UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

November 6, 2017

2017 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS17-RER-16

MEMORANDUM FOR ACS Research and Evaluation Workgroup

From: Victoria Velkoff
Chief, American Community Survey Office

Prepared by: Gregory J. Mills
Survey Analytics and Measures Branch
American Community Survey Office

Subject: Investigating the Use of Cellular Telephone Numbers for the Computer-Assisted Telephone Interview Operation in the American Community Survey

Attached is the American Community Survey Research and Evaluation report, "Investigating the Use of Cellular Telephone Numbers for the Computer-Assisted Telephone Interview Operation in the American Community Survey." The Census Bureau undertook this research to better understand the effects of including cellular telephones along with landlines in the Computer-Assisted Telephone Interview nonresponse followup mode. Using a split panel experimental design methodology, this research first tested the effects of including cellular telephones on several survey productivity and efficiency metrics. Next, it tested the effects of including cellular numbers on telephone contactability. Finally, it looked into differences in demographic characteristics between households interviewed by landline and cellular telephones.

If you have any questions about this report, please contact Gregory Mills at 301-763-6172 or David Raglin at 301-763-4226.

Attachment

cc:
Jennifer Ortman (ACSO)
David Raglin (ACSO)
Jennifer Reichert (ACSO)
Elizabeth May Nichols (CSM)
Asaph Young Chun (DSSD)
Elizabeth Poehler (DSSD)
Anthony Tersine (DSSD)

Page intentionally left blank.

November 6, 2017

Investigating the Use of Cellular Telephone Numbers for the Computer-Assisted Telephone Interview Operation in the American Community Survey

FINAL REPORT

United States Census Bureau

Gregory J. Mills

American Community Survey Office

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The American Community Survey (ACS) is a multi-modal household survey administered by the U.S. Census Bureau. Prior to January 2017, the ACS Computer-Assisted Telephone Interview (CATI) operation exclusively focused efforts on calling landline telephone numbers. However, the number of interviews typically obtained in a CATI month decreased by over 50 percent between 2012 and 2016 (Mills, 2016b). Other large surveys have reported similar trends in recent years (Dutwin & Lavrakas, 2016).

In January 2017, the Census Bureau conducted a split-panel experiment to examine the effects of including cellular telephone numbers in the ACS CATI workload. In this experiment, the test group included a mixture of both landline and cellular numbers, while the control group included only landline numbers. Various outcome measures were then compared between the two groups. This report documents the findings from this experiment.

First, test group cases, which included cases with cellular numbers, had a higher overall CATI contact rate (31.4 vs. 26.3 percent), as well as a higher CATI interview completion rate (12.6 vs. 10.4 percent) compared to the landline-only control group. Second, cases that were called at least once in the test group were more likely to self-respond during the CATI month (12.8 vs. 11.4 percent). This was likely due to a higher contact rate, which was shown to be correlated with higher late self-response return rates as the telephone contact served as an additional reminder to self-respond (Griffin, Fischer, & Morgan, 2001; Nichols, Horwitz, & Tancreto, 2013).

Telephone numbers from the test group had better overall contactability, which included all categories of contact as well as lower rates of disconnected telephone numbers.[1] This was shown when considering all telephone numbers aggregated, as well as when focusing on second-best number specifically.[2] When comparing landline and cellular numbers directly, call attempts to landlines were more likely to lead to a complete interview, but made lower contact overall and had a higher proportion of out-of-service numbers. However, general contactability improved with the addition of cellular telephone numbers.

Finally, there were several demographic differences between people and households included in interviews from landline and cellular telephones. Landline telephone numbers were shown to more often collect interviews from owner-occupied households, single person households, and reach households with residents that are older in age. Cellular numbers were more likely than landlines to reach renter-occupied households and households that included people that were younger and more racially and ethnically diverse.

[1] In this report, the term "contactability" refers to the level of contact a particular telephone number is able to make with a sample unit. It is measured on a hierarchy and ranges from a complete interview, to being disconnected or unable to make contact.

[2] For cases that have more than one highly scored telephone number, the ACS sends two numbers. This equates to about 10 percent of CATI cases.

