

## The Research and Methodology on Staggering the 2020 Census Mailings

Ioana (Julia) Marasteanu<sup>1</sup>

<sup>1</sup>U.S. Census Bureau

### Abstract

In order to avoid potentially overloading data collection systems, such as Census Questionnaire Assistance call centers, the 2020 Census will stretch out each of its five mailings over a one-week period and put them into groups of different mail cohorts instead of delivering all of its mailings to the whole country on a single day. The purpose of this study is to establish a method by which to separate addresses into cohorts, such that there is an even distribution of cohorts throughout the country, the burden on the call centers is minimized, and the response rate is maximized. Using analyses and results from mid-decade testing at the Census Bureau, we examine the following aspects of cohort assignment: 1. The geographic level at which the cohorts should be broken up (e.g., Census tract versus block); 2. Operational constraints; 3. Based on their relationship to response rates, which/how characteristics should be used to determine the breakup and timing of the cohorts; and 4. The distribution of the cohorts. This study uses statistical methods to attempt to improve the efficiency and success of the 2020 Census.

**Key Words:** Census, mail strategy, response rate, adaptive design

Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this data product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release.

## I. Introduction

In March and April of 2020, the U.S. Census Bureau will use the U.S. Postal Service (USPS) to send mailings to the majority of housing units, asking them to respond to the 2020 Census. These housing units will have the option of responding to the census online, by mail or by phone (Census, 2018a). With over 143 million housing units set to receive mailings from the Census Bureau via a postal worker, there is a large burden on call centers (who answer questions from respondents via telephone), the internet instrument (where respondents input their answers to the census online), print vendors (who print the letters, postcards, and paper questionnaires), and USPS (who delivers the mailings to the housing units). The challenge for the Census Bureau, therefore, is to minimize these burdens. In order to do this, the 2020 Census will deliver the mailings to housing units in waves, or cohorts, instead of delivering each mailing to the whole country on a single day. This means that a portion of the population will receive Mailing 1 on one day and follow the mailing schedule (see Background section), a certain portion will receive Mailing 1 on a different day and follow the mailing schedule, etc.

The purpose of this study is to establish a method by which to separate self-responding housing units into mailing cohorts, such that there is an even distribution of cohorts throughout the country; the burden on the call centers, the print vendor, USPS, and the internet instrument is minimized; and the response rate is maximized. Using analyses and results from mid-decade testing at the Census Bureau, we examine the following aspects of cohort assignment: 1. The geographic level at which the cohorts should be assigned (e.g., census tract versus Basic Collection Unit); 2. Operational constraints; 3. Response Rates; and 4. The distribution of the cohorts.

## II. Background: Census Mailings

Housing units will receive up to five mailings (see Table 1) inviting them to self-respond. For the Internet First contact strategy (contact strategies described below), a housing unit will first receive a letter inviting them to complete the census online. Four days later, the housing unit will receive another letter reminding them to complete the census. If the housing unit does not self-respond, it will receive a reminder postcard 10 days after the second mailing. If there is still no response, it will receive another letter plus the paper census questionnaire 13 days after the postcard. If there is still no response, it will receive a final “It’s not too late” postcard 12 days later. The mailing strategy is the same for housing units in the Internet Choice contact strategy, with the exception that the first mailing contains both a letter and the paper census questionnaire. The timing of the five mailings is determined based on mid-decade testing (Phelan, 2016).

**Table 1**

| Contact Strategy | Mailing 1                | Mailing 2<br>4 days later | Mailing 3*<br>10 days later | Mailing 4*<br>13 days later | Mailing 5*<br>12 days later     |
|------------------|--------------------------|---------------------------|-----------------------------|-----------------------------|---------------------------------|
| Internet First   | Letter                   | Letter                    | Postcard                    | Mail Q’nnaire<br>+Letter    | “It’s not too<br>late” postcard |
| Internet Choice  | Mail Q’nnaire<br>+Letter | Letter                    | Postcard                    | Mail Q’nnaire<br>+Letter    | “It’s not too<br>late” postcard |

\*non-respondents, only

Contact strategy is assigned at the tract level. A tract is a census-defined level of geography, which contains, on average, about 1,500 housing units. The Internet Choice contact strategy is assigned

to any tract that, based on American Community Survey (ACS) self-response rates, responds by mail more than by internet and has at least one of the following additional attributes:

- Is a low responding tract when an Internet First strategy is used, based on ACS self-response rates
- Is a tract with higher older population, based on ACS estimates or
- Is a tract with less internet access, based on FCC data

Internet First is assigned to the remaining tracts. The reason for having different contact strategies is that certain areas may have a higher response rate if given the opportunity to complete a paper questionnaire from Mailing 1 rather than waiting to receive one in Mailing 4 (Census, 2018b).

