Outline

- Background on Nonresponse Followup (NRFU) and Administrative Record files
- Three Current Research Examples and Results

Results:

Field Work Reduction Summary

National Summaries of Population and Housing

Housing Unit-level Agreement of Population and Occupancy Status

Population and Occupied Housing summary for States and Counties

Analysis of Demographics: Age, Hispanic Origin and Race

- Research on Weighing Trade Off Between Cost Savings versus Accuracy
- Future Work

Research Questions

- How can we use Administrative Records to replace NRFU contacts?
- How many interview/contact attempts can be projected to be reduced?
- Can imputation methods be used to account for unresolved data due to curtailment?
- What happens to accuracy under different scenarios of NRFU curtailment and Administrative Records usage?
- How much does curtailing NRFU reduce cost?

2010 NRFU Execution Costs

Excludes Infrastructure Costs

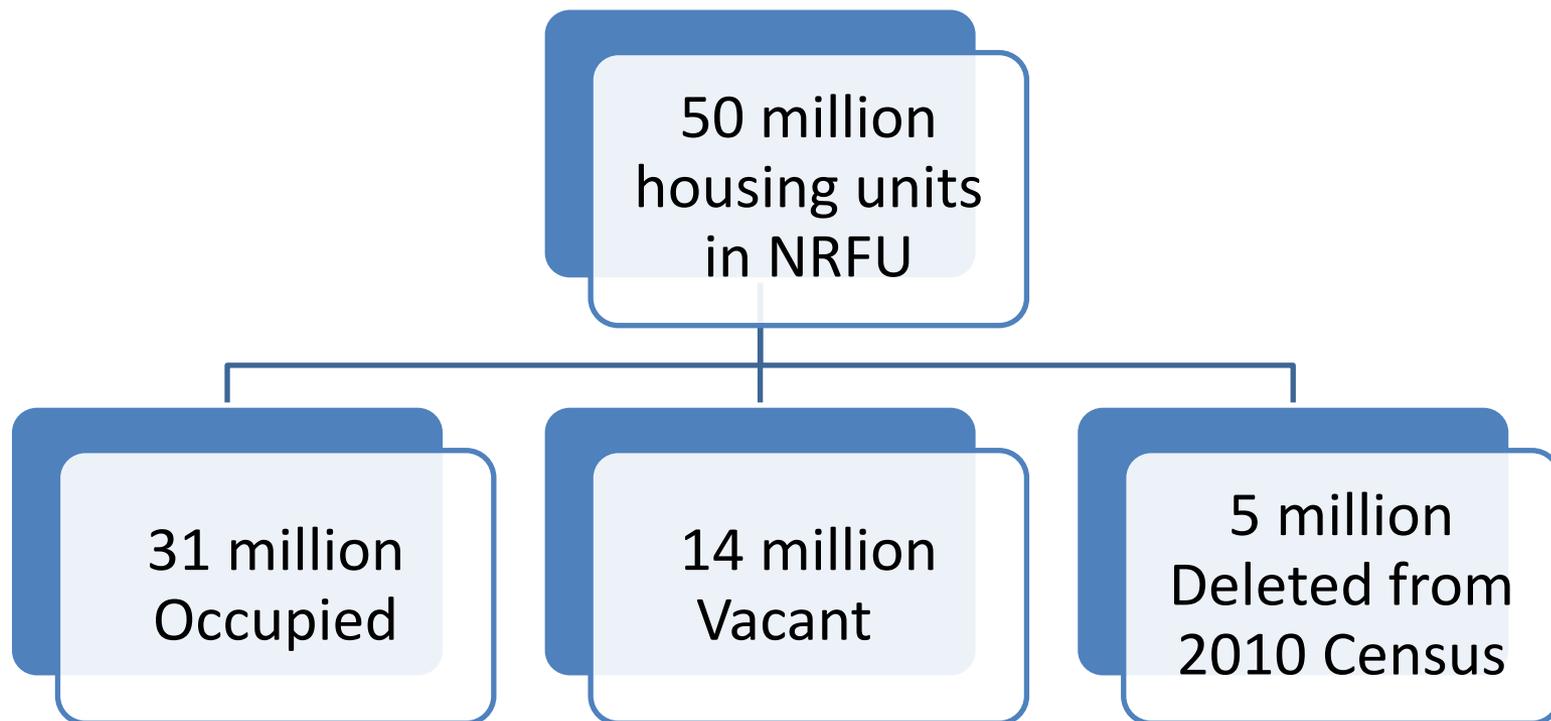
- Salary, training, benefits and mileage

Operation	Total Cost (in billions)	Average Cost per Unit
NRFU	1.59 billion	\$34
NRFU Reinterview	0.10 billion	\$106
Vacant Delete Check	0.28 billion	\$32
NRFU Residual	0.04 billion	\$58
Total	2.01 billion	\$35

- Additional overhead costs as well

NRFU Background

2010 NRFU Housing Unit Statuses



NRFU Background

2010 NRFU Field Operation Contacts

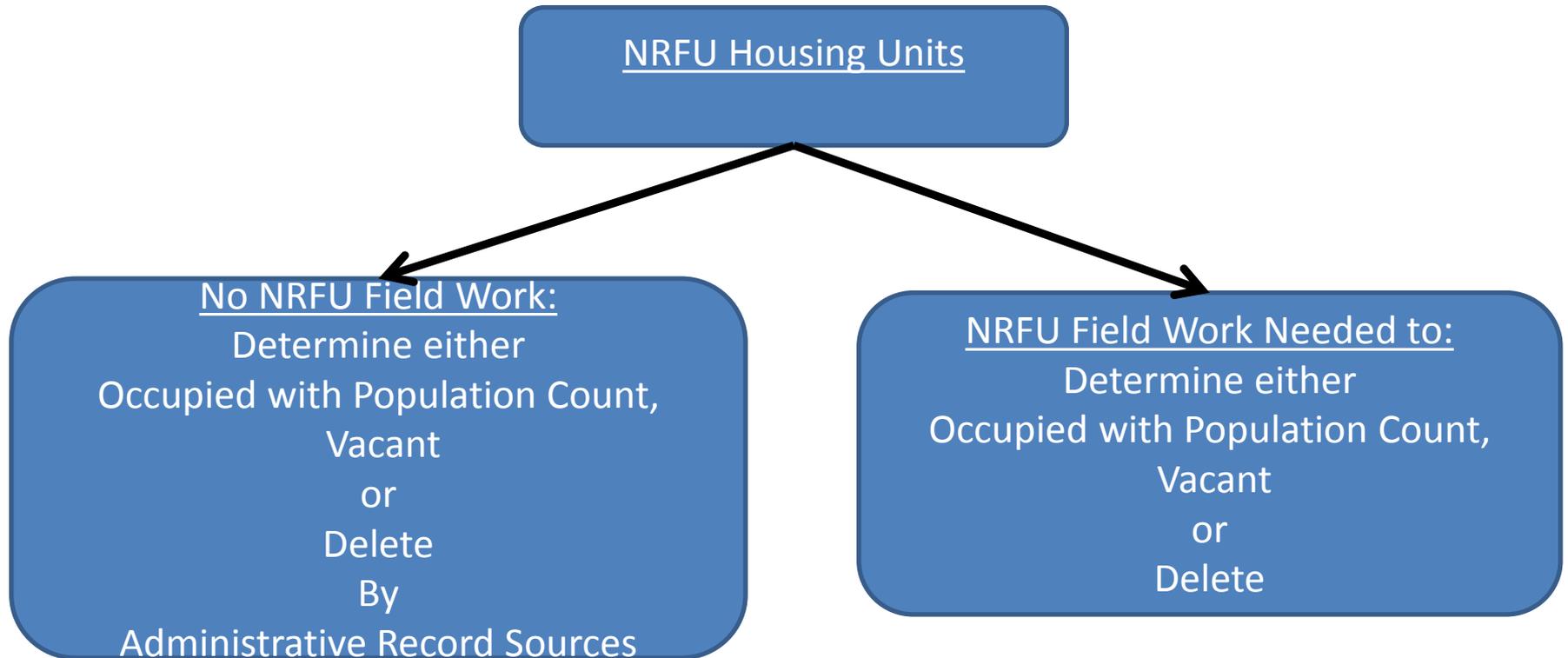
Number of Contact Attempts Made to NRFU Field Operation Housing Units

Number of Contact Attempts	Percent	Cumulative Percent
1	41%	41%
2	25%	66%
3	16%	82%
4	9%	91%
5	4%	95%
6	5%	100%

Source: Table 18 of 2010 Census NRFU Operations Assessment

Note: Rounded percentages shown. Unknown or 0 contacts excluded.

Background Conceptual Approach



Initial Approaches and Research

Researching the direct usage of administrative records for individual housing units

- Composite approach
 - Along the lines of the Census Match Study
 - Use different sets of files
- Rule-based approach
 - Hierarchy of rules about which records to use.
 - Simplest example is unit is occupied if a 1040 return filed for unit.

Administrative Record Sources

Building on:

- Statistical Administrative Records System (StARS)
- 2010 Census Match Study

- StARS files
 - 7 Federal files
- 2010 Census Match Study files
 - Additional 4 Federal and 9 Commercial Files
- Recent Files
 - Additional 1 Federal, 1 State and 1 Commercial File

Building and Using a Person Composite

Assigning person records from multiple administrative record sources to NRFU-eligible housing units

Steps

- Initial checks
- Determine how likely is it that the administrative record would be counted at this NRFU-eligible unit
- Determine which NRFU unit to assign a person if associated with multiple NRFU Units
- Rank housing units
- Determine threshold where composite population count is used instead of NRFU field work.

Building and Using a Person Composite

- Simple example for determining occupied units
- 101 Main Street did not respond to the 2020 Census
- What is a way that we can use multiple administrative sources to determine the population count for 101 Main Street?

Building and Using a Person Composite 2020 Challenge

Administrative Sources

101 Main Street

John Smith

Betty Smith

Steve Smith

Robert Smith

Building and Using a Person Composite 2020 Challenge

Administrative Sources

101 Main Street

John Smith →

Betty Smith →

Steve Smith →

Robert Smith →

Census

101 Main Street

Building and Using a Person Composite Initial Checks

- Did the person record get assigned a Protected Identification Key (PIK)?
- Is the person still alive?
- Was the person already counted by a census self-response?
- Do we have change of address information from the USPS?

Building and Using a Person Composite

Initial Check Example

- What happens if the Smith family was enumerated on a mail return at 500 Broad Street?

Administrative Sources

101 Main Street

~~John Smith~~

~~Betty Smith~~

~~Steve Smith~~

~~Robert Smith~~

Census

101 Main Street

Building and Using a Person Composite

Predicted Likelihood of Census Enumeration

Instead of being counted somewhere else, the Smith family passed the initial checks.

2020 Administrative Sources

101 Main Street

John Smith 

Betty Smith 

Steve Smith 

Robert Smith 

2020 Census

101 Main Street

Use 2010 Administrative records and census data to determine the likelihood that the Census would enumerate that person at the same address.

Building and Using a Person Composite

2010 Administrative Records and Census Results

Administrative Sources

101 Main Street

Alan Brown

Mary Brown

Joe Brown

Census

101 Main Street

Alan Brown

Mary Brown

Joe Brown

102 Main Street

Fred Gray

Wilma Gray

Peggy Gray

102 Main Street

Fred Gray

Wilma Gray

103 Main Street

Bill Thomas

103 Main Street

Building and Using a Person Composite

Do you want to exclude any 2010 Census NRFU Housing Units from the analysis?