1. INTRODUCTION

The American Community Survey (ACS) is a multi-modal household survey carried out by the U.S. Census Bureau. After a self-response period, nonresponse followup is attempted with a Computer-Assisted Telephone Interview (CATI), and then with a Computer-Assisted Personal Interview (CAPI) for a subsample of cases. Since its inception, the CATI mode has exclusively focused efforts on calling landline telephones, which in recent years has been hampered by decreasing efficiency in obtaining interviews and a higher share of sample units that are unreachable by telephone. This research is part of a broader effort to improve the efficiency and cost-effectiveness of the ACS CATI operation

Between 2012 and 2016, the number of interviews typically obtained in a CATI month decreased by over 50 percent (Mills, 2016b). This is a common trend among telephone surveys, and particularly for landline telephone-based surveys. Several large survey firms that use landlines report dramatic declines in response rates over the past decade, from an average of 15.7 percent in 2008 to 9.3 percent in 2015 (Dutwin & Lavrakas, 2016). Landline coverage has also decreased in recent years, with almost 50 percent of adults and 60 percent of children living in households that did not have a landline telephone during the first half of 2016 (Blumberg & Luke, 2016). These factors suggest that at least part of the decline in CATI productivity might be due to exclusively calling landline telephone numbers.

This report explores the effect of including cellular telephone numbers in the CATI workload, using a two-group, randomized experimental design in the December 2016 ACS panel.[3] The test group included a mixture of both landline and cellular numbers, while the control group included only landline numbers, which is the current ACS CATI methodology. These two groups were proportionally represented in the CATI workload, enabling a comparison of their respective productivity and efficiency metrics.

2. BACKGROUND

The ACS sample for each year is split into twelve roughly equal-sized panels that loosely correlate to the twelve calendar months of the year. During each panel's first month of data collection, respondents are encouraged to self-respond via an internet instrument, and later in the month via a paper questionnaire. Prior to October 1, 2017, if a response was not received and a landline telephone was matched to the address, a CATI interviewer would attempt to contact a respondent in the housing unit during the second month of data collection via telephone.[4] At the end of the CATI month, a subsample of cases for which there was still no completed interview was selected for CAPI.

As of January 2017, the CATI operation fielded 95,000 cases each month, which were selected from a ranked list of sample unit/telephone number matches based on the probability of the

[3] The January 2017 CATI workload corresponds to the December 2016 ACS panel, as the CATI mode takes place in the second month of a panel's data collection period.

[4] Beginning in March 2017, the CATI workload underwent a series of size reductions due to its overall continuing inefficiency. The ACS CATI nonresponse followup operation was eliminated on October 1, 2017.

telephone number being associated with the sample unit.[5] Various third-party sources of telephone number data were used to identify sample unit/telephone number matches and a quality score was assigned to each pair. This score was based on factors such as the particular data source's propensity to match respondent-provided phone numbers in previous ACS panels and the quantity and recency of sources in which a particular telephone number was confirmed. This scoring process was carried out for both the test and control groups. In the model, the actual scores were less meaningful than the relative ranking of matched cases. After the self-response month was completed, cases that had self-responded were removed from the workload and the top-ranked 95,000 remaining cases were sent to CATI. From its inception, the CATI operation attempted to call only landlines. In an effort to improve declining CATI productivity, the Census Bureau tested including cellular telephones in the ACS CATI workload in January 2017, the results of which are detailed in this report.

3. LITERATURE REVIEW

From its inception, the ACS CATI operation called only landlines, because the ACS sample is address-based. However, the proportion of the population with access to a landline telephone has been steadily falling. In 2015, the National Center for Health Statistics estimated that almost 50 percent of U.S. adults and 60 percent of U.S. children live in households without landline telephone coverage, which is up from only 10 percent in 2005 (Blumberg & Luke, 2016). The obvious effect of this has been a sharp decline in the capacity for the ACS CATI operation to maintain its past efficiency. Response and contact rates in the CATI operation have been falling for some time (Mills, 2016b), which is a trend seen across telephone surveys (Groves & Couper, 2012). For example, several large survey firms that use landlines report dramatic declines in response rates over the past decade, from an average of 15.7 percent in 2008 to 9.3 percent in 2015 (Dutwin & Lavrakas, 2016). Landline coverage also tends to favor certain demographic groups. For example, a 2015 survey conducted by the Pew Research Center showed that occupants of cellular-only households are more likely to be younger, Hispanic, poor, and reside in urban areas (McGreeney, 2016).