Approximately 80% of the country will receive the Internet First contact strategy and 20% will receive the Internet Choice contact strategy. The concept of cohort will only apply to housing units that receive the Internet First contact strategy. The housing units that will receive the Internet Choice contact strategy will all follow the same mailing schedule.

### **III. Methodology:**

The first step is to determine the desired number of cohorts and the desired number of housing units assigned to each. The next step is to determine which housing units are assigned to which cohorts based on operational constraints and other considerations, such as response propensity and level of geography. These steps are described in detail below.

#### **Number of Cohorts and Allocation:**

The census mailings will take place in March and April. The following operational constraints need to hold when setting a mailing schedule:

- Mailing 1 of any cohort cannot land on a Monday. This is because call volumes peak on Mondays as well as on days that mailings are delivered. We want to spread the call response over a few days rather than have it peak on a Monday.
- Mailings should not be delivered on the weekend.
- Mailings should provide respondents with sufficient time to respond before Census Day, which is April 1, 2020.
- Mailings should not overlap with each other, as this will increase the operational burden.

Given the above constraints, the maximum number of cohorts attainable is four. In order to determine the desired number of cohorts, we used an internal response model that predicts daily and hourly incoming call activity based on historical response patterns and the size of each mailing (Census, 2015). The model uses mid-decade test response data and takes into consideration the effects that mailing delivery dates, day of week, and holidays have on response rates. Comparing different model estimates using different inputs, we find that assigning four cohorts, as shown in Figure 1, with the housing unit allocation shown in Table 2 provides for the smoothest call volume peaks. We also find that it is optimal to mail to Internet Choice housing units at the same time as Internet First, Cohort 2 housing units.

**Figure 1**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| March 2020 |            |             |    |            |             |    | April 2020 |            |             |            |            |           |    |
|------------|------------|-------------|----|------------|-------------|----|------------|------------|-------------|------------|------------|-----------|----|
| Su         | M          | T           | W  | Th         | F           | Sa | Su         | M          | T           | W          | Th         | F         | Sa |
| 1          | 2          | 3           | 4  | 5          | 6           | 7  | 29         | 30         | 31          | 1          | 2<br>M3C3  | 3<br>M3C4 | 4  |
| 8          | 9          | 10          | 11 | 12<br>M1C1 | 13<br>M1C2* | 14 | 5          | 6          | 7           | 8<br>M4C1  | 9<br>M4C2* | 10        | 11 |
| 15         | 16<br>M2C1 | 17<br>M2C2* | 18 | 19<br>M1C3 | 20<br>M1C4  | 21 | 12         | 13         | 14          | 15<br>M4C3 | 16<br>M4C4 | 17        | 18 |
| 22         | 23<br>M2C3 | 24<br>M2C4  | 25 | 26<br>M3C1 | 27<br>M3C2* | 28 | 19         | 20<br>M5C1 | 21<br>M5C2* | 22         | 23         | 24        | 25 |
| 29         | 30         | 31          |    |            |             |    | 26         | 27<br>M5C3 | 28<br>M5C4  | 29         | 30         | 1         | 2  |

