



# **2020 Census Program Management Review**

## **2013 National Census Contact Test**

**Dave Sheppard, Contact Frame Team Lead**  
**Frank McPhillips, Non-ID Processing Team Lead**

September 24, 2013



# Outline for Today's Presentation

---

- Purpose of the 2013 National Census Contact Test
- Definitions
  - Supplemental Contact Frame
  - Non-ID Processing
- Results from the Test
- Next Steps

# 2013 National Census Contact Test (NCCT)

## Purposes and Definitions

# 2013 NCCT: Purpose

---

- Contact Frame Quality
  - Evaluate the quality of phone and email contact information acquired from commercial sources
- Improving Non-ID Processing
  - Test proposed enhancements to automated processing of census responses lacking a preassigned census identification number



# Supplemental Contact Frame

## What Is It?

---

- Built from commercial sources
- Provides additional contact information
- Supplements our address frame



# Supplemental Contact Frame

## How Can We Use It?

---

- Encourage self-response & aid nonresponse followup
- First step: link it with our address frame
  - The address frame for the Decennial Census is the Master Address File (MAF)
  - Phone numbers and emails in the vendor data were linked to MAF records wherever the vendor address matched a MAF address



# Supplemental Contact Frame

## Off to a good start ...

---

- The contact frame has been built and now includes data from five commercial sources
  - Phone numbers: Both landline and cell
  - Email addresses
- Four years of data – 2010-2013
  - Some sources are only from 2010, others from multiple years
  - Some sources are obtained annually, some are obtained quarterly

# Supplemental Contact Frame Initial Metrics – Phone Numbers

	#	universe	%
Only 2010 Census HU MAFIDs with 1+ phone numbers	105.4 million	131.7 million	80%
Only 2010 Census NRFU HU MAFIDs with 1+ phone numbers	30.4 million	49.8 million	61%

- In addition, 42% of all 2010 Census respondent provided phone numbers are in the contact frame at the same address
  - Baseline: From files with a vintage of 2010 only

# What the Contact Frame is: Initial Metrics – Phone Numbers (Cont.)

- Regionally, % of addresses with 1+ phones

<b>Midwest</b>	<b>85%</b>	<b>Northeast</b>	<b>76%</b>
West	78%	South	80%

- By state, % of addresses with 1+ phones

<b>Highest Coverage</b>	<b>Iowa</b>	<b>88%</b>
Lowest Coverage	Alaska	43%

Using 2010 sources only

# Supplemental Contact Frame

## Phones per Address

- The number of phone numbers associated with a housing unit = 2.4

Phones per address	
Phone Numbers	% of addresses
0	20 %
1	21%
2	21%
3	15%
4	10%
5+	14%

Using 2010 sources only

# Supplemental Contact Frame

## Addresses per Phone

- The number of housing units associated with a phone number = 1.4

Addresses per Phone	
Addresses/phone	% of addresses
1	70%
2	22%
3	6%
4	2%
5+	1%

Using 2010 sources only



# Non-ID Processing Overview

---

- Compares addresses from responses lacking a Census ID to the Census Bureau address inventory
- Opportunity to link response data to existing MAF record and/or assign to a census block for tabulation



# Non-ID Processing Overview (Cont.)

---

- Outcomes
  - Matched to a Geocoded Record in the MAF
  - Matched to Ungeocoded Record, but Subsequently Geocoded
  - Matched to an Ungeocoded Record, Unable to Geocode
  - Nonmatch/Geocoded (must be verified)
  - Nonmatch/Ungeocoded
- Key Goals for 2020
  - Increase amount of cases resolved during automated processing
  - Significantly reduce cost by reducing manual processing & fieldwork



# Non-ID Processing Overview (Cont.)

---

- The NCCT was an opportunity to test address enhancement as part of automated Non-ID Processing
- Three steps to address enhancement
  - Standardization
  - Confirmation
  - Supplementation



# Non-ID Processing Overview (cont.)

---

- Standardization
  - Preparation for address matching and geocoding
  - SAS DataFlux and in-house Geography Division software



# Non-ID Processing Overview (Cont.)

---

- Confirmation
  - Exact matches or minor variations
  - Gives us higher confidence in the address going to MAF matching and geocoding that follows



# Non-ID Processing Overview (Cont.)

---

- Supplementation
  - For addresses not confirmed
  - May result in the addition or correction of city-style address data elements
  - Uses additional match variables

# 2013 National Census Contact Test (NCCT)

## Background



# 2013 NCCT - Test Background

## Survey Operations

---

- Data collected at the Census Bureau's telephone centers
  - Hagerstown, MD,
  - Jeffersonville, IN
  - Tuscon, AZ
- Operation methodology includes:
  - Phone Number Service Check before Operation
  - Mailing Advanced Letter with UAA results
  - Outbound and Inbound Interviewing
  - Telephone Questionnaire Assistance
  - Computer Assisted Telephone Interviewing