For these reasons, many surveys that include a telephone response mode have increased efforts to include cellular telephone numbers in their sample frame (Dutwin & Lavrakas, 2016). The Census Bureau tested cellular phones in the ACS CATI operation with the hopes of improving CATI response rates. However, past work cautions against assuming that including cellular phones in a sample frame will inherently improve productivity in CATI. First, although refusal rates are rising among landlines (Dutwin & Lavrakas, 2016), research consistently shows higher rates of refusals and fewer refusal conversions among cellular phones compared to landlines, in addition to higher nonresponse rates in general (AAPOR, 2010). This could be due to differences in the way people treat cellular phones and landlines. For example, past work suggests that about one-third of interviews completed on cellular telephones are completed while the respondent is away from home.

On the other hand, the ACS might be shielded from some of these more widely documented issues. For one, initial analyses suggested that even if fully implemented, cellular phones would

[5] About 700 of the 95,000 cases were from Puerto Rico, which were sent regardless of their model score or rank.

likely replace a landline as the best telephone number for 15 to 30 percent of cases in the workload.[6] Second, although some reshuffling of the ranked telephone numbers would occur upon making cellular numbers eligible, it has been well established that the lowest scored landline numbers in the CATI workload had very low productivity in terms of making contact or completing a CATI interview (Mills, 2016b). Replacing landline numbers that were shown to have low productivity minimizes the associated opportunity cost of no longer calling them (due to potentially being ranked lower than the top 95,000 cases). Thus, even if the newly included cellular phones had higher than expected nonresponse or noncontact rates, it is likely that the landline cases they would have replaced would also have had elevated nonresponse, nullifying any overall negative effects on CATI productivity. Lastly, some researchers have suggested that making contact with a case in a CATI followup mode spurs additional self-response (Griffin, Fischer & Morgan, 2001; Nichols, Horwitz, & Tancreto, 2013). The contact in these cases served as an additional reminder to complete the survey. These additional responses were a benefit of the CATI operation, as they were cheaper to obtain than a CATI interview and also reduced the number of cases that go to CAPI for additional followup.

4. RESEARCH QUESTIONS AND ANALYSIS METHODOLOGY

4.1 Test Group Construction

The ACS tested the use of cellular telephone numbers for the January 2017 CATI operation. To do this, a test group was created out of six methods panels groups, which equates to one-fourth of the panel's total cases. Each of the ACS monthly panels is separated into 24 methods panel groups, which allows the ACS to form randomized experimental groups and follow them throughout the data collection period. The six methods panel groups that were assigned to the test group used the experimental methodology of including both landline and cellular numbers, while the remaining 18 methods panel groups that were assigned to the control group included only landline numbers. Since the test group was made up of one-fourth of the overall panel size, it was also allocated one-fourth of the CATI workload. Typically, the top scoring 95,000 cases were sent to the CATI operation. In this test, the CATI workload was comprised of the top approximately 70,775 scored cases from the control group (three-quarters of the total stateside workload), along with the top 23,525 scored cases from the test group (one-quarter of the total stateside workload).[7]

The randomized experimental design of the methods panel groups allowed for the statistical comparison of various outcomes and characteristics between the control and test groups.[8] Cases in the control group were matched with the top-scoring landline number, as had been customarily

[6] It should be noted that in the CATI operation there was not 100 percent certainty of whether a number belonged to a landline or cellular telephone. For one, telephone number mobility enables numbers to shift from one type to another rather seamlessly. Second, call forwarding services enable people to receive calls from a landline, for example, on a cellular telephone. Lastly, there was a small degree of error in the telephone number data, where some numbers were incorrectly flagged as landline or cellular due to being old data or some other reason.

[7] The roughly 700 cases that were sent to Puerto Rico were not part of this experiment. Thus, the sum of the test and control cases as described here is 94,300.

