

## 2013 National Census Contact Test

**Timothy D. Stewart**  
**United States Census Bureau**

**Disclaimer:** This report is released to inform interested parties of ongoing research and to encourage discussion. Any views expressed on statistical or methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

### Abstract

To meet the strategic goals and objectives for the 2020 Census, the United States Census Bureau must make fundamental changes to the design, implementation, and management of the decennial census. The Census Bureau designed the 2013 National Census Contact Test (NCCT) to collect data from household members in order to validate landline and cell phone numbers as well as email addresses associated with the household's address that were obtained from multiple commercial sources. In January 2013, we conducted telephone interviews with respondents using a web-based Computer Assisted Telephone Interviewing (WebCATI) instrument.

In this paper, we present the 2013 NCCT study design, several methods such as validating phone numbers prior to telephone interviews, and some preliminary results that begin to assess the quality of our contact frame. The ultimate goal of the Census Bureau for this study is to utilize our contact frame to initiate contact with respondents via landline or cell phone and/or email in an overall effort to reduce future census costs while maintaining or improving quality.

**Keywords:** contact frame, telephone interviewing, CATI, administrative records

### 1 Introduction

To meet the strategic goals and objectives for the 2020 Census, the Census Bureau must make fundamental changes to the design, implementation, and management of the decennial Census. These changes must build upon the successes and address the challenges of the previous Censuses while also meeting standards for cost containment, quality, flexibility, innovation, and disciplined and transparent acquisition decisions and processes. Through the 2020 Census Research and Testing Program, research teams are identifying and analyzing options to improve planning and operations for the next decennial census. The Contact Frame team was established to link household contact information to address data. This paper presents results from the first test to use the contact frame.

#### 1.1 Research Purposes

The purposes of this research are to describe the methodological and analytical approaches to evaluate the quality of a supplemental contact frame, composed of contact information from different businesses (vendors), and to present some 2013 National Census Contact Test (NCCT) preliminary results. Another purpose of this test, which is beyond the scope of this paper, is to test proposed enhancements to automated processing of census responses lacking a pre-assigned census identification (ID) number, also known as "Non-ID cases."

The 2013 NCCT team, comprised of the Contact Frame team and the Non-ID Processing team, conducted this research. The Contact Frame team is responsible for validating the alternative contact frame of phone numbers and email addresses associated with an address. If a contact frame of sufficient quality can be established, the Census Bureau plans to use administrative records data to facilitate contacting respondents via telephone (i.e., cell phone or landline), text messages, and/or email in an overall effort to encourage self response by mail or internet and/or to contact non-respondents prior to a personal visit. The Non-ID Processing team is determining ways to enhance the automated processing of Non-ID cases. Non-ID processing includes both matching to records in the Master Address File (MAF), the Census Bureau's national inventory of addresses, and geocoding, which, in this case, involves assigning an address to a census block. For this paper, we focus on the efforts done by the Contact Frame team with some brief mention of efforts done for the Non-ID Processing team.

## **1.2 Background of Relevant Prior Work**

In the initial efforts to assemble a contact frame, the Census Bureau conducted the Information Reseller (IR) project in 2006, employing a contractor to assess the IR market and recommend a dataset for purchase. Because of that project, the Census Bureau acquired commercial data. This vendor file was tested for use as a sample frame for the Census Bureau's demographic surveys and was evaluated for other census and survey uses.

In 2009, the Census Bureau sought additional commercial data for use in the 2010 Census Match Study (Rastogi and O'Hara, 2012), which compared 2010 Census results with administrative records and commercial data. A new acquisition process was initiated to obtain commercial sources of person and address data, as well as roster information, alternate contact information (i.e., cell phone numbers and emails), and historical addresses if available. The commercial sources obtained their information from various public records such as telephone white pages, magazine subscription records, and utility bill records. The Census Bureau conducted a full and open competition with the vendors, requiring the submission of test files for three states. Each vendor's person and address data were validated and assessed. The process resulted in contract awards to five vendors, all of whom provided telephone and email address data.

The Census Bureau has not yet initiated contact with respondents using email or text messaging. However, telephone interviewing is a commonly used tool for the Census Bureau staff conducting interviews for the regular demographic surveys, such as the Current Population Survey. Outbound calls are placed for the current surveys for addresses in sample using Computer Assisted Telephone Interviewing (CATI) instruments. The Contact Frame team is planning to review the paradata collected during CATI portion of the American Community Survey (ACS) to extend the analysis from the 2013 NCCT and to inform future 2020 Census tests.