\*Internet Choice follows the same schedule as Cohort 2

**Table 2\***

DRB approval number CBDRB-FY19-ACSO00J2-B0019

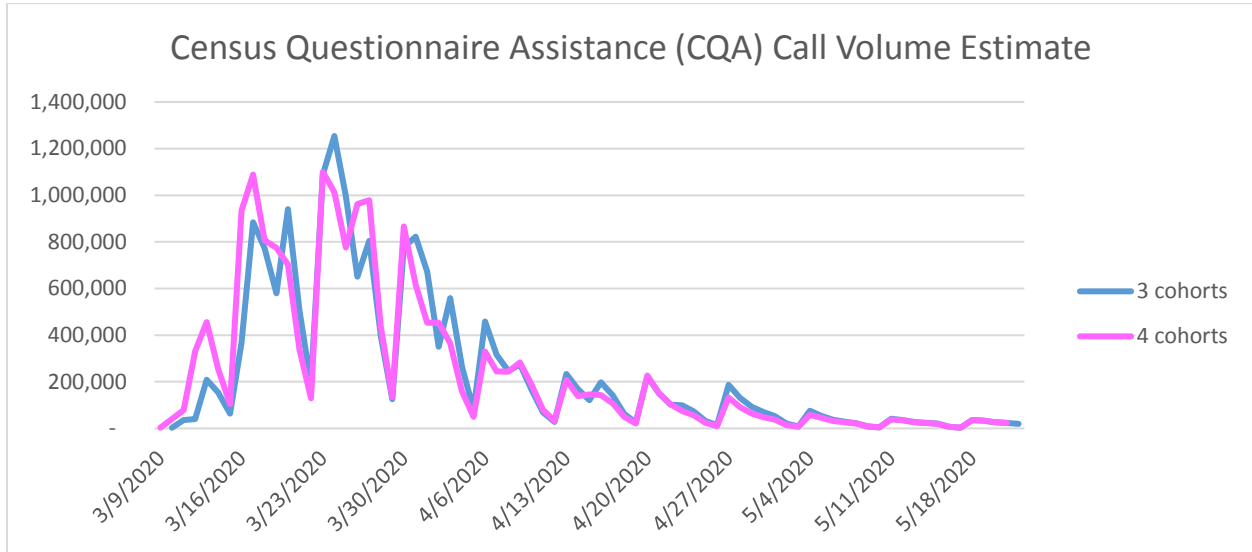
| Mailing Strategy        | Allocation |
|-------------------------|------------|
| Internet Choice         | 22%        |
| Internet First Cohort 1 | 23%        |
| Internet First Cohort 2 | 18%        |
| Internet First Cohort 3 | 20%        |
| Internet First Cohort 4 | 17%        |

\*The Internet Choice allocation is not an estimate, but a calculation of the number of Internet Choice housing units as a percentage of total housing units in scope for this paper

The following graph illustrates a comparison of call volume estimates generated by the internal model with three versus four cohorts, with their respective optimal allocations. It shows how four cohorts, as opposed to three, provide for smoother call volume peaks:

**Figure 2**

DRB approval number CBDRB-FY19-ACSO00J2-B0019



**Assignment of Housing Units to Cohorts:**

When determining which housing units go into which cohorts, we take in to account Early Non-Response Follow Up (ENRFU) status, In Field Address Canvassing (IFAC) status, projected response rates, geographic level, and distribution of demographic characteristics. These considerations are described in detail below.

**ENRFU:**

Housing units which do not self-respond end up forming the Non-Response Follow Up (NRFU) universe. Starting in mid-May, enumerators are sent to the housing units in the NRFU universe to collect the missing information (housing units may self-respond at any time, even after NRFU operations have begun). In areas such as college towns, however, the majority of residents (e.g., students living off campus) may have already left for the summer by the time the NRFU operation begins (Census, 2018a). As a result, they need to be counted via NRFU enumeration before all self-response mailings are completed, and are assigned to an operation called Early NRFU, which starts in early April. Because this is a population that needs to be reached earlier, it is assigned to Cohort 1.

**IFAC:**

The print vendor is tasked with printing and addressing hundreds of millions of mailings. They will begin printing and addressing mailings in November of 2019. A census operation that poses a challenge to the print vendor task is IFAC. During IFAC, address canvassers compare the most current list of addresses to the actual addresses in the field, and make corrections as necessary (e.g., the address list may reference “Apt. A” when it is actually “Apt. 1” or the census address list does not have a new house that was built recently). Areas are designated as IFAC if we expect that these corrections may need

to be made (e.g., areas that are experiencing a lot of growth, and hence a lot of new properties). Because IFAC corrections are not reflected until the final address file, which is delivered in January, and because the printing and mailing process is first-in-first-out, this is a population that needs to be reached later (Census, 2018a). Therefore, IFAC areas are assigned to Cohorts 3 and 4 (unless they are in ENRFU) and are the last set of addresses that the print vendor will receive.