Examples to remove:

- Count imputation
- Occupied status based on proxy response
- PIK not assigned to all census records
- Census records are duplicate enumerations
- Affirmative response to undercoverage or overcoverage questions
- Population count response not equal to number of data-defined records
- Census Followup (CFU) cases

Building and Using a Person Composite

2010 Analysis Removal Example

Administrative Sources

101 Main Street

Alan Brown

Mary Brown

Joe Brown

Census

101 Main Street

Alan Brown

Mary Brown

Joe Brown

102 Main Street

Fred Gray

Wilma Gray

Peggy Gray

102 Main Street

Fred Gray

Wilma Gray

~~103 Main Street~~

~~Bill Thomas~~

~~103 Main Street~~

~~Count Imputation~~

Building and Using a Person Composite

Predicted Likelihood of Census Enumeration

Assign predicted likelihood of census enumeration to each record

2020 Administrative Sources

101 Main Street

John Smith



Betty Smith



Steve Smith



Robert Smith



2020 Census

101 Main Street

Building a Person Composite

Predicted Likelihood of Census Enumeration

Prediction of how likely that the Census would enumerate that person at that NRFU-eligible address

Prediction take into account:

- Administrative record sources have that person at the same census address in 2010?
- Administrative record sources have that person at a different census address in 2010?
- Do sources indicate unit may have been unoccupied in 2010?
- How far from April 1st census day is the administrative source record?
- What are the demographic characteristics of the person?
- What are the Master Address File characteristics of the address?

Building and Using a Person Composite

Predicted Likelihood of Census Enumeration

Assign predicted likelihood of census enumeration to each record

Administrative Sources

101 Main Street

John Smith



Betty Smith



Steve Smith



Robert Smith



Census

101 Main Street

Building and Using a Person Composite

Multiple NRFU Housing Units

What if a person is found on administrative record files for multiple NRFU housing units?

- Person is placed at the housing unit with the highest predicted likelihood.

Robert Smith

- Associated with 2 NRFU-eligible Addresses: 101 Main Street and 700 Center Street
- 700 Center Street has a higher predicted likelihood of him being counted there.

Building and Using a Person Composite

Assigning Records to Housing Units

Administrative Records Composite for 101 Main Street

Robert Smith

- Associated with 2 NRFU-eligible Addresses: 101 Main Street and 700 Center Street
- 700 Center Street has a higher predicted likelihood of him being counted there.

Administrative Record Composite

101 Main Street

John Smith 

Betty Smith 

Steve Smith 

Census

101 Main Street

Building and Using a Person Composite

Ranking Housing Units

Composite

101 Main Street

John Smith 

Betty Smith 

Steve Smith 

Census

101 Main Street

Rank housing units by weakest likelihood

- Prediction with the smallest value.

Building and Using a Person Composite Ranking Housing Units

Order Housing Units by Minimum Person Predicted Value



Highest

Lowest

Building and Using a Person Composite Ranking Housing Units

Order Housing Units by Minimum Person Predicted Value



Do you use the 101 Main St. administrative records or do NRFU work?

Building and Using a Person Composite

Determining Threshold for Usage

Two Main Ways Shown Today

- Pre-determined magnitude of NRFU workload reduction
- Weighing tradeoff between cost savings and accuracy

Building and Using a Person Composite

Determine Threshold by Weighing Tradeoff

Potential Loss of Accuracy

- One measure is count agreement
- Used composite household when we shouldn't have

Potential Loss of Cost Savings

- Did not use composite household when we could have

Method to weigh these against each other

- Equal
- Emphasize one more

Current Research

Will show one rule-based and
two composite-based approaches

Showing preliminary research

Additional approaches are continuing to be
researched and developed

Current Research Example #1: IRS/UAA

Internal Revenue Service 1040 returns

- End of April
- 12 or fewer Protected Identification Keys (PIKs) assigned

United States Postal Service (USPS) Undeliverable-as-Addressed (UAA) for Initial Mailing

- Several reasons for being UAA
- Vacant reason only

Current Research Example #1: IRS/UAA

Remove From NRFU Fieldwork

1. IRS Return filed before April 30 with 12 or fewer people filed and no vacant UAA code - **Occupied (14.5 million units)**
2. No IRS Return filed before April 30 and vacant UAA code – **Vacant (5.6 million units)**

Total: 20 million housing units removed (40% removed from workload)

Include In NRFU Fieldwork

1. IRS Return filed before April 30 and vacant UAA code – **Conflicting Information**
2. IRS Return filed before April 30 with more than 12 people and no vacant UAA code – **Large administrative record count**
3. No IRS Return filed before April 30 and not vacant UAA code – **No information**

Current Research Example #2: Federal/State/Commercial Composite

- 10 Federal, 1 State and 8 Commercial Files
- Target occupied units
- Uses approach to equally weigh trade off between population count accuracy and cost-savings

- 11.8 million occupied units
- 24% reduction in workload

Current Research Example #3: Federal/State or Government Composite

Federal/State or Government Composite

- Use PIKs on 10 Federal and 1 State File
- Same initial checks
- Target occupied as described
- Target vacant units by similar approach of analyzing unoccupied units
- Emphasize reduction in costs by removing more housing units. What if we remove 14.5 million occupied and 5.6 million vacant just like example #1?
- 20 million total
- 40% reduction in workload

Definition: Primary Allocated Units

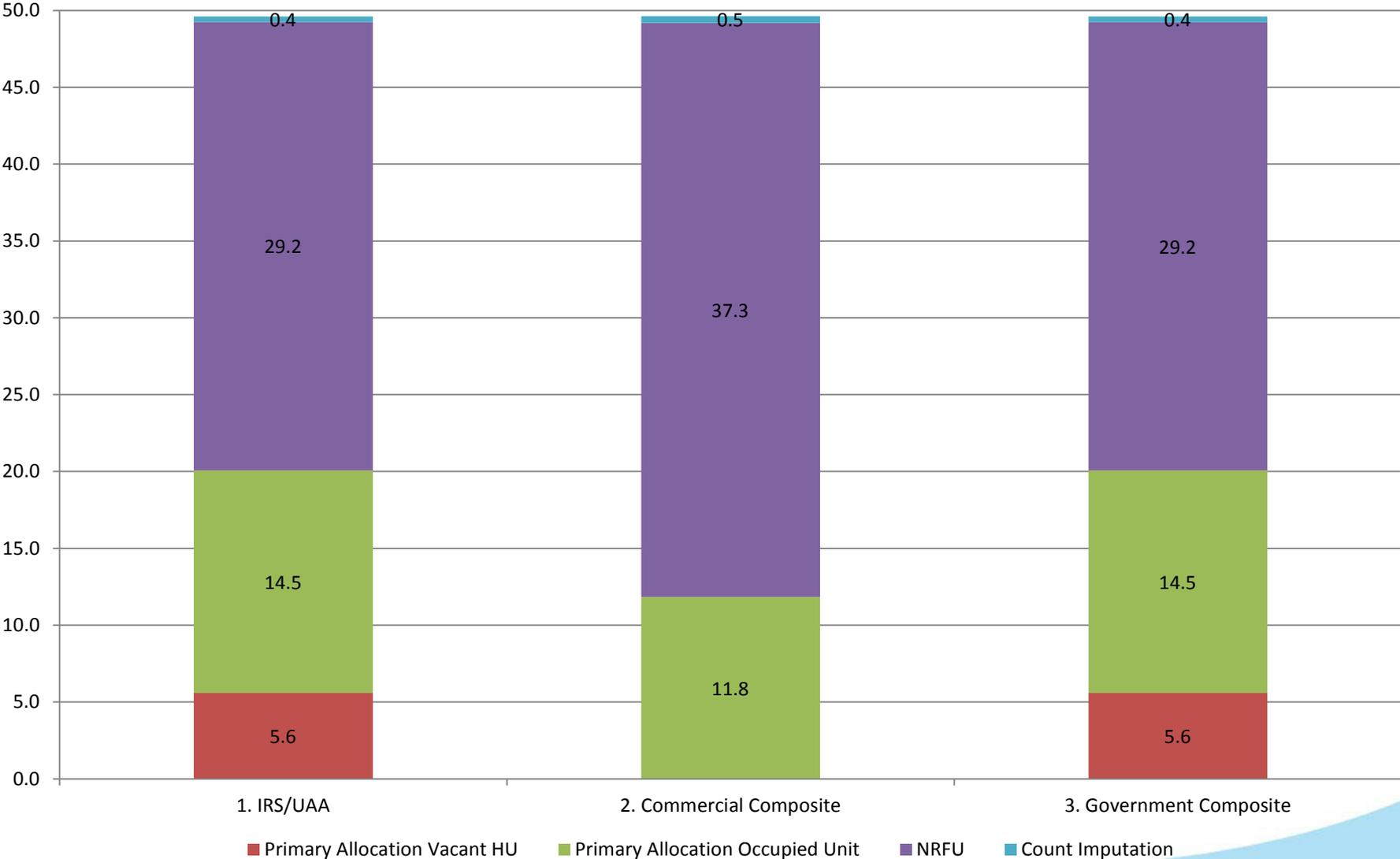
These are NRFU Units where administrative records were used to determine occupancy status (occupied, vacant or delete) and the population count if occupied.