The Contact Frame team is investigating the literature on non-mail contact by other statistical and survey organizations. For example, once individuals opt-in to a National Center for Health Statistics (NCHS) program, they communicate via text messages. The team has been reviewing external studies to monitor market research on cell phone use, broadband penetration, and social networking trends. In addition, the team has been

following the progress of the National Science Foundation (NSF) – Census Research Network (NCRN) grantees who are studying similar topics.

## **2 Assumptions**

When we contacted households by phone for this test, we asked respondents about their household members and the phone numbers and email addresses used by household members. The survey asked that the respondents provide this kind of information about themselves as well as other household members. The results of cognitive testing indicated the respondents might be reluctant or unwilling to provide this information for others in the household (Smirnov, 2012), so we assumed that the results we saw would be a lower bound in terms of validation of phone numbers and email addresses.

Given the absence of any documentation on a similar sort of effort or study, the 2013 NCCT team determined the best strategy for sampling would be to err on the side of being conservative in estimating response rate. Therefore, the team assumed a 25 percent response rate out of the 40,000 sampled housing units, or 10,000 housing units. Based on an evaluation of potential sample sizes, the team felt 10,000 completed interviews would enable the team to make determinations about the quality and comprehensiveness of the commercial datasets in considering them for future 2020 testing efforts.

## **3 2013 NCCT Methodology**

### **3.1 Creation of Contact Frame Database**

The Contact Frame team has been developing a supplemental comprehensive contact frame database in the past few years. The team, through an extensive acquisition process, purchased 18 different commercial databases from five businesses, which contained addresses, phone numbers, and email addresses along with other person and household characteristics. The five businesses used various public records such as telephone white pages, utility bills, magazine subscriptions, and change of address notifications to create their databases. We acquired these databases across three vintage years (2010-2012). The addresses on the commercial databases were linked to an extract from the MAF of the addresses that were considered valid residential addresses at the time of the match. The MAF identification (ID) numbers were attached to the records that could be matched. The contact frame database is organized at the MAFID-phone number unit of analysis with each row of data containing flags to denote what data source(s) that record came from. Its database design easily facilitates data additions, edits, and extractions as we collect other data sources over time. Using data obtained in 2010 and 2011, the contact frame database had over 120.2 million MAFIDs with one or more phone numbers.

### **3.2 Sample Design**

As part of the 2020 Census Research and Testing Program, the 2013 NCCT is the first test to support supplemental contact frame research as well as research on enhancing the automated processing of Non-ID cases. The sample frame consisted of addresses from vendor files, which also contained supplemental contact information (i.e., phone numbers and emails). In order to address the research questions of both the Contact Frame team and the Non-ID Processing team, two panels were established.

The first panel was a systematic random sample of addresses provided in the vendor files that matched to housing unit (HU) records in the July 2012 MAF extract, which was about 77.5 percent of the total vendor address inventory. The second panel included records from the vendor datasets for which there was not a matching address in the MAF extract. In addition, to reduce burden on respondents, the ACS cases for the 2012 calendar year and the first six months of the 2013 calendar year were excluded from the sampling frame. Both panels are described in detail below.

### **Panel 1**

The Contact Frame team has been interested in the matched addresses because of the practical application of the contact data for the decennial census. The team's goal is to associate phone number(s) and/or email address(es) with living quarters addresses in the Census inventory in order to facilitate various contact strategies during the 2020 Census. This panel included housing units in the mailout/mailback areas and military areas of the 50 states and the District of Columbia. Group quarters and housing units in any other types of enumeration areas (e.g., Update/Leave, Remote Alaska, etc.) were not included in this panel.

Based on power analysis calculations, the NCCT team observed that a sample of 32,000 HUs for Panel 1 permitted us to detect differences among vendors of around two to three percent. This was based on a 25 percent survey completion rate and accounted for the overlap among vendors in which many addresses had the same phone numbers provided.