[8] For the purposes of this paper, contact and completion rates differ from response rates in that they are unweighted and are predominantly used for operational, data collection purposes. Completion rates cannot be assumed to be reflective of the population in general, just the cases that ACS sends to CATI.

done. The test group was matched with the top scoring telephone number, which could have been either a cellular or landline number. Finally, it is also important to note that in the test group, any included cases with cellular numbers replaced the cases with the lowest quality landline numbers. Thus, any group-level differences were likely due to the combination of including cellular numbers and excluding the lowest quality landline numbers in the test group.

4.2 Research Questions

The research questions for this paper can be categorized into three themes. The first and overarching theme of this research was to see how the addition of cellular telephones to the CATI workload affected productivity and efficiency. Answers to these questions were determined by comparing case-level interview completion rates, CATI response rates, and late self-response return rates between the test and control groups; and relate to Research Questions 1, 2, and 3.[9] The second set of research questions examined the overall quality of the telephone numbers in terms of contactability and completion rates, and also their associated household demographic characteristics and contact propensity at various times of day. These questions were analyzed at the telephone number level in Research Questions 4 through 9. Third, Research Questions 10 and 11 center on how the contact centers adjusted to additional calling protocol introduced with the cellular telephones, such as asking if the respondent can safely talk. The specific research questions are introduced below, and methodology is detailed in the next section.

- Research Question 1. What are the unweighted CATI contact, completion, and completion-given-contact rates for the control and test groups? Are the rates statistically different between the control and test groups?

- Research Question 2. What are the weighted CATI response rates of the control and test groups? Are the rates statistically different between the control and test groups?

- Research Question 3. How do unweighted late self-response rates compare between the control and test group, among cases that were called at least once? Are the rates statistically different between the control and test groups?

- Research Question 4. What is the distribution of contactability in the control and test groups? Are the contactability rates statistically different between the control and test groups? How does contactability vary by model score?

- Research Question 5. What is the distribution of contactability for second-best numbers in the control and test groups? Are the contactability rates statistically different between the control and test groups? How does contactability vary by model score?

[9] Late self-responses are those that are received via internet or mail after the self-response time-period. The self-response period is roughly four weeks after the first mailing.

- Research Question 6. What is the distribution of contactability among landline and cellular telephone numbers? Are the contactability rates statistically different between landline and cellular telephone numbers? How does contactability vary by model score?

- Research Question 7. For cellular and landline telephone numbers, how is CATI contactability distributed across number of call attempts? Specifically, what is the distribution of call attempt totals within CATI outcomes? What is the distribution of final CATI case outcomes by the number of call attempts?

- Research Question 8. Do call attempts from cellular and landline telephones make contact at different rates at different times-of-day? Are these rates statistically different for landline and cellular numbers?

- Research Question 9. Do demographic characteristics differ between households that completed an interview with a cellular phone and those who completed an interview on a landline phone in terms of age, race, Hispanic origin, family type, tenure, and household income?

- Research Question 10. How often do people report safety or privacy issues? At what rate did interviewers utilize a new software feature enabling the addition of a new telephone number? Are these rates statistically different for landline and cellular numbers?

- Research Question 11. What are the main findings from the contact center debriefing with the supervisors and monitors regarding the cellular phone training and operational questions?

4.3 Data and Metrics

This section includes descriptions of the data and metrics used to answer the research questions detailed above. Variances were calculated using the Successive Differences Replication method with replicate weights or unweighted replicate factors where appropriate, which is the standard method used in the ACS (see U.S. Census Bureau, 2014, Chapter 12). The variance for each rate and difference was calculated with the formula below. The standard error of the estimate (X_0) is the square root of the variance:

$$\mathrm{Var}(X_0) = \frac{4}{80}\sum_{r=1}^{80}(X_r - X_0)^2$$

where:

X_0 = the estimate calculated using the full sample,
X_r = the estimate calculated for replicate r.

For each rate, the difference was calculated between the test and control groups and margins of error for both rates and the difference between the rates. The rates of the test and control groups were statistically different if the margin of error of the difference was smaller than the respective difference. All statistical tests were performed using two-tailed tests for differences at the $\alpha = 0.10$ level. To statistically compare unweighted rates between the test and control groups,

margins of error were constructed by creating replicate factors by dividing replicate weight by the base weight.

- Research Question 1. What are the unweighted CATI contact, completion, and completion-given-contact rates for the control and test groups? Are the rates statistically different between the control and test groups?