Example 1: 20 million primary allocated

Example 2: 11.8 million primary allocated

Example 3: 20 million primary allocated

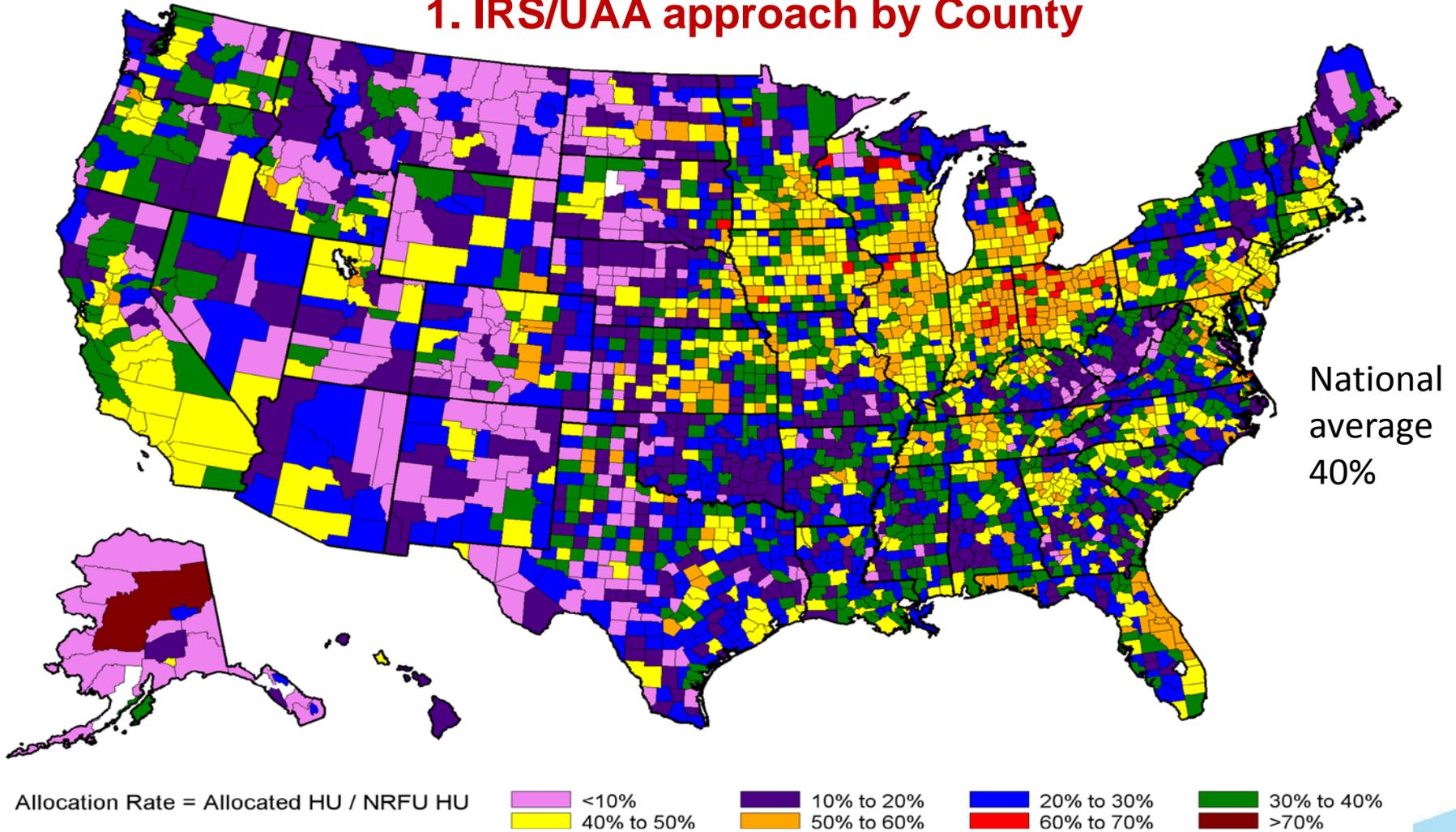
Summary of NRFU Workload Reduction



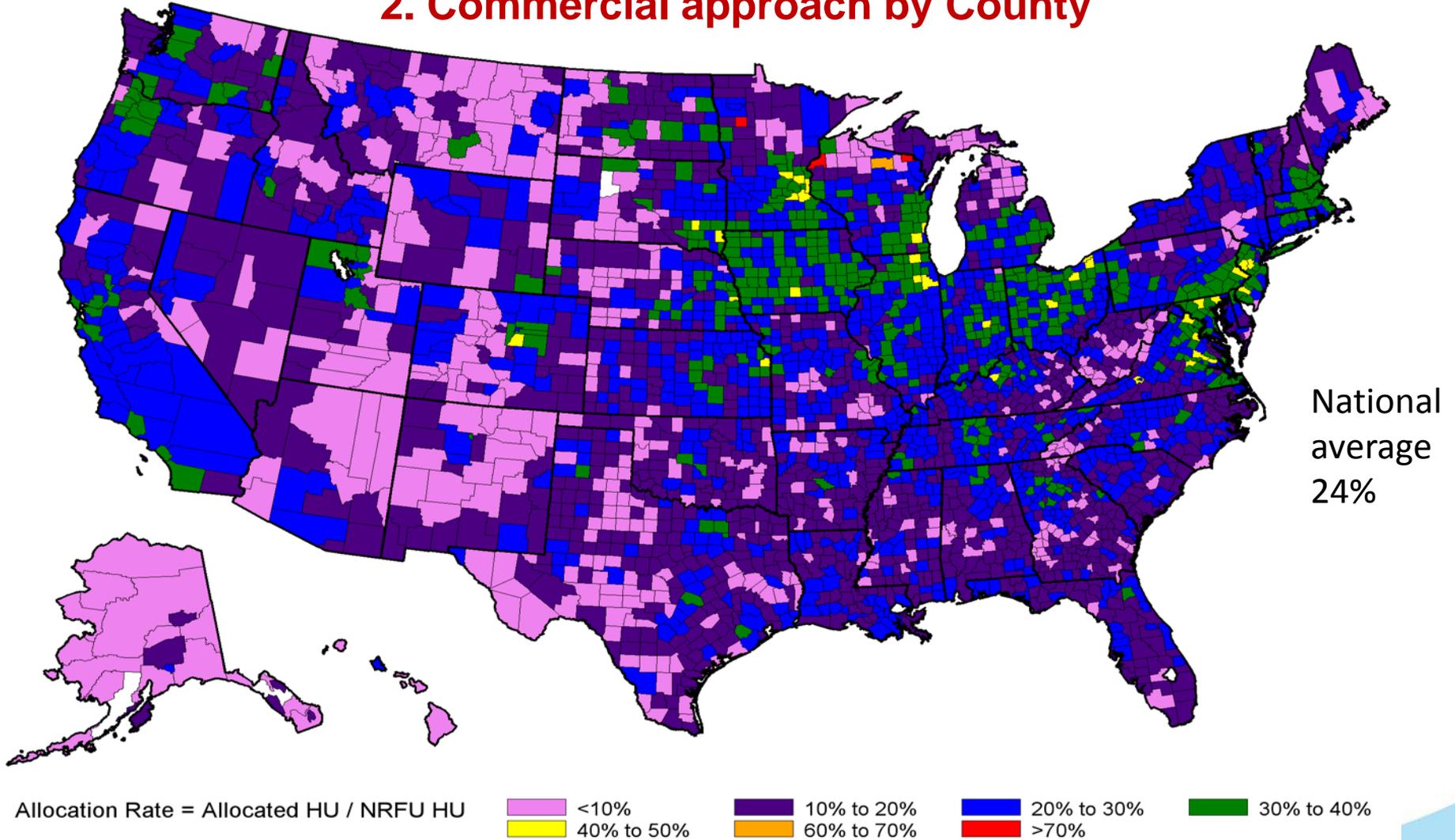
Reduction in Fieldwork by County

- Show maps for each of the 3 Examples

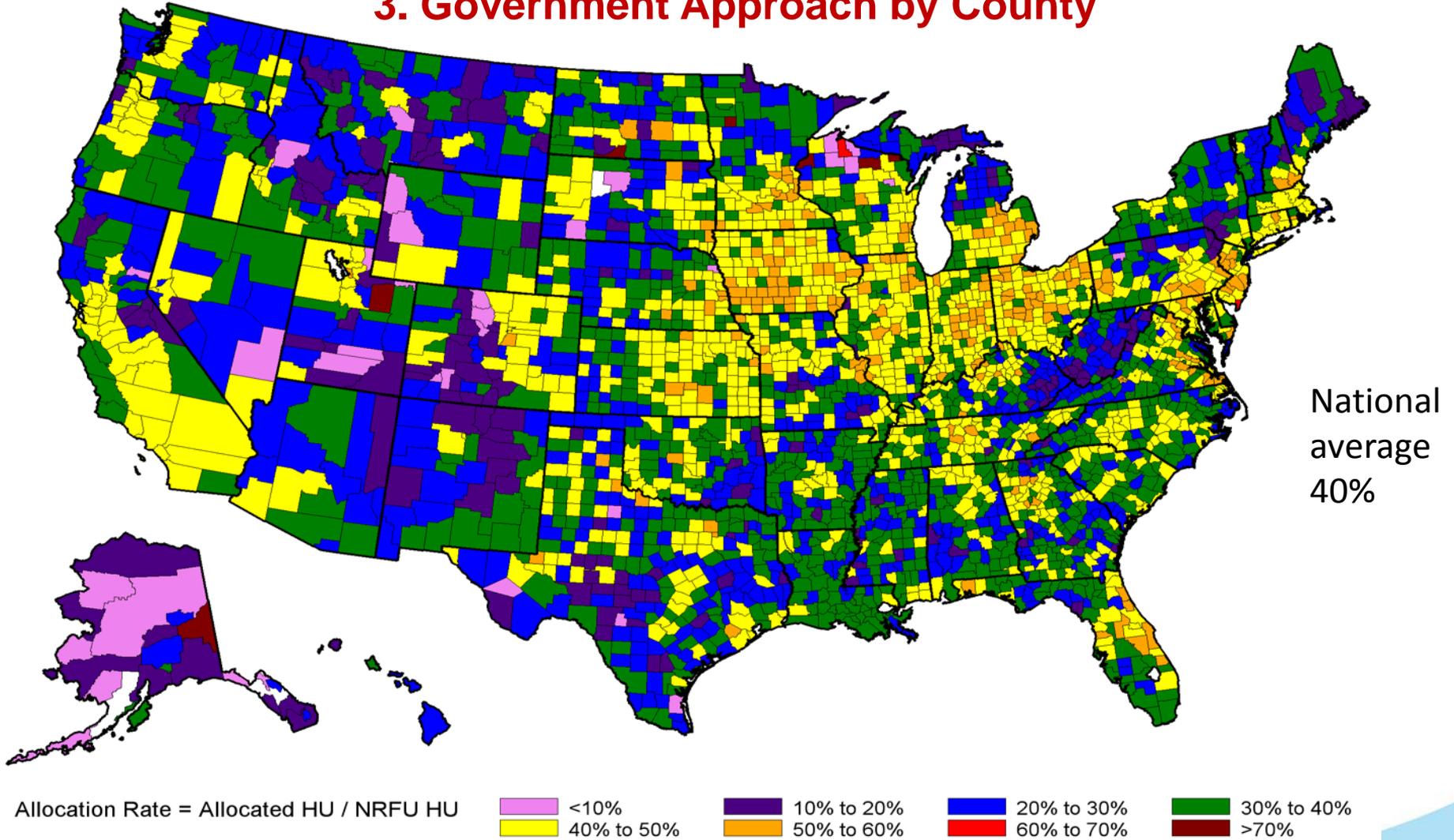
Percent of 2010 NRFU Housing Units Removed From Workload for 1. IRS/UAA approach by County



Percent of 2010 NRFU Housing Units Removed From Workload for 2. Commercial approach by County



Percent of 2010 NRFU Housing Units Removed From Workload for 3. Government Approach by County



Reduction in Fieldwork by County Summary

- Reductions can differ by approach
- Reductions are not uniform across country

Census Coverage Results for the Three Examples

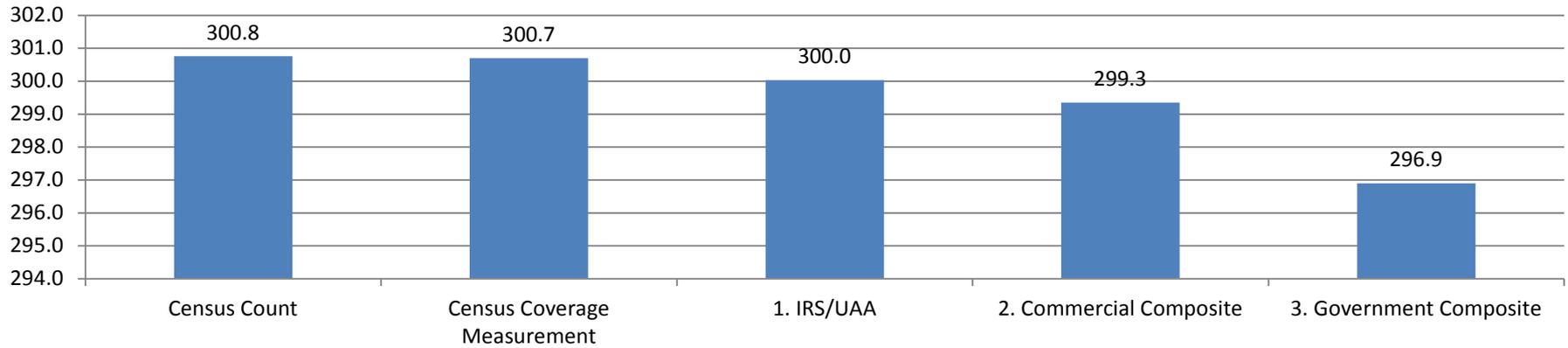
Please keep in mind

Changes to the 3 examples approaches can produce different census coverage results.