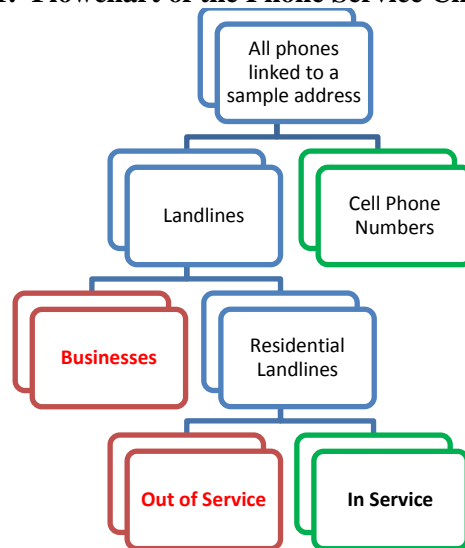
### **Panel 2**

The addresses that did not match to the MAF extract were of greater interest to the Non-ID Processing team. Based on the respondent's answers, the team investigated whether these vendor-supplied address were missing from the MAF extract (i.e., if the respondent confirmed the address) or if the match was unsuccessful for other reasons. If the respondent did not confirm the vendor-supplied address, the interviewer attempted to collect a new address. Both the confirmed addresses as well as the new addresses were matched to the MAF/TIGER database once the interviewing was complete.

Panel 2 cases permitted us to detect differences among vendor data at around five to eight percent. The Panel 2 cases were primarily intended to provide an opportunity to test proposed enhancements to Non-ID Processing. Also, while the test was not intended to make comparisons for these cases, it was possible to obtain some useful information by including a small number of them in the study.

### **Final Sample**

In order to obtain our final sample of 40,000 housing units in two panels, we sent an initial combined sample of 50,000 housing units with associated phone numbers through a phone service check process. This check first identified phone numbers as landlines or cell phone numbers. If they were identified as cell phone numbers, the process stopped because we could not autodial these cell phone numbers to check if they were in service because of Federal law. If they were identified as landline numbers, then a vendor compared them against phone numbers in their business registry. Business phone numbers were deemed invalid; all other phone numbers were classified as residential landline phone numbers. The residential landline phone numbers were further checked by autodialing the phone numbers and listening to the tones to determine whether or not they were out of service. Figure 1 below summarizes the phone service check process.

**Figure 1: Flowchart of the Phone Service Check Process**

This phone check allowed us to have a more efficient telephone interviewing operation, by eliminating phone numbers that were out of service or business phone numbers. Eliminating out of service and/or business phone numbers placed some housing units out of the sample because they had no remaining valid phone numbers. After this process, we were left with 42,538 addresses which had one or more phone numbers not determined to be out of service or associated with a business; 34,394 (81%) in Panel 1 and 8,144 (19%) in Panel 2. These two panels were then each subsampled down to the desired final sample of 40,000, split by the same proportion of address in each sample before the reduction; Panel 1 was 32,340 (81%) and Panel 2 was 7,660 (19%).

Table 1 summarizes the total number of available phone numbers in the final sample. Around 53.2 percent of the overall 90,646 phone numbers were landline phone numbers and 46.8 percent were cell phone numbers.

**Table 1: Analysis of Valid Phone Numbers in Final Sample**

Stratum	Number of Valid Phone Numbers	Percent of Overall Number of Valid Phone Numbers
Landline Phone Numbers	48,250	53.2%
Cell Phone Numbers	42,396	46.8%
<b>Overall Total</b>	<b>90,646</b>	<b>100.0%</b>

Source: 2013 NCCT Sample Universe File

### 3.3 2013 NCCT Survey Operations

The Census Bureau contact centers in Hagerstown, Maryland; Jeffersonville, Indiana; and Tucson, Arizona, conducted the interviewing for this project. The operational methodology included an advance letter, outbound interviewing, and telephone questionnaire assistance (TQA).

### 3.3.1 Advance Letter

Prior to the start of telephone interviewing, an advance notification letter was sent to all addresses in the sample on January 3, 2013. The letter was intended to improve the response rate, as previous studies (Ford, 1967; Link, 2005) have shown. The letter:

- informed respondents that the NCCT is mandatory,
- provided the Office of Management and Budget (OMB) number,
- provided information about the survey,
- and included a telephone number that respondents may call if they had any questions that were not answered in the letter.

### 3.3.2 WebCATI Operations

Interviewing took place starting on January 7, 2013, and continued through January 31, 2013, for a total of almost four weeks. The 40,000 housing units in sample were managed using the WebCATI survey management system. WebCATI provided centralized management of the sample cases and routed calls to the participating contact centers. WebCATI provided basic case status information, set codes for non-contact call outcomes, allowed interviewers to make appointments, and set case priority in the calling queue. The project team set parameters to determine how the WebCATI assessor determined the case priority.

Once a phone number had been reached and the interviewer was speaking with someone for an interview, WebCATI loaded the survey instrument. The interview was conducted, and upon completion, data was returned through WebCATI. If the interview was not completed, the interviewer could set a callback appointment, mark the interview as a refusal, or assign other outcomes as necessary. WebCATI provided survey data and operational data on call outcomes to the 2013 NCCT team.