### **Projected Response Rate:**

Results from a small scale test indicates that sending two mailings before Census Day results in a higher response rate than sending only one mailing before Census Day (Census, 2018c). Using this and the fact that an earlier cohort gives housing units more opportunity to self-respond before the start of NRFU (which is not staggered), we place tracts that have lower projected response rates into earlier cohorts. The response propensity is determined using an ordinary least squares regression model. Some of the core independent variables used to model response include ACS socioeconomic and demographic information such as age, household size, and internet and telephone accessibility.(Census, 2019).

### **Level of Geography:**

When deciding the geographic level at which to assign cohorts, we take into account how the above considerations can best be captured, as well as the burden on the print vendor and USPS. Because IFAC is assigned at the census Basic Collection unit (BCU, which contains anywhere from one to a few thousand housing units and is smaller than a census tract) level, and because there are very few census tracts that do not have BCUs in IFAC, we chose to assign cohorts at the BCU level. We do, however, need to take into consideration the fact that splitting cohorts based on larger, more contiguous areas than BCUs would reduce the burden on the print vendor and USPS. In order to account for this, we avoid putting BCUs that are located within the same tract in different cohorts when possible.

While we initially looked at more sophisticated methods of assigning cohorts (e.g., propensity score matching and balancing the effects of projected response and area size), we ultimately decided that the best approach is to assign cohorts based on simple operational constraints, such as the ones mentioned above.

## **Estimates and Results**

### **Determine the number of housing units in each cohort**

First, we count the total number of housing units that we will mail to (estimated at 143.2 million as of July, 2019), and then look at the predetermined distribution of cohorts to determine how many households should go into each cohort. Estimates presented below are rounded for disclosure avoidance purposes.

**Table 3**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| Contact Strategy | Pre-determined percentage of households | Number of desired households (based on the 143.2 million estimate) |
|------------------|---|--|
| Internet Choice  | 22%                                     | 31.4 million *   |
| Cohort 1         | 23%                                     | 32.9 million   |
| Cohort 2         | 18%                                     | 25.8 million   |
| Cohort 3         | 20%                                     | 28.6 million   |
| Cohort 4         | 17%                                     | 24.3 million   |

\*This number will not be based on an estimate – it will be the number of self-responding housing units in Internet Choice

### Assign Cohorts 1 and 2

We then determine the population eligible for Cohorts 1 and 2. To do this, we delete all observations that have the Internet Choice contact strategy, as well as all observations that are in IFAC, but NOT in ENRFU. This becomes our population eligible for Cohorts 1 and 2. Because we want to keep tracts together as much as possible, we split the above population into cohorts at the tract level.

First, we put all tracts that have at least one BCU in ENRFU into Cohort 1. We then subtract the number of housing units in these tracts from the desired number of housing units in Cohorts 1 and 2 (32.9 million + 25.8 million) to determine the remaining number of housing units that still need to be assigned to Cohorts 1 and 2 from the eligible population. We want to have a disproportionately large number of low responders in Cohorts 1 and 2, so we pick from the tracts with the lowest average response rate until the desired number of housing units is (approximately) achieved. The tracts from the population eligible for Cohorts 1 and 2 that do not make this cutoff will be referred to as leftovers, and will be used later in the process.

We then assign the chosen tracts to Cohort 1 and Cohort 2. We sort by variables of interest (census region, state, bilingual mailing materials assignment, number of housing units) in order to get an even distribution between the two cohorts. We then use systematic random sampling (where, given sorted data, the procedure picks an observation from the beginning of the data set, and then the end of the data set, and then the beginning of the data set, etc., which prevents the possibility that all observations selected for one cohort are at the top of the data set/have disproportionately high or low values of certain variables) to pick the tracts that go into Cohort 2. The remaining tracts go into Cohort 1, along with the ENRFU tracts.