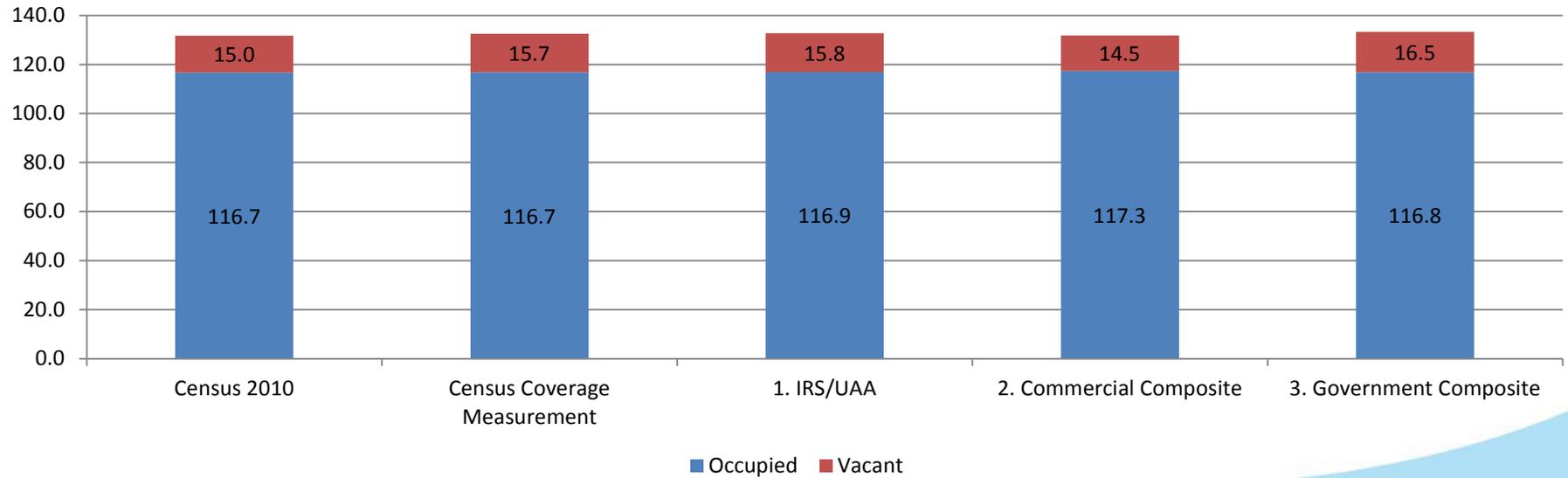
Examples:

- For #1 IRS/UAA, IRS households with over 12 are sent to the field. If the cutoff was 11, 10, 8 or 6 then you will see different results.
- For the #3 Federal, we are reducing the workload by 40% to target more occupied and vacant units. This is an example of implications reducing more workload on census quality.

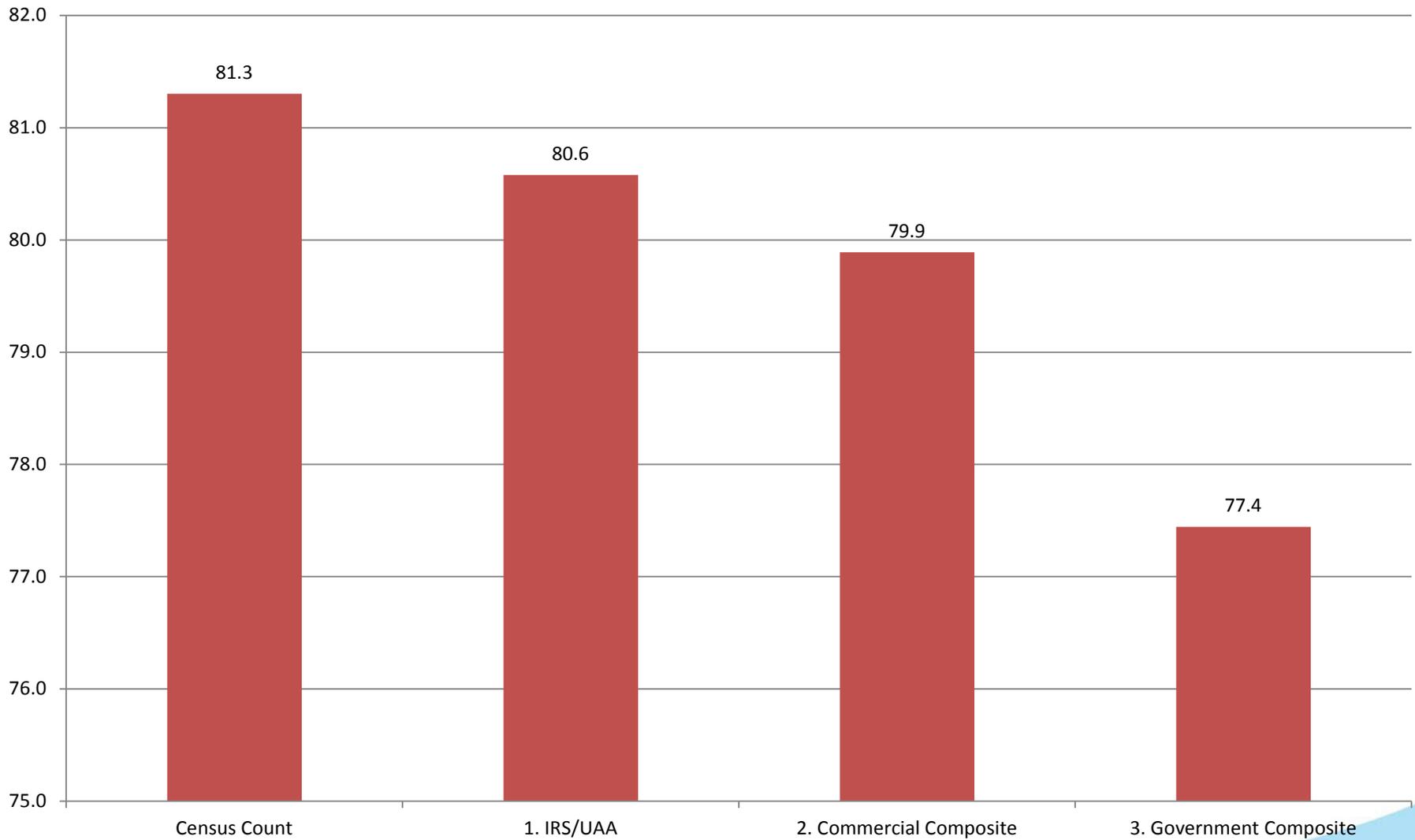
National Household Population



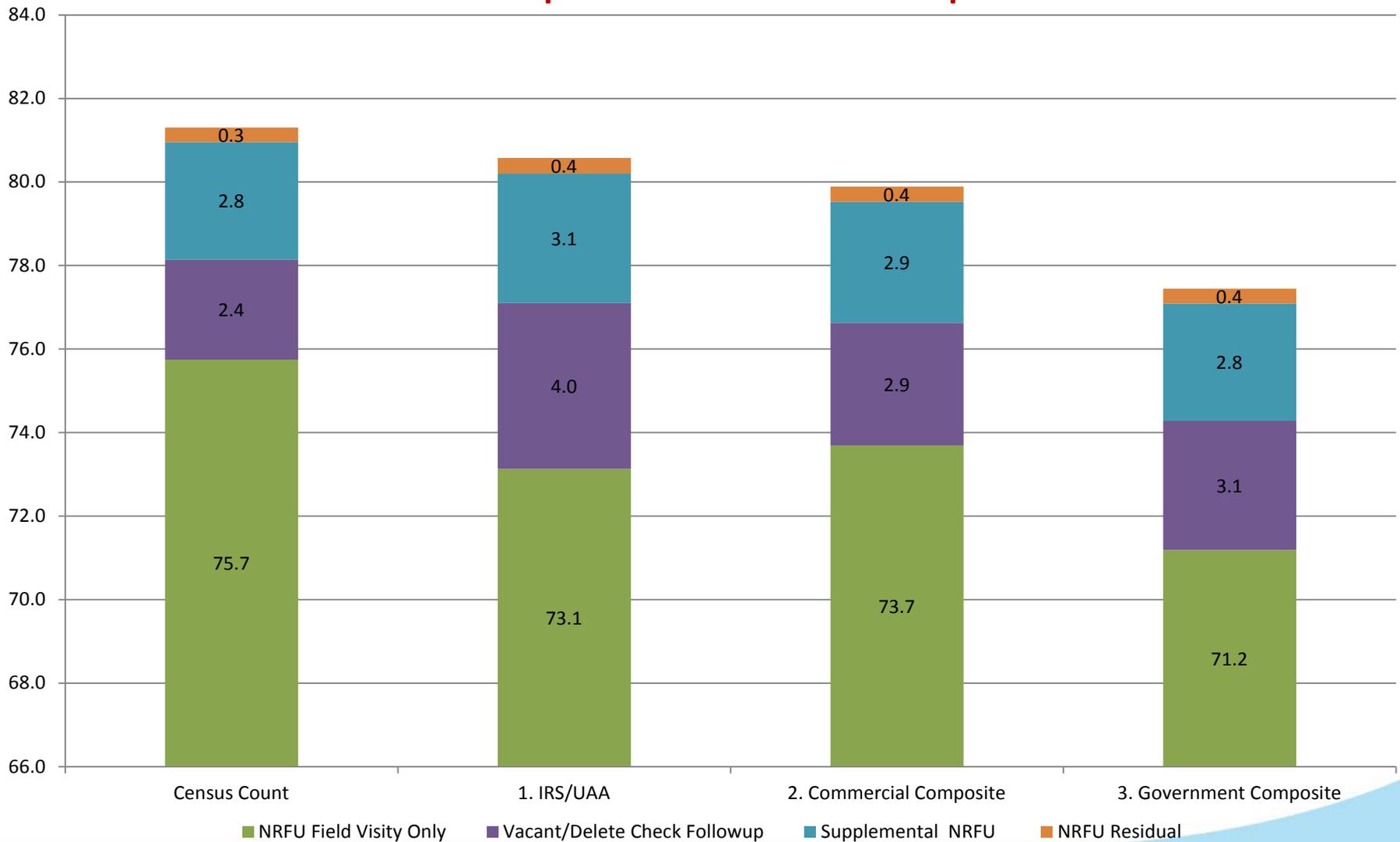
Comparison of Census Coverage by Housing Unit Status



NRFU Population by Example



NRFU Population by 2010 NRFU Field Operation Grouping: NRFU Field Operation and NRFU VDC Operation