Many of the 2013 NCCT cases had more than one telephone number; there was an average of three phone numbers per case. Interviewers were given the option to toggle between phone numbers, so information on as many numbers as possible could be obtained during the interview. Since respondents are more likely to refuse completing telephone surveys, we increased the number of refusals to three, from the standard of two for most surveys, to allow for an additional phone number to be tried.

### 3.3.3 Survey Instrument and Questionnaire Format

The WebCATI instrument for the 2013 NCCT was programmed using BLAISE software by the Technologies Management Office (TMO) of the United States Census Bureau. The 2013 NCCT questionnaire asked questions about the following topics:

- Whether or not the respondent received the advance letter (if not, the interviewer must read a confidentiality statement)
- Address verification
  - If respondent lived at the sampled address
  - If respondent did not live at the sampled address, were they familiar with the address?
  - If respondent was familiar with the sampled address, describe how respondent was familiar with the address
  - If respondent did not live at the sampled address, collect their current address

- Tenure information (e.g., owned with or without mortgage, rented, etc.)
- Length of time at address
- Roster
- Demographics of roster members (relationship to respondent, sex, age, Hispanic origin, race)
- Cell phone questions for roster members (age 18+ and live at sampled address)
  - If they had a cell phone
  - If they used the cell phone for texting
  - If they accessed the internet using the cell phone
  - Collect cell phone numbers
- If household had landline phone (if they lived at sampled address)
  - Collect landline phone numbers
  - If phone numbers were unlisted or unpublished
- If there were any other numbers (work phone or additional phone line) for household members and what those numbers were (if they lived at sampled address)
- Email questions for roster members (age 18+ and lived at sampled address)
  - How often email was checked
  - Collect email addresses (if email was checked at least once a week)
  - If they did not provide an email address, questions about the email domain (i.e., MSN, Hotmail, AOL, etc.)
- Additional question for respondents – did they own a tablet (e.g., iPad)?
- Respondent's attitude regarding the potential collection of Global Positioning System (GPS) coordinates for their location for use in correctly assigning them to a tabulation area.

### **3.3.4 Telephone Questionnaire Assistance**

The advance letter for the 2013 NCCT contained a telephone number that respondents could call if they had any questions about the 2013 NCCT survey. This phone number connected the respondent with an interviewer at one of the three call centers. The interviewer would be available to answer questions from the respondent and would try to conduct the interview with the respondent while they were on the phone. A Case ID would be included on the advance letter that the interviewer could use to locate the respondent's case in WebCATI to conduct the interview and assign the appropriate case status. Even if the respondent did not verify the address provided by the interviewer, but instead provided a different address, the Case ID remained the same. These newly-collected addresses would be researched by the Non-ID Processing team by matching them to the MAF/TIGER database and attempting to assign them to census blocks.

## **4 2013 NCCT Preliminary Results**

The Contact Frame team reports some preliminary results from this test. An official Census Bureau document with full results from this test will be released in the future. Our results here are in the following three major areas:

- General survey and demographic results
- Phone number validation results
- Email validation results

#### 4.1 General Survey and Demographic Results

Table 2 outlines the survey completion results of the two different panels. We achieved our expected goal of 25 percent completion rate.

**Table 2: Survey Completion Rate**

Stratum	Number of Sampled Cases	Number of Interviewed Cases	Row Percent of Sampled Cases Interviewed
Panel 1	32,340	8,367	25.9%
Panel 2	7,659	1,872	24.4%
<b>Overall Total</b>	<b>39,999*</b>	<b>10,239</b>	<b>25.6%</b>

Source: 2013 NCCT Instrument Data File

\*One case was removed because it was later discovered that it had a mailing address outside of the United States.

Then, for Panel 1, 61.3 percent ( $\pm 0.3$ ) of the respondents who completed the interview confirmed the address, as shown in Table 3. We focus only on the Contact Frame team's Panel 1 results for the remainder of this paper.

As a side note for Tables 3 through 10, the missing values were subtracted from the overall total and the resulting total became the denominator for determining the percentages; they are denoted as \*\*.