# 2013 NCCT - Test Background Questionnaire

---

Two main flows based on address verification at start of interview:

- Address verified by respondent
  - Collect the household roster, demographics, landline phones, cell phones, email addresses
  - Phone and emails in the contact frame are validated during post processing
- Address not verified by respondent
  - Collect current address
  - Ask if respondent is familiar with the sample address & how
  - Collect household roster and demographics only
  - Current address was used for Non-ID Processing



# 2013 NCCT - Test Background Risks/Limitations

---

- 25% response rate assumption
- Final sample of 40,000 needed
  - For vendor addresses that matched to a MAF record, we detected differences among sources of about 2-3%



# 2013 NCCT - Test Background

## Initial Sample Design

---

- Sample was selected from Mailout / Mailback areas of the 50 states and Washington, D.C.
- Limited to addresses with alternative contact frame phone data
- Initially, we oversampled 50,000 addresses to ensure we would have enough sample with working phone numbers



# 2013 NCCT - Test Background

## Preparing for Non-ID Processing

---

- Assembled an administrative records composite consisting of commercial and federal sources
- Employed matching and geocoding system used for evaluating partner-provided address data for Geographic Support Services work

# 2013 National Census Contact Test (NCCT)

## Phone Service Check

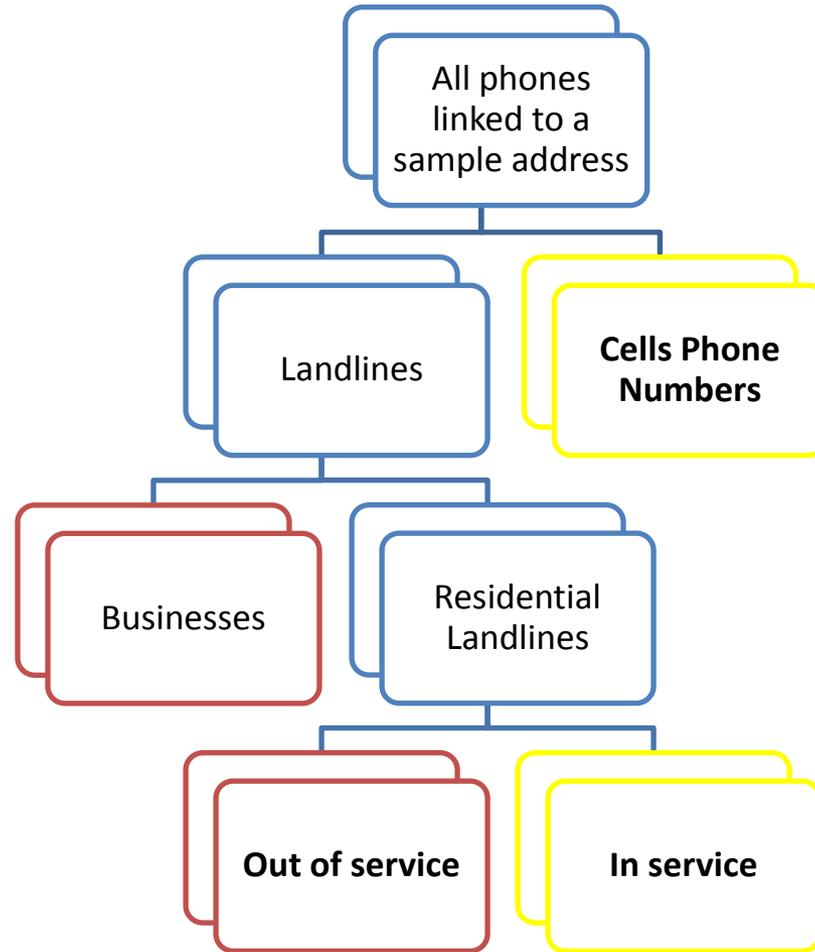


# 2013 NCCT: Phone Service Check Overview

---

- Phone numbers were provided by commercial vendors in 2010, 2011, and 2012
- A preliminary pre-validation stage was conducted
  - To eliminate phone numbers that are out of service
  - To eliminated sample units without potentially valid phone number
- A commercial vendor provided this service for us

# 2013 NCCT: Phone Service Check Results



# 2013 NCCT: Phone Service Check Results (Cont.)

	Phone numbers	% of phone numbers
Cell phone numbers	45K	30% of phone numbers were POSSIBLY GOOD
Landlines in service	51K	34% of phone numbers were POSSIBLY GOOD
Landlines out of service & businesses	53K	35% of BAD phone numbers removed
Overall total	149K	96K remain (64%) 53K removed (35%)

**Possibly good phone numbers** are **landline phones that are in service or cell phones.**

Of 50,000 addresses, 42,538 were found to have one or more phone numbers that were possibly good. These were subsampled down to 40,000.