### Assign Cohorts 3 and 4

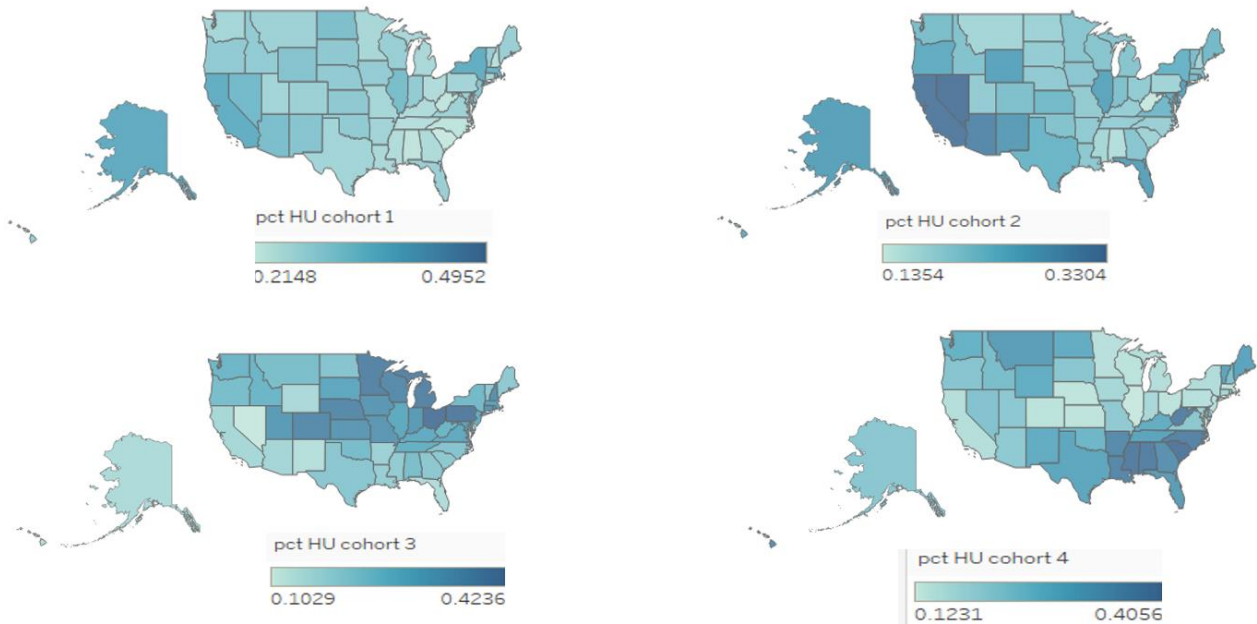
The eligible population for Cohorts 3 and 4 consists of the BCUs that are in IFAC, but NOT in ENRFU, as well as the leftovers from Cohorts 1 and 2. We first reunite tracts that were defined as leftover when assigning cohorts 1 or 2 with their corresponding BCUs that were not eligible for cohorts 1 or 2 because of IFAC status. We put these into Cohort 3. The remaining eligible tracts will only have BCUs that are in IFAC. Similar to the Cohorts 1 and 2 assignment, we use systematic random sampling to pick the tracts that go into Cohort 4, and then put the remainder into Cohort 3.

**Final cohort assignment**

We merge the Cohorts 1 and 2 assignments with the Cohorts 3 and 4 assignments and the Internet Choice population to get the final cohort assignment. Figure 3, below, illustrates the distribution of cohorts.

**Figure 3**

DRB approval number CBDRB-FY19-ACSO00J2-B0019



We then check to make sure the following assumptions hold:

- The housing unit numbers should closely match those in Table 3
- IFAC should not be present in Cohorts 1 or 2, unless the housing unit is also in ENRFU
- ENRFU should only be present in Internet Choice or Cohort 1
- Contact Strategy for Cohorts 1-4 should be Internet First
- Projected Response Rate should be lower for Cohorts 1 and 2 than for other cohorts

**Table 4: Final Cohort Assignment**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| Cohort          | Number of Housing Units |
|-----------------|-------------------------|
| Internet Choice | 31.2 million            |
| 1               | 32.9 million            |
| 2               | 25.8 million            |
| 3               | 28.9 million            |
| 4               | 24.3 million            |



**Table 5: Census Regions**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| Cohort          | Census Region | Number of Housing Units |
|-----------------|---------------|-------------------------|
| Internet Choice | midwest       | 6.5 million             |
| Internet Choice | northeast     | 4.4 million             |
| Internet Choice | south         | 16.5 million            |
| Internet Choice | west          | 3.8 million             |
| 1               | midwest       | 7 million               |
| 1               | northeast     | 6.6 million             |
| 1               | south         | 10.4 million            |
| 1               | west          | 8.8 million             |
| 2               | midwest       | 5.2 million             |
| 2               | northeast     | 4.5 million             |
| 2               | south         | 8.6 million             |
| 2               | west          | 7.5 million             |
| 3               | midwest       | 8.9 million             |
| 3               | northeast     | 6.3 million             |
| 3               | south         | 8.4 million             |
| 3               | west          | 5.3 million             |
| 4               | midwest       | 3.8 million             |
| 4               | northeast     | 3.2 million             |
| 4               | south         | 12.2 million            |
| 4               | west          | 5.1 million             |