**Table 3: Analysis of Address Confirmation**

Stratum	Number of Interviewed Households	Percent of Overall Interviewed Households (Std. Error)
Confirmed Address	5,078	61.3% (0.2)
Did not Confirm Address	3,212	38.7% (0.2)
Missing	77**	---
<b>Overall Total</b>	<b>8,367</b>	<b>100.0%</b>

Source: 2013 NCCT Instrument Data Files

Of those who did not confirm the address, the interviewed respondents were asked how familiar they were with the sampled address as shown in Table 4. Around 26.3 percent ( $\pm 1.8$ ) of these respondents reported being familiar with the address. A large percentage (17.2 percent ( $\pm 1.0$ )) of these respondents who did not confirm the address but contact frame associated with that phone number also reported, "I used to live there."



**Table 4: Analysis of Respondents' Familiarity with the Sampled Address but Did not Confirm Address**

Responses	Number of Interviewed Respondents	Percent of Overall Interviewed Respondents (Std. Error)
<b>Familiar with Sampled Address</b>	<b>843</b>	<b>26.3% (1.1)</b>
I used to live there	552	17.2% (0.6)
My parents live there	37	1.2% (0.0)
My children live there	85	2.6% (0.5)
Another relative lives there	43	1.3% (0.1)
I use this address for billing purposes	1	<0.1% (<0.1)
I use this as a shipping address	1	<0.1% (<0.1)
Other (Specify based on responses)	123	3.8% (0.5)
Refuse	1	<0.1% (<0.1)
<b>Not Familiar with Sampled Address</b>	<b>2,366</b>	<b>73.7% (1.1)</b>
<b>Missing</b>	<b>3**</b>	<b>---</b>
<b>Overall Total</b>	<b>3,212</b>	<b>100.0%</b>

Source: 2013 NCCT Instrument Data Files

Of those who confirmed the address, the majority of the respondents were homeowners at 88.2 percent ( $\pm 0.8$ ) as shown in Table 5. Most of them (88.9 percent ( $\pm 0.5$ )) have lived at the sampled address for more than 5 years as shown in Table 6.

**Table 5: Tenure (Confirmed Address)**

Tenure	Number of Interviewed Households	Percent of Overall Interviewed Households (Std. Error)
<b>Owned</b>	<b>4,304</b>	<b>88.2% (0.5)</b>
Owned with a mortgage or a loan	2,263	46.4% (0.8)
Owned without a mortgage or a loan	2,041	41.8% (0.9)
<b>Rented</b>	<b>514</b>	<b>10.5% (0.5)</b>
<b>Occupied without payment of rent</b>	<b>62</b>	<b>1.3% (0.2)</b>
<b>Missing</b>	<b>198**</b>	<b>---</b>
<b>Overall Total</b>	<b>5,078</b>	<b>100.0%</b>

Source: 2013 NCCT Instrument Data Files

As a side note for the reader, a major limitation of this paper is the nonresponse bias. It is apparent that the respondent demographics were quite different from that of the general United States population. However, it is not clear how much the bias impacts the phone and email validation results later in this paper. It is recommended to keep this bias in mind for the remainder of this paper.

**Table 6: Length of Time Respondent Stayed at the Address (Confirmed Address)**

Length of Time Respondent Stayed at the Address	Number of Interviewed Households	Percent of Overall Interviewed Households (Std. Error)
Less than one year	34	0.7% (0.0)
One to two years	133	2.7% (0.1)
More than two years to five years	379	7.7% (0.3)
More than five years	4,389	88.9% (0.3)
Missing	143**	---
<b>Overall Total</b>	<b>5,078</b>	<b>100.0%</b>

Source: 2013 NCCT Instrument Data Files

Other notable demographic characteristics of households where respondents confirmed the address are that they were likely to be White (85.2 percent ( $\pm 0.5$ )), non-Hispanic (93.1 percent ( $\pm 0.5$ )), over age 65 (34.7 percent ( $\pm 1.3$ )), over age 50 (62.9 percent ( $\pm 0.7$ )), and female (51.6 percent ( $\pm 0.3$ )). See Tables 7 through 10 for detailed household demographics. Future analysis will provide information on the characteristics of people who did not confirm a phone-residential address link.