\* Included phone numbers from 2010-2012 vintages

# 2013 NCCT: Phone Service Check Results (Cont.)

- Phone numbers per sample address:

- Before validation: 3
- After validation: 2.25

- Final phone type mix:  
53% landline / 47% cell

Number of phone numbers per case	% of cases BEFORE validation	% of cases AFTER validation
1	32%	43%
2	21%	25%
3	16%	15%
4	11%	8%
5	7%	4%
6+	12%	5%

\* Included phone numbers from 2010-2012 vintages



# 2013 National Census Contact Test (NCCT)

## General Results



# 2013 NCCT: General Results

## Completion Rate

---

All addresses were sent an advance letter and all sample units received at least one call

- 26% completed cases
  - About 61% verified the address provided by interviewer
  - Of those who did not verify it as their current address, about a quarter were familiar with the address.
    - Of those familiar with the address,
      - About 65% used to live there
      - About 20% said a relative lived there

For Successful MAFID match cases



# 2013 NCCT: General Results

## Respondent Characteristics

---

- Respondent demographics (unedited data) for confirmed address complete cases
- Race: 85% reported White alone, 6% reported Black alone
- Origin: 7% reported to be of Hispanic Origin
- Age: 35% of HH members were 65+ years old
- Tenure: 88% owners
- Respondent demographics were quite different than the general population of the United States. Nonresponse bias is a limitation of the 2013 NCCT results.

For Successful MAFID match cases



# 2013 National Census Contact Test (NCCT)

## Non-ID Processing Results

# 2013 NCCT Non-ID Results

## Address Enhancement - Standardization

Address Type	As Collected by Interviewer		GEO Standardized		DataFlux/ GEO Standardized	
	Number	Percent	Number	Percent	Number	Percent
Complete City-Style	3,515	92.2%	3,630	95.2%	3,639	95.4%
Complete Rural Route	3	0.1%	0	0.0%	0	0.0%
Complete P.O. Box	28	0.7%	15	0.4%	15	0.4%
<b>Complete Subtotal</b>	<b>3,546</b>	<b>93.0%</b>	<b>3,645</b>	<b>95.6%</b>	<b>3,654</b>	<b>95.8%</b>
Incomplete City-Style	218	5.7%	100	2.6%	88	2.3%
Incomplete Rural Route	17	0.4%	6	0.2%	10	0.3%
Incomplete P.O. Box	2	0.1%	0	0.0%	0	0.0%
Incomplete	30	0.8%	62	1.6%	61	1.6%
<b>Incomplete Subtotal</b>	<b>267</b>	<b>7.0%</b>	<b>168</b>	<b>4.4%</b>	<b>159</b>	<b>4.2%</b>
<b>Total</b>	<b>3,813</b>	<b>100.0%</b>	<b>3,813</b>	<b>100.0%</b>	<b>3,813</b>	<b>100.0%</b>

- Increasing success as more enhancement processes added

# 2013 NCCT Non-ID Results

## Address Enhancement: Confirmation & Supplementation

Respondent-provided addresses		With IRS		Without IRS		Increase
		No.	Percent	No.	Percent	Ratio
Only GEO Standardized	confirmed	2,853	74.8%	295	7.7%	10:1
	supplemented	122	3.2%	51	1.3%	2:1
DataFlux/GEO standardized	confirmed	3,138	82.3%	320	8.4%	10:1
	supplemented	81	2.1%	37	1.0%	2:1

# 2013 NCCT Non-ID Results

## MAF Matching and Block Geocoding

Matching and Geocoding Result	GEO Standardized/ Not AR Matched		GEO Standardized/ AR Matched		DataFlux/ GEO Standardized/ AR Matched	
	No.	Percent	No.	Percent	No.	Percent
Matched/Geocoded	2,895	75.9%	2,908	76.3%	3,065	80.4%
Matched /Not Geocoded	<u>40</u>	<u>1.0%</u>	<u>40</u>	<u>1.0%</u>	<u>41</u>	<u>1.1%</u>
<b>Matched</b>	<b>2,935</b>	<b>77.0%</b>	<b>2,948</b>	<b>77.3%</b>	<b>3,106</b>	<b>81.5%</b>
Not Matched/Geocoded	138	3.6%	137	3.6%	145	3.8%
Not Matched/Not Geocoded	<u>740</u>	<u>19.4%</u>	<u>728</u>	<u>19.1%</u>	<u>562</u>	<u>14.7%</u>
<b>Not Matched</b>	<b>878</b>	<b>23.0%</b>	<b>865</b>	<b>22.7%</b>	<b>707</b>	<b>18.5%</b>
<b>Total</b>	<b>3,813</b>	<b>100.0%</b>	<b>3,813</b>	<b>100.0%</b>	<b>3,813</b>	<b>100.0%</b>