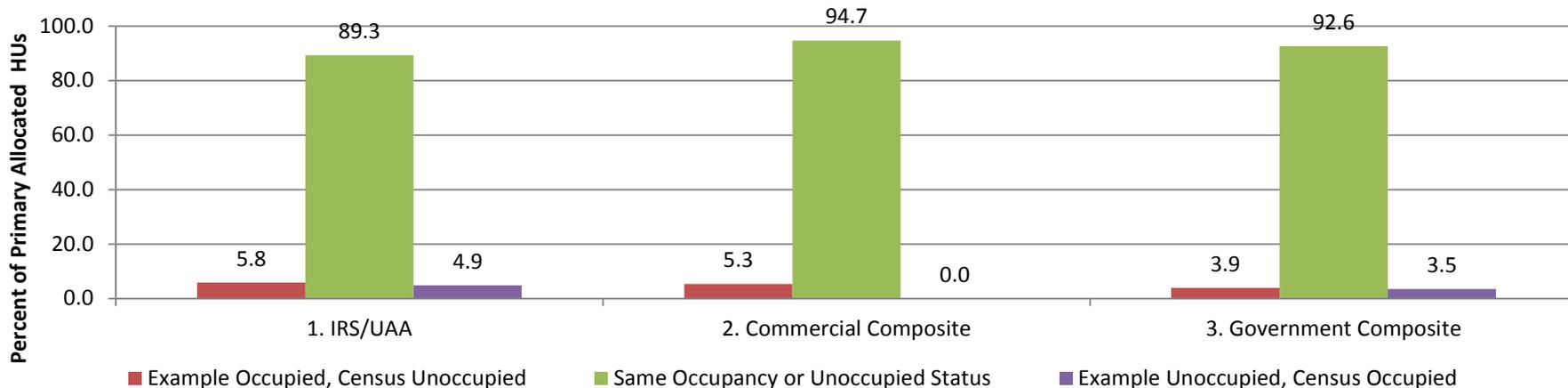
National Census Coverage Summary

- Different approaches can result in different coverage results

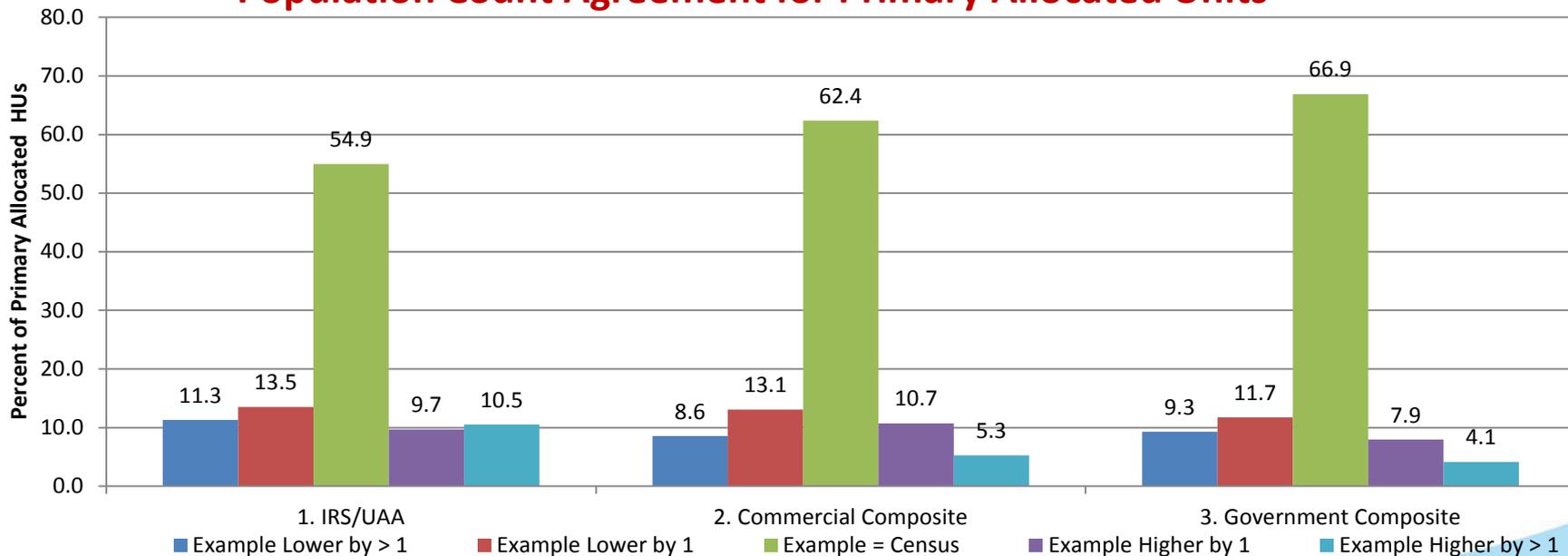
Housing Unit-level comparisons of Occupancy Status and Population

- Show some comparisons.
- Additional comparisons based on proxy status and others are also being investigated.

Occupied or Unoccupied Status Agreement for Primary Allocated Units



Population Count Agreement for Primary Allocated Units



Housing Unit-level comparisons

Administrative Records Usages may not have same result as 2010 Census result

Composite approaches have better agreement.

- Models have dependent variable of an Administrative Record person matching a census record in the same NRFU unit.
- IRS/UAA does not do that.

Percent Difference for States and Counties

- Distribution of Percent Difference

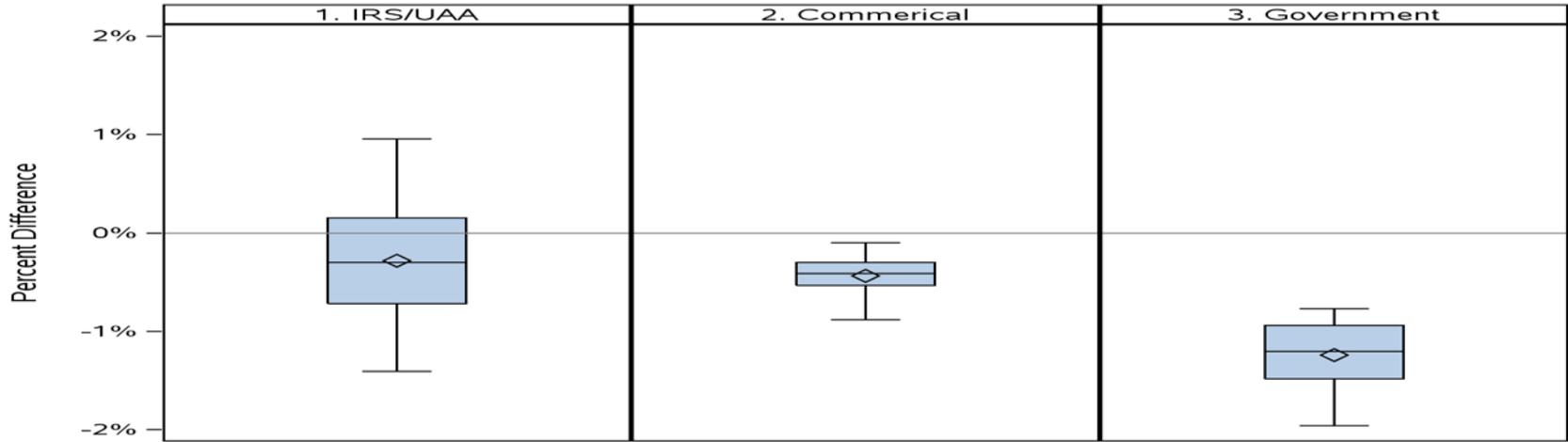
$$\text{Percent Difference} = \frac{\text{Example} - \text{Census}}{\text{Census}}$$

- Total Household Population
- Occupied Housing Units

- States
- Counties by Size Groupings

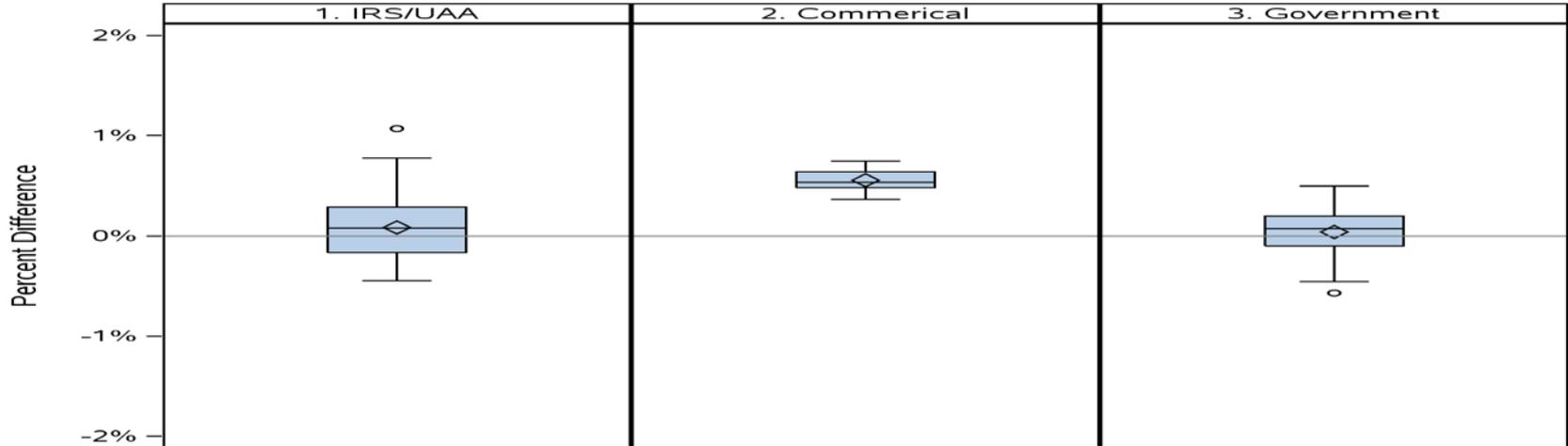
- Additional analysis being done for tracts
- Additional analysis will compare to CCM results and other evaluation sources