**Table 7: Race (Confirmed Address)**

Race	Number of Roster Members	Percent of Overall Roster Members (Std. Error)
White	9,058	85.2% (0.3)
Black or African American	640	6.0% (0.1)
American Indian or Alaska Native	42	0.4% (0.0)
Asian	377	3.5% (0.2)
Native Hawaiian or Other Pacific Islander	23	0.2% (<0.1)
Other Race	230	2.2% (0.1)
Multiple	257	2.4% (0.2)
Missing	237**	---
<b>Overall Total</b>	<b>10,864</b>	<b>100.0%*</b>

Source: 2013 NCCT Instrument Data Files

**Table 8: Hispanic Origin (Confirmed Address)**

Hispanic Origin	Number of Roster Members	Percent of Overall Roster Members (Std. Error)
Yes	736	6.9% (0.3)
No	9,922	93.1% (0.3)
Missing	206**	---
<b>Overall Total</b>	<b>10,864</b>	<b>100.0</b>

Source: 2013 NCCT Instrument Data Files

**Table 9: Age (Confirmed Address)**

Age	Number of Roster Members	Percent of Overall Roster Members (Std. Error)
<b>Under 5</b>	207	1.9% (0.1)
<b>5-9</b>	405	3.8% (0.2)
<b>10-14</b>	501	4.7% (0.1)
<b>15-19</b>	575	5.4% (0.1)
<b>20-24</b>	359	3.3% (0.3)
<b>25-29</b>	249	2.3% (0.3)
<b>30-34</b>	248	2.3% (0.2)
<b>35-39</b>	279	2.6% (0.1)
<b>40-44</b>	515	4.8% (0.1)
<b>45-49</b>	648	6.0% (0.1)
<b>50-54</b>	870	8.1% (0.2)
<b>55-60</b>	1,035	9.6% (0.6)
<b>60-64</b>	1,114	10.4% (0.4)
<b>65 and over</b>	3,725	34.7% (0.8)
<b>Missing</b>	134**	---
<b>Overall Total</b>	<b>10,864</b>	<b>100.0%*</b>

Source: 2013 NCCT Instrument Data Files

**Table 10: Sex (Confirmed Address)**

Sex	Number of Roster Members	Percent of Overall Roster Members (Std. Error)
Male	5,097	47.3% (0.2)
Female	5,678	52.7% (0.2)
Missing	89**	---
<b>Overall Total</b>	<b>10,864</b>	<b>100.0%*</b>

Source: 2013 NCCT Instrument Data Files

#### 4.2 Phone Validation Results

Before presenting the phone validation results, we need to discuss the limitations of our results. Not every telephone number associated with a sampled housing unit was attempted. If there was a completed interview, regardless of address verification, it rendered the other phone numbers ineligible for further contact attempts. In other words, the other available phone numbers could be the correct number to contact for an address, but we would never know from this test. There was not enough time or resources to call every phone number associated with every sampled housing unit due to the budget.

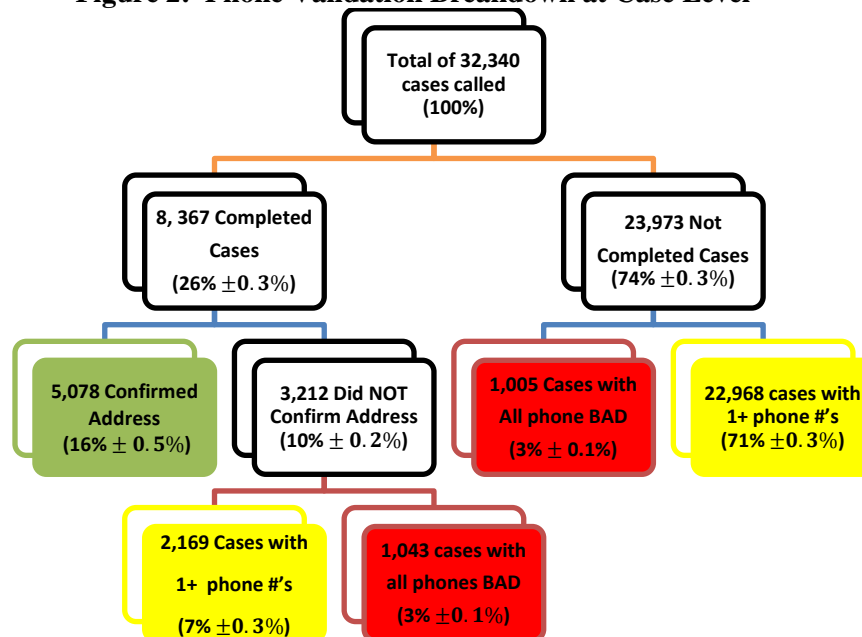
Table 11 presents a summary of phone number used for the interviewed cases, regardless whether or not the address was confirmed. Nearly 70 percent ( $\pm 1.0$ ) were landline phone numbers and almost 30 percent ( $\pm 1.0$ ) were cell phone numbers. Thirty-four respondents called the Telephone Questionnaire Assistance (TQA) line and completed the interview.