**Table 6: Bilingual**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| Cohort          | Bilingual | Number of Housing Units |
|-----------------|-----------|-------------------------|
| Internet Choice | 0         | 25.9 million            |
| Internet Choice | 1         | 5.2 million             |
| 1               | 0         | 29.7 million            |
| 1               | 1         | 3.2 million             |
| 2               | 0         | 22.9 million            |
| 2               | 1         | 2.9 million             |
| 3               | 0         | 28.6 million            |
| 3               | 1         | 0.4 million             |
| 4               | 0         | 22.7 million            |
| 4               | 1         | 1.7 million             |

**Table 7: IFAC**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| Cohort          | IFAC | Number of Housing Units |
|-----------------|------|-------------------------|
| Internet Choice | 0    | 18.9 million            |
| Internet Choice | 1    | 12.3 million            |
| 1               | 0    | 31.9 million            |
| 1               | 1    | 1.1 million             |
| 2               | 0    | 25.8 million            |
| 3               | 0    | 16.7 million            |
| 3               | 1    | 12.2 million            |
| 4               | 1    | 24.4 million            |

**Followup: Details about Cohort 1\***

|          | No IFAC      | IFAC         |
|----------|--------------|--------------|
| No ENRFU | 1 million    | 0            |
| ENRFU    | 0.07 million | .014 million |

\*no IFAC = 1 in cohort 2

**Table 8: ENRFU**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| Cohort          | ENRFU | Number of Housing Units |
|-----------------|-------|-------------------------|
| Internet Choice | 0     | 30.6 million            |
| Internet Choice | 1     | 0.6 million             |
| 1               | 0     | 30 million              |
| 1               | 1     | 4 million               |
| 2               | 0     | 25.8 million            |
| 3               | 0     | 28.9 million            |
| 4               | 0     | 24.3 million            |

**Table 9: Projected Response Rate**

DRB approval number CBDRB-FY19-ACSO00J2-B0019

| cohort          | mean projected response rate |
|-----------------|------------------------------|
| Internet Choice | 54.6%                        |
| 1               | 65.8%                        |
| 2               | 65.3%                        |
| 3               | 82.1%                        |
| 4               | 64.4%                        |

## Conclusion

Due to the high number of self-responding housing units that are set to receive mailings from the Census Bureau via USPS, census operations have to take into account the burden that this places on call centers, USPS, the Internet Instrument and the print vendor. In this study, we address this challenge, and develop a method by which to assign self-responding households to mailing cohorts for the 2020 Census. Using analyses and results from mid-decade testing at the Census Bureau, we assign cohorts such that there is an even distribution of cohorts throughout the country, the operational burden is minimized, and the response rate is maximized.

## References

- U.S. Census Bureau (2015). External Demand Model for Internet, Telephony, and Paper, Version 2.1 (Internal document)
- Phelan, J. (2016). 2020 Research and Testing: 2015 National Content Test Optimizing Self-Response Report. U.S. Census Bureau: 2020 Census Program Internal Memorandum Series 2016.57.i. November 22, 2016.
- U.S. Census Bureau (2018a). 2020 Operational Plan, Version 4. Retrieved from: <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan4.pdf>. September 12, 2019
- U.S. Census Bureau (2018b). Determining Internet Choice Areas for the 2020 Census, Version 2 (Internal document)
- U.S. Census Bureau (2018c). 2020 Research and Testing: The Effect of the Mail Delivery Date on Survey Login Rates and Helpline Call Rates Analysis Report. (Internal Document)
- U.S. Census Bureau (2019). 2020 Census Predictive Models and Audience Segmentation Report, Version 1.5. Retrieved from: <https://www2.census.gov/programs-surveys/decennial/2020/program-management/census-research/predictive-models-audience-segmentation-report.pdf> September 12, 2019