Box Plots of State Total Household Population Percent Difference by Example



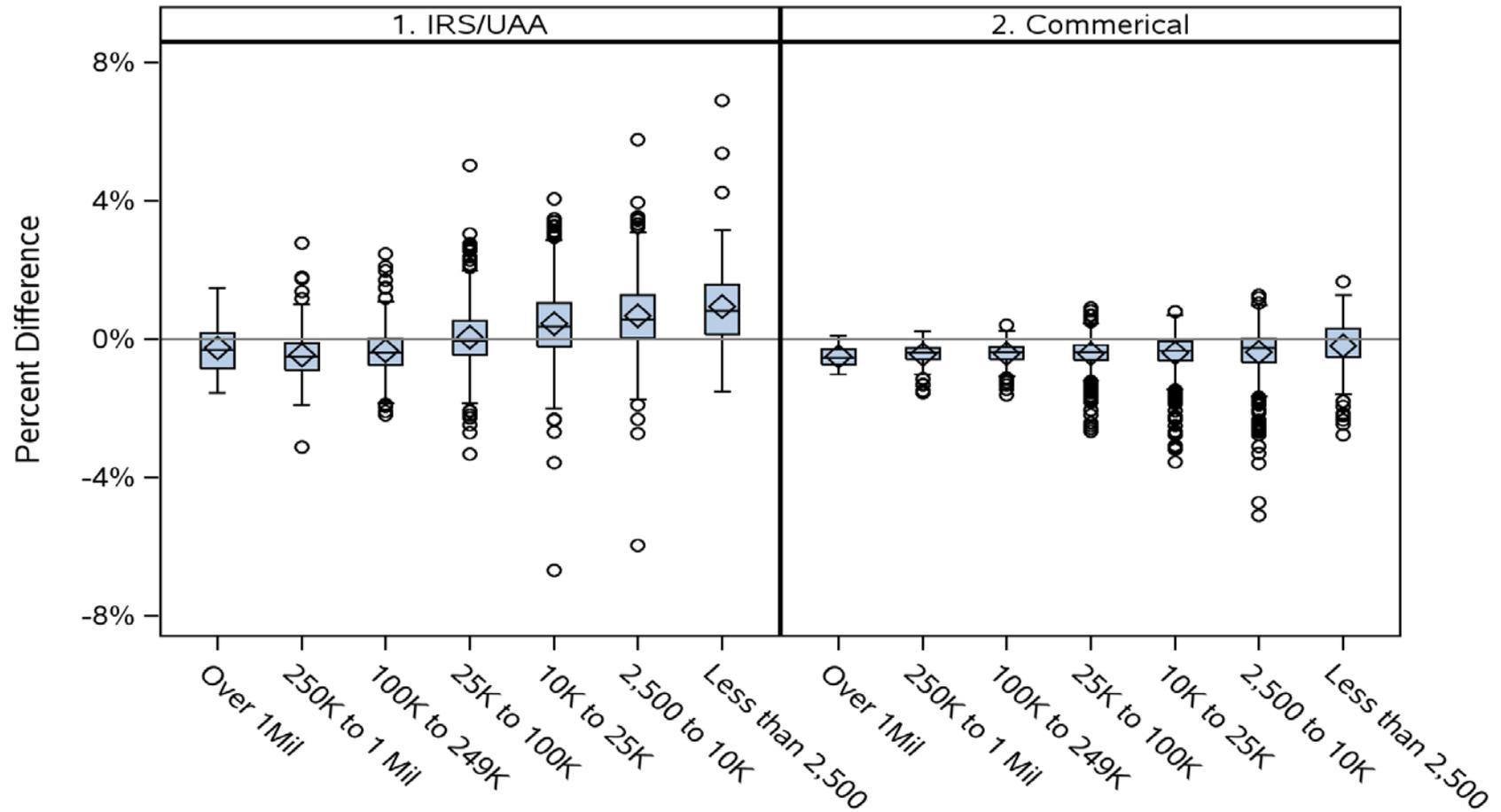
Percent Difference = $100 * (\text{Example} - \text{Census}) / (\text{Census})$.
 Positive percent difference indicates Example count larger than Census. Negative percent difference indicate Example count smaller than Census.

Box Plots of State Occupied Housing Unit Percent Difference by Example



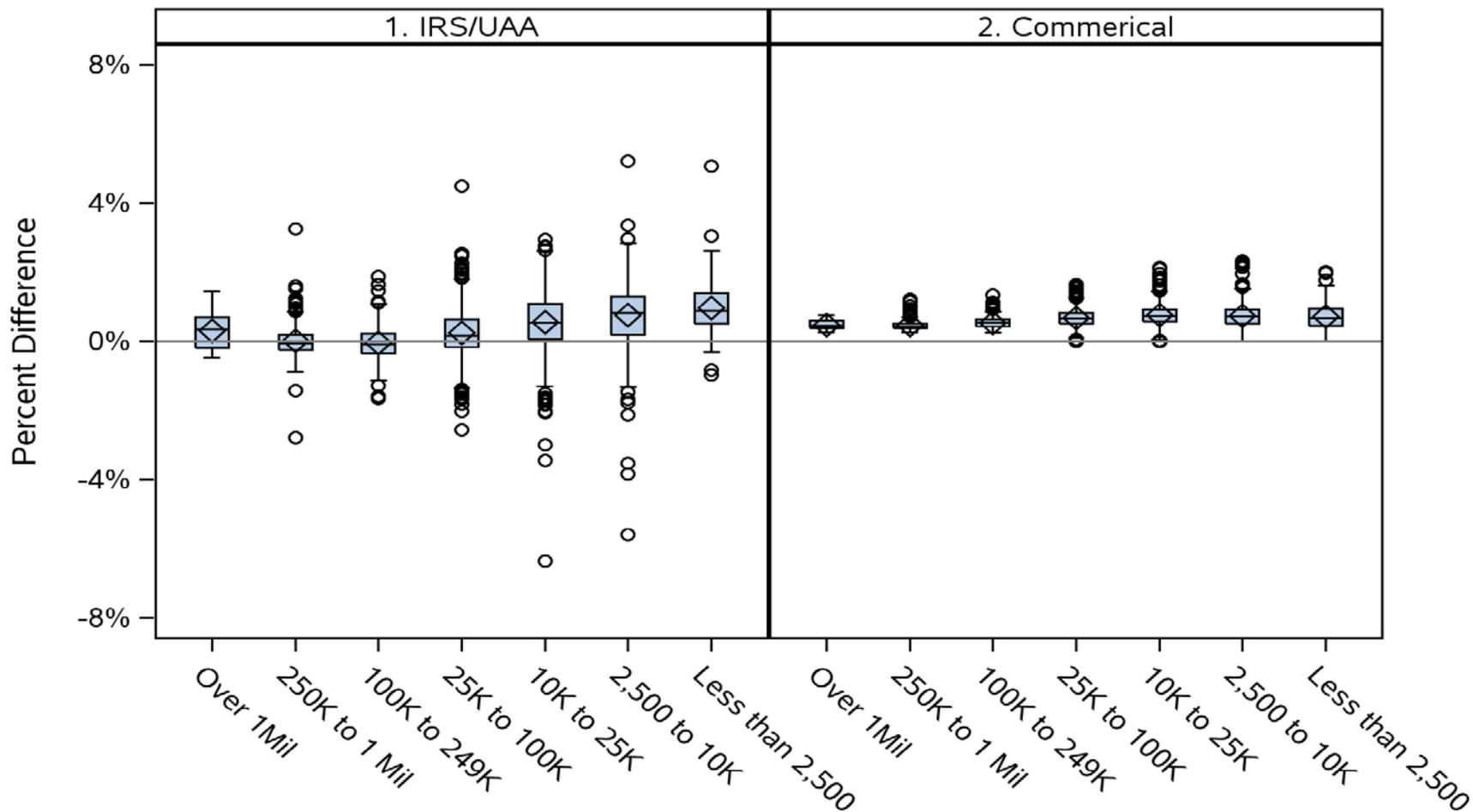
Percent Difference = $100 * (\text{Example} - \text{Census}) / (\text{Census})$.
 Positive percent difference indicates Example count larger than Census. Negative percent difference indicate Example count smaller than Census.

Box Plots of Total Household Population by County Size



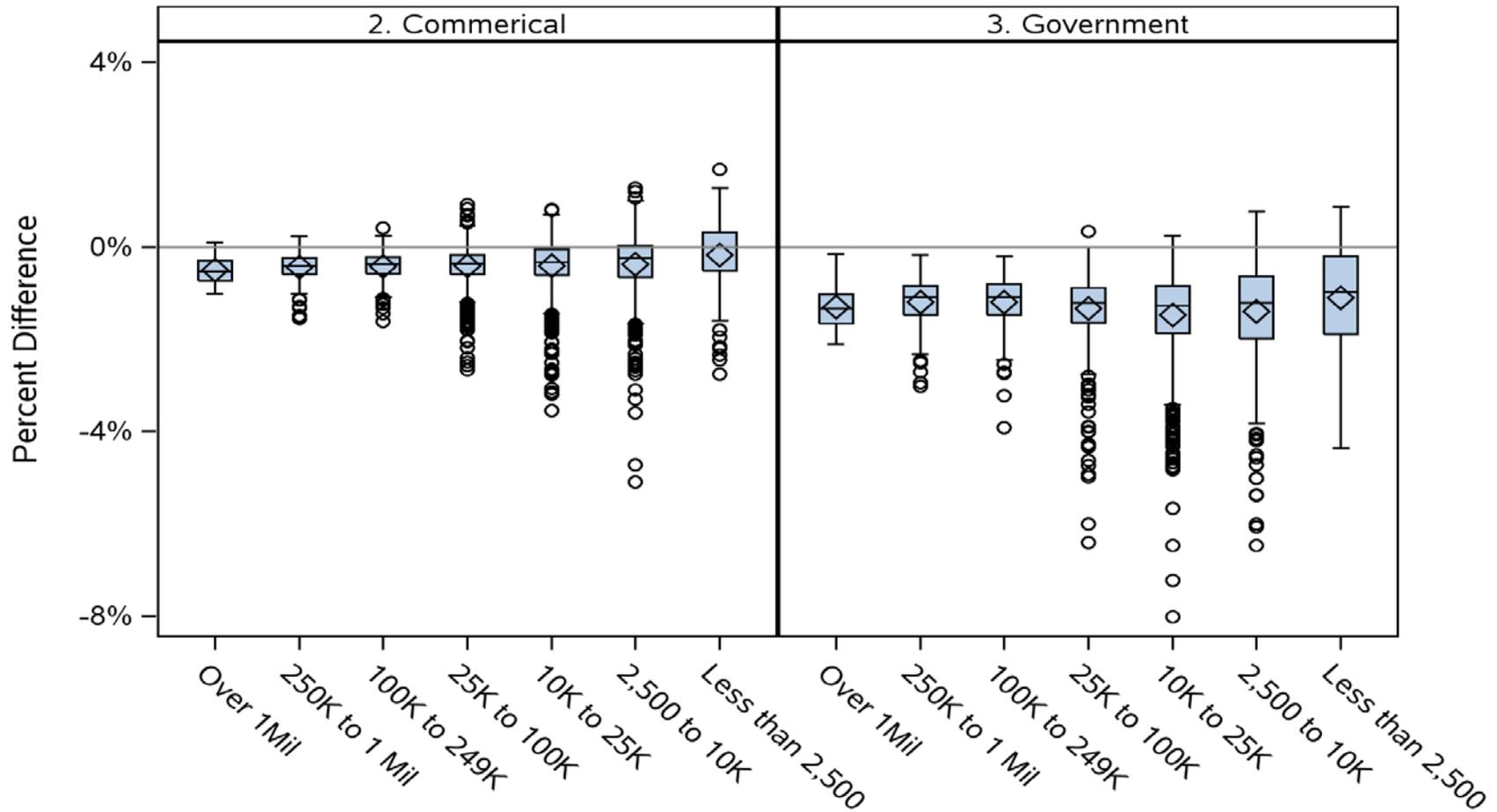
Percent Difference= $100 \times (\text{Example} - \text{Census}) / (\text{Census})$. Positive percent difference indicates Example count larger than Census count. Negative percent difference indicate Example count smaller than Census count.