**Table 11: Analysis of Phone Number Used for Completed Interview**

Stratum	Number of Interviewed Cases	Percent of Overall Number of Interviewed Cases (Std. Error)
<b>Interview Phone Number also in Frame</b>	<b>8,330</b>	<b>99.6% (0.5)</b>
Landline Phone Number	5,848	69.9% (0.6)
Cell Phone Number	2,482	29.7% (0.6)
<b>Respondent Called the TQA Line</b>	<b>34</b>	<b>0.4% (&lt;0.1)</b>
<b>Missing Phone Number but Interviewed</b>	<b>3</b>	<b>&lt;0.1% (&lt;0.1)</b>
<b>Overall Total</b>	<b>8,367</b>	<b>100.0%</b>

Source: 2013 NCCT Instrument Data Files

Now, we examine the phone validation results at the housing unit (case) level in Figure 2. Starting at the top, all 32,340 cases were contacted by phone. At the next level, the number of cases completed was 26 percent ( $\pm 0.3$ ) and not completed was 74 percent ( $\pm 0.3$ ). Of the completed cases, we were able to confirm a good phone-residential address link for 5,078 cases (or 16 percent ( $\pm 0.5$ ) of the overall 32,340 cases attempted). In a similar manner, we also know that 3,212 cases did not confirm the address (or 10 percent ( $\pm 0.2$ ) of the overall 32,340 cases attempted). Of those 3,212 cases, there were 2,169 cases with one or more other phone numbers that were not attempted but were potentially the correct phone-residential address link (or 7 percent ( $\pm 0.3$ ) of the overall 32,340 cases attempted), and 1,043 cases with no other phone number available (or 3 percent ( $\pm 0.1$ ) of the overall 32,340 cases attempted). Of the not completed cases, we had 22,968 cases with one or more other phone numbers that were not attempted but were potentially the correct phone-residential address link (or 71 percent ( $\pm 0.3$ ) of the overall 32,340 cases attempted) and 1,005 cases with no other phone numbers available (or 3 percent ( $\pm 0.1$ ) of the overall 32,340 cases attempted).

**Figure 2: Phone Validation Breakdown at Case Level**

Source: 2013 NCCT Instrument Data Files

Combining all of the results, we know the following:

- 16 percent ( $\pm 0.3$ ) of cases had a good phone-residential address link
- 78 percent ( $\pm 0.3$ ) of cases had a potentially good phone-residential address link if we attempted all other phone numbers until we contact the right address
  - 63 percent ( $\pm 0.5$ ) of the cases had at least one landline in service phone number
  - 15 percent ( $\pm 0.5$ ) of the cases had only cell phone numbers left
- 6 percent ( $\pm 0.3$ ) of cases had a bad phone-residential address link.

In summary, the good news was that over 26 percent ( $\pm 0.3$ ) of the households we called answered the phone and responded. Around 30 percent ( $\pm 1.0$ ) of those who completed the interview were conducted through the respondent's cell phone. This is promising because more and more households are using cell phones. The bad news was that there were many cases where we were not able to determine if we had a correct phone-address link. We were also left with many phone numbers about which we have no information or know only that they were in service.

Another point to consider was that the phone numbers were randomly sorted prior to the interviews, so it was possible that an older, potentially outdated phone number was used to initiate the interview and failed to establish a good phone-address link. This was done to allow comparisons across vendors in our later analysis. The Contact Frame team is working on a prioritization algorithm to order phone numbers when multiple numbers are provided by vendors. The team can see that phone numbers present across multiple years of contact frame data are more likely to be correct than phone numbers from just the most current year. Likewise, phone numbers from the most current year are more likely to be correct than three year old phone numbers that only appeared in the 2010 dataset. We hypothesize that we would get a higher response rate than 25 percent with an optimal prioritization algorithm if we continue to toggle through the phone numbers until one number reaches a respondent at the right address.

#### 4.3 Email Validation Results

The 2013 NCCT provided less information about email addresses than phone numbers. There was no pre-survey validation method for email addresses like there was for phone numbers, and we did not contact respondents by email. Only 41.1 percent ( $\pm 0.8$ ) of the households who confirmed the address had respondents who provided email addresses as shown in Table 12.