- Match rate progressively increased the more the addresses were enhanced



# 2013 National Census Contact Test (NCCT)

## Phone Validation

# Phone Validation Case Level

- Results of last contact attempt for each of the 40,000 cases

Disposition of final contact attempt by case	%
<b>Message Left</b>	<b>31%</b>
<b>Completed case</b>	<b>24%</b>
Refusal/immediate hang up	11%
Answering Machine – but no message left	10%
Number not in service	9%
Ring no answer	8%
Other	7%

- In more than half of the cases, we left a message or conducted an interview



# Phone Validation Case Level

---

- For this national sample, we found
- Good phones-address link 16%
- Potentially good link 78%
- Bad phone-address link 6%

Potentially good links are when we have remaining phone numbers we were not able to follow up with in this study

\*Excludes addresses with only landline phones found to be out of service by Service Check



# Phone Validation: Phone Level

---

- About 79K phone numbers were associated with the 32K sample units
  - 43K phones called 1+ times
  - 36K phones never called
    - 8K not called because # is associated with a case already completed
    - 28K never called & associated with incomplete cases (limitation of the control system)
- Overall, 68K calls attempted

# Phone Validation: Phone Level Overall

---

- For this national sample, we found
  - Good phone-address combo 14%
  - Potentially Good (in service) 56%
  - Potentially Good (unknown service) 24%
  - Bad phone-address combo 6%

Potentially good links are phone numbers we were not able to follow up with in this study and get a final address confirmation

\*Excludes landline phones found to be out of service by Service Check



# 2013 National Census Contact Test (NCCT)

## Email Validation



# 2013 NCCT: Email Validation

---

We also evaluated the email addresses from our vendor data which we were able to link to our living quarters address frame

# 2013 NCCT: Email Validation Initial Metrics

Overall, we obtained:

	#	universe	%
2010 Census HU MAFIDs with 1+ email addresses	51.7 million	131.7 million	39%
2010 Census NRFU HU MAFIDs with 1+ email addresses	12.1 million	49.8 million	24%

Using 2010 sources only

# 2013 NCCT: Email Validation Initial Metrics (Cont.)

- Regionally, % of addresses with 1+ emails

<b>Midwest</b>	<b>42%</b>	<b>Northeast</b>	<b>39%</b>
West	37%	South	39%

- By state, % of addresses with 1+ emails

<b>Highest State</b>	<b>Ohio</b>	<b>47%</b>
Lowest State	Vermont	22%

Using 2010 sources only



# 2013 NCCT: Email Validation

## Email Validation Methods

---

- Methodology of the 2013 NCCT was less well suited to learn about email validity
  - No pre-survey validation method
  - Did not contact respondents by email
- Respondents were asked about their email use and to provide addresses they check at least once a week

# 2013 NCCT - Email Validation Results

- Results of validation for both the 2012 NCT and the 2013 NCCT

	2012 NCT	2013 NCCT
Response mode	Internet	Telephone
Response rate	38%	26%
% of respondents who provided 1+ email addresses	92%	25%
% of addresses where 1+ respondent provided emails were found on the contact frame at that address	10%	16%

# Next Steps in Contact Frame Research



# Contact Frame: Next steps

---

- Keep growing the contact frame
- Learn more about data we have
- Continue to validate the frame
- Develop in-house abilities

# Next steps: Contact Frame

## Keep growing the frame

---

- Continually refurbish the contact frame with more current contact information - Ongoing
  - Refresh current commercial sources
  - Obtain contact data collected from other surveys (NSCG, ACS, ...)
    - Respondent reported
    - Results of successful contacts
  - Obtain state level data sources (SNAP and drivers license applications, etc...)



# Next steps: Contact Frame

## Continue to validate the frame

---

- Conduct internal test of contact validity
- Provide contact information for upcoming tests
- Learn more using data we already have
- Investigate other methods of email validation



# Next steps: Contact Frame

## Develop in house abilities

---

- Develop a process to identify cell/landline status
- Evaluate conducting phone service checks in house
- Determine characteristics of correctly matched phones to order multiple phone numbers associated with an address



# Next Steps in Non-ID Research



# Next Steps

## Non-ID Processing

---

- Continue detailed analysis of 2013 NCCT data to fine tune processing
- Address enhancement for 2010 workload
- Production during field tests

# Thank you

Please direct any questions to:

[20rpo.program.management@census.gov](mailto:20rpo.program.management@census.gov)