Box Plots of Total Occupied Housing Units by County Size



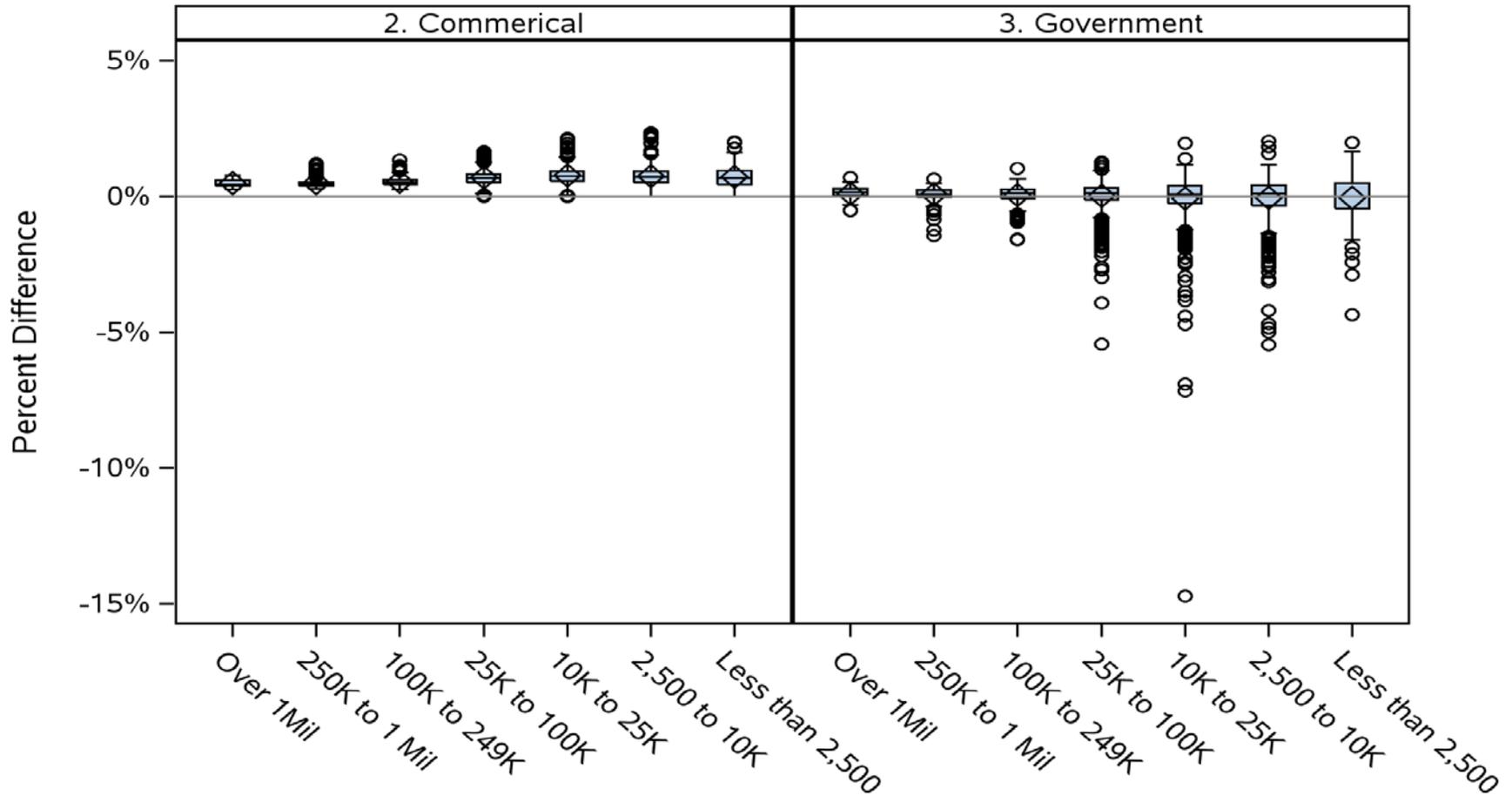
Percent Difference= $100 * (\text{Example} - \text{Census}) / (\text{Census})$. Positive percent difference indicates Example count larger than Census count. Negative percent difference indicate Example count smaller than Census count.

Box Plots of Total Household Population by County Size



Percent Difference= $100 * (\text{Example} - \text{Census}) / (\text{Census})$. Positive percent difference indicates Example count larger than Census count. Negative percent difference indicate Example count smaller than Census count.

Box Plots of Total Occupied Housing Units by County Size



Percent Difference= $100 * (\text{Example} - \text{Census}) / (\text{Census})$. Positive percent difference indicates Example count larger than Census count. Negative percent difference indicate Example count smaller than Census count.

Percent Differences for States and Counties Summary

- Different examples of usage can produce different results for states and counties
- County differences can vary by size

Analysis of Demographics

Working on characteristic imputation methods for NRFU simulations so that we can evaluate demographics

At this point, developed some evaluation methods using administrative records demographic characteristics and characteristic imputation of the Primary Allocated Occupied Units in an example

Age:

For Primary Allocated Units, how close is the age distribution to the 2010 Census results?

IRS/UAA Example 1

- For the 14.5 million occupied units, 41.5 million records have a PIK.
- For the 40 million records, we get age from Numident or other AR files.
- Calculate the distribution of age using these 40 million IRS/UAA records.

- For the SAME 14.5 million units, we get the 2010 Census results.
- 40 million total records in 2010 with edited and imputed age.
- Calculate the distribution of age using these 40 million Census 2010 records.

- Compare the distributions to see if we are seeing similar or different age distributions.

- Did this for all three examples using their primary allocated occupied units.

AGE: For Primary Allocated Units, is the distribution the same or different as compared to 2010 Census?

Comparison of Age Percentages for the Three Example Approaches to Census for Primary Allocated Units

	1. IRS/UAA 14.5 million occupied units			2. Federal/State/Commercial Composite 11.8 million occupied units			3. Federal/State Composite 14.5 million occupied units		
Age group	Census 2010 (40 million) (Percent)	Simulation (40 million) (Percent)	Increase or Decrease in Percentage Points vs. Census	Census 2010 (30 million) (Percent)	Simulation (28 million) (Percent)	Increase or Decrease in Percentage Points vs. Census	Census 2010 (37 million) (Percent)	Simulation (35 million) (Percent)	Increase or Decrease in Percentage Points vs. Census
0-9	18.5	21.0	2.5	16.8	16.3	-0.5	17.2	16.9	-0.3
10-17	13.5	14.2	0.7	13.1	12.1	-0.9	13.1	12.3	-0.8
18-29	21.5	22.6	1.1	16.6	16.3	-0.3	17.5	17.3	-0.2
30-49	30.0	28.5	-1.5	31.2	32.6	1.4	31.0	32.1	1.1
50+	16.5	13.7	-2.8	22.3	22.8	0.4	21.2	21.4	0.2
Total	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0

Census Coverage for Age

Distribution of Age in Primary Allocated Housing Units for different approaches can differ from the Census 2010 distributions

Hispanic Origin: For Primary Allocated Units, what do we see for Hispanic Origin as compared to the 2010 Census results?

IRS/UAA Example 1

- For the 14.5 million occupied units, 41.5 million person records
- Obtain Hispanic Origin of the record by either of two ways:
- First we use a Previous Census Bureau Title 13 response if available based on PIK assignments
- If Previous is not available, we then used a quick Hispanic Origin imputation using IRS 1040 Primary Filer Last Name
- Working on more complete methods for Hispanic origin imputation
- Examine the Hispanic Origin using these 40 million IRS/UAA records

- For the SAME 14.5 million units, we get the 2010 Census results
- 40 million total records in 2010 with edited and imputed Hispanic Origin
- Calculate the distribution of Hispanic using these 40 million Census 2010 records

- Compare the distributions to see how the distributions compare

- We have done this for the 1. IRS/UAA and 2. Federal/State/Commercial example

Hispanic Origin

Preliminary Results for Hispanic Origin of Primary Allocated Units

		1. IRS/UAA 14.5 million occupied units People (millions) % Hispanic		2. Federal/State/Commercial 11.8 million occupied units People (millions) % Hispanic		3. Federal/State 14.5 million occupied units People (millions) % Hispanic	
Census 2010 results in the Primary Allocated Units		40.2	23%	29.6	16%	37.0	17%
Approach Results	Past Title 13 Response (1) Primary Filer Surname Imputation (2)	23.5	16%	18.6	10%	22.4	11%
	Subtotal (1+2)	40.5	20%	27.1	12%	33.3	14%
	Primary Filer Imputation Not Possible Under This Methodology (3)	1.0	n/a	1.0	n/a	1.4	n/a
	Total (1+2+3)	41.5	n/a	28.2	n/a	34.6	n/a

Note: Primary Filer Surname Imputation is being used to get a quick estimate of Hispanic Origin

Preliminary Results for Hispanic Origin

- Potential difference as compared to Census
- Examine using additional files to help
- Continue development of characteristic imputation methods for missing cases by DSSD, CSRM and other team members

Race Analysis: Missing Data

Percent of Race Information Available for Primary Allocation Records

	1. IRS/UAA 14.5 million occupied units	2. Federal/State/Commercial 11.8 million occupied units	3. Federal/State 14.5 million occupied units
Past Census Title 13 Response	57%	66%	65%
Other Administrative Records	15%	14%	14%
None Available	29%	20%	21%

Implications for Race

- Potential differences as compared to Census due to large amount missing data
- Utilize additional sources to compensate
- Continue development of characteristic imputation methods for missing cases

Research on Weighing Tradeoff between Cost Savings and Accuracy

Potential Loss of Accuracy

- One measure is count agreement
- Used composite when it did not agree with census count