**Table 12: Analysis of Households whose Respondents Provided Email Addresses**

Stratum	Number of Interviewed Households	Percent of Overall Interviewed Households (Std. Error)
Provided Email Addresses	2,087	41.1% (0.5)
Did not Provide Email Addresses	2,991	58.9% (0.5)
<b>Overall Total</b>	<b>5,078</b>	<b>100.0%</b>

Source: 2013 NCCT Instrument Data Files

Table 13 contains a breakdown of the matching results of respondent-provided email addresses and the email addresses within the contact frame database. Around 16.1

percent ( $\pm 2.8$ ) of the households provided email addresses that matched to the contact frame email addresses; 15.4 percent ( $\pm 0.7$ ) of households had one email address that matched and 0.7 ( $\pm 0.1$ ) percent of households had two email addresses that matched. The remaining households either did not have any emails provided by the vendors (28.5 percent ( $\pm 2.0$ )) or did not match to any emails provided by the vendors (55.4 percent ( $\pm 1.8$ )).

**Table 13: Analysis of Matches between Respondent-Provided Email Addresses and Vendor-Provided Email Addresses**

Stratum	Number of Households with Email Address Matches	Percent of Overall Households with Email Address Matches (Std. Error)
<b>No Match</b>	<b>1,751</b>	<b>83.9% (1.7)</b>
Vendor Emails Available	1,157	55.4% (1.1)
No Vendor Email Available	594	28.5% (1.2)
<b>Match</b>	<b>336</b>	<b>16.1% (1.7)</b>
1 Match	322	15.4% (0.4)
2 Matches	14	0.7% (<0.1)
<b>Overall Total</b>	<b>2,087</b>	<b>100.0%</b>

Source: 2013 NCCT Instrument Data Files

Obviously, this test design was far from optimal to validate email addresses. The Contact Frame team plans to conduct more research on email addresses by comparing email addresses collected from other Census Bureau surveys with email addresses within the contact frame database.

## 5 Conclusions and Future Research Plans

In conclusion, the 2013 NCCT survey compared respondent-provided phone numbers and email addresses with our supplemental contact frame to evaluate the newly built contact frame. Our ability to determine the quality of our contact frame was limited due to respondents not being willing to share all of their available phone numbers and email addresses. The Contact Frame team learned a limited amount about phone numbers and the characteristics of those who were likely to respond to the phone survey, but less about email addresses. More research is needed for those areas.

The Contact Frame team is developing the following future research goals as a result of the 2013 NCCT analysis:

- Develop a prioritization algorithm to order the phone numbers most likely to have a correct phone- residential address link.
- Conduct phone number and email address verifications with other available Census surveys such as the 2010 Census, 2010-2012 American Community Survey, and the 2012 National Census Test.
- Conduct analysis of phone-residential address links at lower levels of geography (state, county, block, and tract-levels) which will require looking at larger datasets.
- Investigate other phone number and email address sources (commercial or government/administrative records sources) to improve the demographic and geographic coverage of phone numbers and email addresses.

- Conduct analysis of residential address-email address links. Perhaps, develop a prioritization algorithm to order the email addresses most likely to have a correct residential address- email address link. Also, analyze these links at lower level of geography.

## 6 References

Ford, N. 1967. The Advance Letter in Mail Surveys. *Journal of Marketing Research Vol. 4, No. 2*, pp. 202-204.

Link, M. and A. Mokdad. 2005. Advance Letters as a Means of Improving Respondent Cooperation in RDD Studies: A Multi-state Experiment. *Public Opinion Quarterly* 69:572-587.

Niosi, M. 2012. 2010 Census Non-ID Processing Assessment Report. *2010 Census Planning Memoranda Series No. 201 (Reissue)*, July 13, 2012.

Rastogi, S. and A. O'Hara. 2012. 2010 Census Match Study. *2010 Census Match Study*. Available online at <[http://www.census.gov/2010census/pdf/2010\\_Census\\_Match\\_Study\\_Report.pdf](http://www.census.gov/2010census/pdf/2010_Census_Match_Study_Report.pdf)>

Smirnov, M. and Scalon, P. 2013. "2013 National Census Contact Test (NCCT) Cognitive Testing Report." *Research and Methodology Directorate, Center for Survey Measurement Study Series (Survey Methodology #2013-19)*. U.S. Census Bureau. Available online at <<http://www.census.gov/srd/papers/pdf/ssm2013-19.pdf>>.

Stewart, T. 2013. 2013 National Census Contact Test Study Plan. *DSSD 2020 Decennial Census R&T Memorandum Series #K-02*, (forthcoming).

Tomaszewski, C. 2012. 2010 Census Evaluation of Address List Maintenance Using Supplemental Data Sources. *2010 Census Planning Memoranda Series* (forthcoming).