Potential Loss of Cost Savings

- Did not use composite household when it agreed with census count

Method to weigh these against each other

- Equal
- Emphasize one more

Research on Weighing Tradeoff between Cost Savings and Accuracy

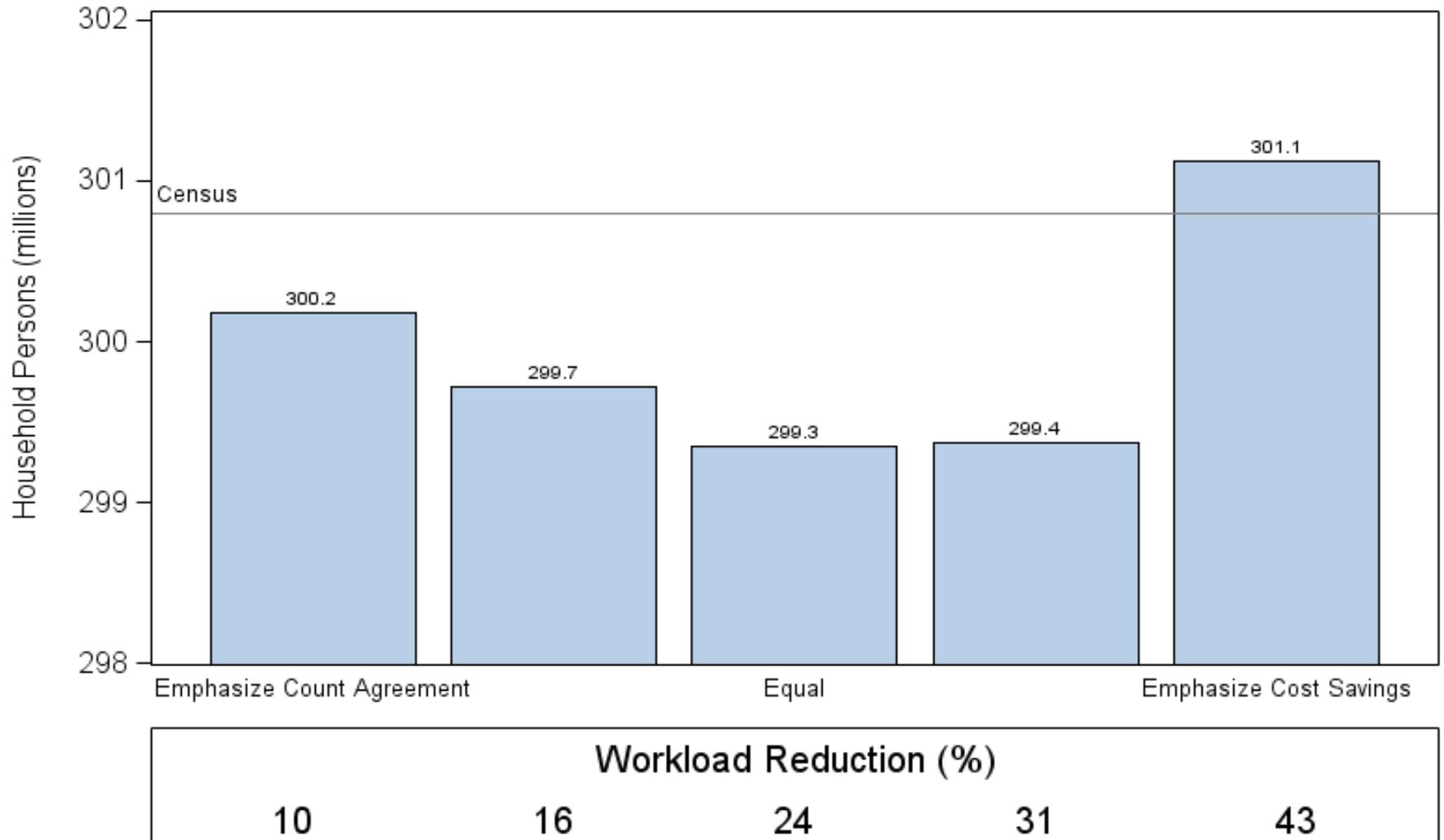
Minimize Potential Loss of Cost Savings

Minimize Potential Loss of Accuracy Measured by Count Agreement

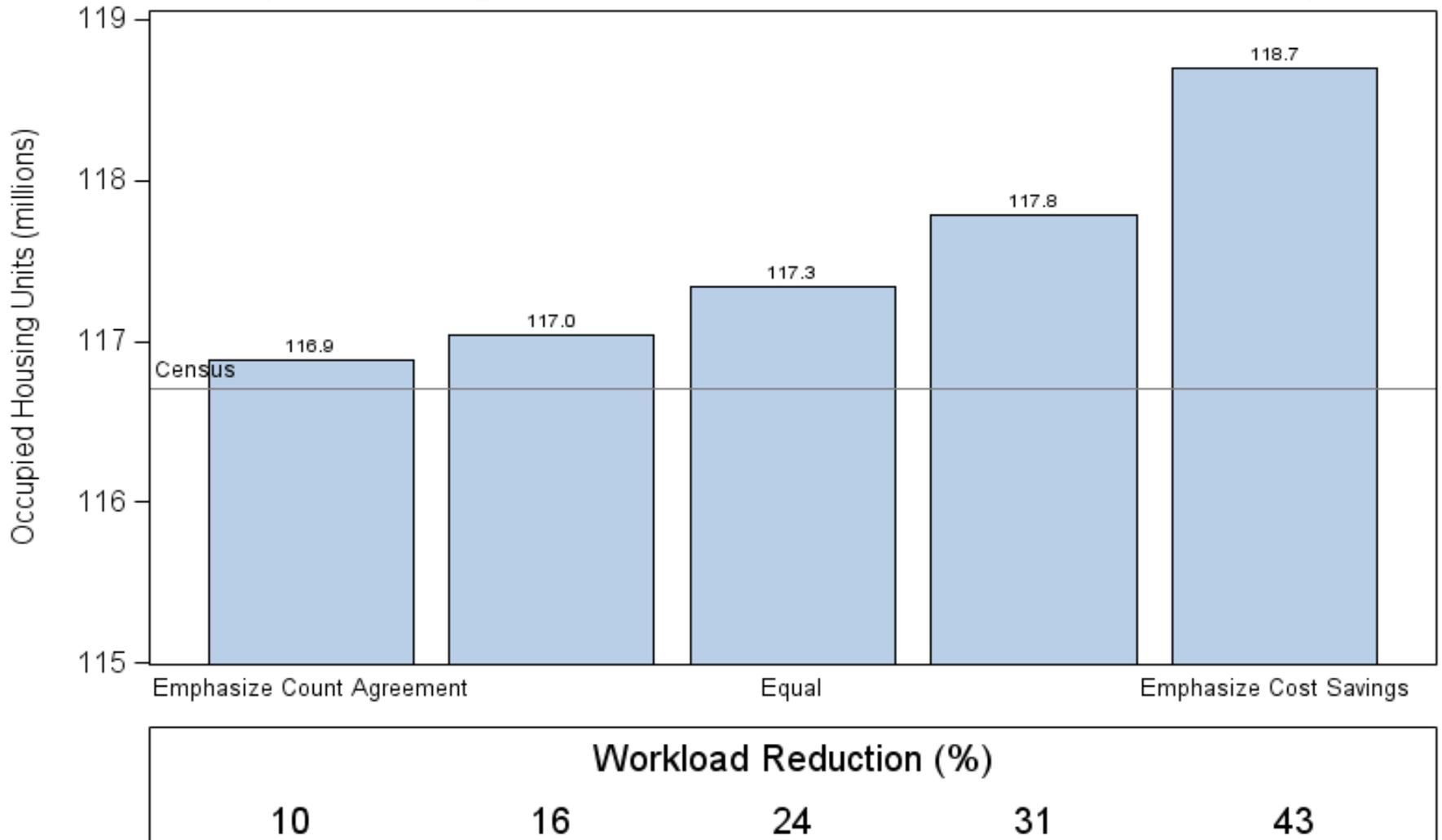
Examples weigh these against each other using the Commercial composite

- Equal
- Emphasize one more

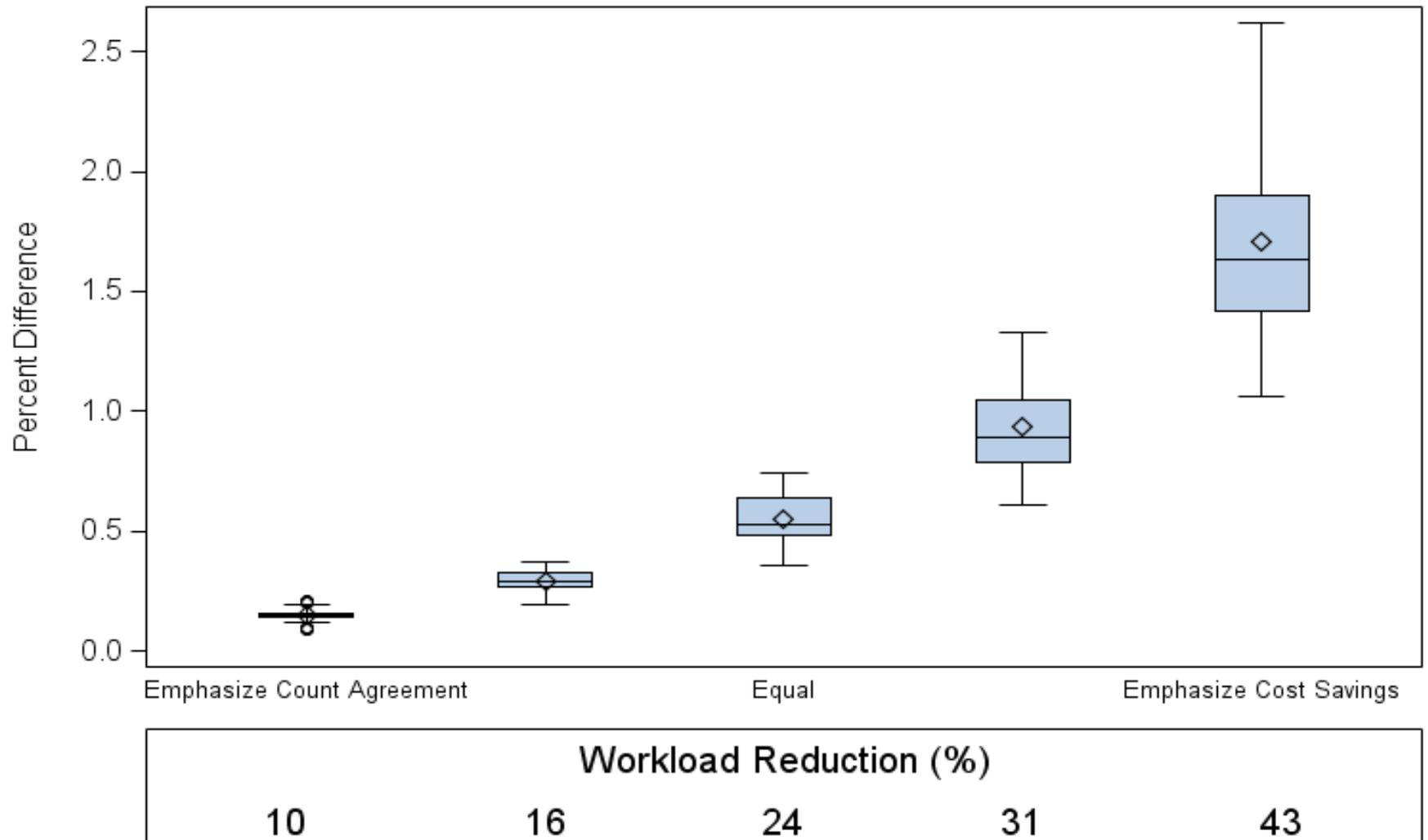
Total Household Persons for Five Cutoffs of the Commercial Composite



Occupied Housing Units for Five Cutoffs of the Commercial Composite



Distribution of State Percent Differences of Occupied Units Using the Commercial Composite



Future Work

- Integrate with NRFU contact strategy research, field reengineering and others
- Understanding similarities and differences between examples
- Enhance/develop characteristic imputation methods for Hispanic Origin, Race and other characteristics
- Estimate coverage and cost savings of different approaches
- Continue research of current or possibly acquire new administrative records sources

2020 Census Program Management Review

Thank You

Administrative Record Sources

StARS Federal Files

Internal Revenue Service

- Individual Income Tax Returns 1040
- Information Returns 1099 (W2, 1099, etc.)

Administrative Record Sources

StARS Federal Files (Cont.)

Department of Housing and Urban Development (HUD)

- Tenant Rental Assistance Certification System (TRACS) file
- Public and Housing Information Center (PIC) file

Administrative Record Sources

StARS Federal Files (Cont.)

Center for Medicare and Medicaid Services (CMS)

- Medicare Enrollment Database (MEDDB)

Indian Health Service (IHS)

- Patient Registration System

Selective Service System (SSS)

- Registration System

Administrative Record Sources

2010 Census Match Study Federal Files

Department of Housing and Urban Development (HUD)

- Computerized Homes Underwriting Management System (CHUMS)

Social Security Administration (SSA)

- Supplemental Security Record (SSR) File
- Death Master File

United States Postal Service (USPS)

- National Change of Address (NCOA) File

Administrative Record Sources

2010 Census Match Study Commercial Files

Experian

- In-Source (INS) File
- End-Dated Records (EDR) File

Targus

- Federal Consumer
- Pure Wireless
- National Address File

Administrative Record Sources

2010 Census Match Study Commercial Files (Cont.)

Veteran Service Group of Illinois (VSGI)

- Name and Address Resource Consumer file
- TrackerPlus file

InfoUSA

Melissa Data Base Source

Administrative Record Sources

Additional Sources

New York Supplemental Nutrition Assistance Program (SNAP)

USPS Undeliverable As Addressed (UAA)

Corelogic