A contact was defined as having reached the sampled housing unit via telephone at least once during the CATI month. This included the call outcomes of a completed interview, a noninterview verified contact, and a refusal. The interview completion rate is the rate at which at least a sufficient partial interview was obtained at a housing unit.[10] The third metric measures the rate of completion among cases with which contact was made.

Data to answer this research question came from the CATI transaction file, which contains paradata about the CATI interviews, and was delivered to the ACS each month at the termination of CATI calling. The CATI transaction file contains a log of call and contact history for every case for the entire month of calling. While the transaction file does not distinguish between landline and cellular telephone numbers specifically (although it does provide the actual telephone number dialed for each call attempt), number type information was received with the telephone number files, which was then merged to the transaction file to facilitate the analysis.

All measures for Research Question 1 were case-level and unweighted. For contact and completion rates, the denominator included all CATI cases that received at least one call attempt and did not culminate with a self-response. Late self-responses were removed from the denominators for each of these measures. The denominator for the completion-given-contact rate included all contacted cases, which was the numerator in the contact rate.

$$\text{[1.1] Contact Rate} = \frac{\text{\# of cases that made contact at least once}}{\text{\# of cases that received at least 1 call attempt} - \text{\# of late self-responses}} * 100$$

$$\text{[1.2] Completion Rate} = \frac{\text{\# of cases that received a sufficient partial or completed CATI interview}}{\text{\# of cases that received at least 1 call attempt} - \text{\# of late self-responses}} * 100$$

[10] In the ACS, a sufficient partial interview is incomplete, but progressed through the basic demographic and housing questions, and at least to the first detailed question of the first person (place of birth).

$$\text{[1.3] Completion-given-contact Rate} = \frac{\text{\# of cases that received a sufficient partial or completed CATI interview}}{\text{\# of cases that made contact at least once} - \text{\# of late self-responses}} * 100$$

- Research Question 2. What are the weighted CATI response rates of the control and test groups? Are the rates statistically different between the control and test groups?

This measure is similar to the completion rate from Research Question 1 with a few differences. The response rate excluded: 1) cases that were deemed to be businesses, 2) vacant and temporarily occupied households, and 3) CATI cases that were subsampled out and excluded from CAPI. Finally, CATI response rates were weighted for the probability of selection, including CAPI subsampling, whereas CATI completion rates were unweighted.[11] Data for this research question came from the CATI case status file, which was also delivered monthly after the termination of the CATI month.

$$\text{[2.1] Final CATI Response Rate} = \frac{\text{\# of cases received a sufficient partial or completed interview}}{\text{\# of cases in CATI workload} - \text{\# of late self-responses} - \text{\# of cases that are subsampled out and excluded from CAPI} - \text{\# of cases deemed to be businesses, vacants, temporarily occupied, or otherwise ineligible}} * 100$$

- Research Question 3. How do unweighted late self-response rates compare between the control and test group cases among cases that were called at least once?

The denominator included all cases that were called at least once during the CATI month. Late self-responses that were received before calling began or before the case was called by a CATI interviewer were excluded from both the numerator and denominator. Self-response rates in this analysis were unweighted.

$$\text{[3.1] Late Self-Response Rate} = \frac{\text{\# of cases that received a late self-response}}{\text{\# of cases where at least 1 call attempt made contact}} * 100$$

[11] In the CAPI mode, roughly two of three cases that have yet to respond are excluded in order to more cost effectively perform further nonresponse followup. To account for this when calculating the CATI response rate, cases that are excluded from the CAPI mode are removed from the CATI response rate denominator, and the remaining nonresponding cases are given additional weight, which is reflective of how the more general ACS response rate is calculated.

- Research Question 4. What is the distribution of contactability in the control and test groups? Are the contactability rates statistically different between the control and test groups? How does contactability vary by model score?

- Research Question 5. What is the distribution of contactability for second-best numbers in the control and test groups? Are the contactability rates statistically different between the control and test groups? How does contactability vary by model score?

- Research Question 6. What is the distribution of contactability among landline and cellular telephone numbers? Are the contactability rates statistically different between landline and cellular telephone numbers? How does contactability vary by model score?

For Research Questions 4 through 6, this paper used an index of contactability (also described in Mills, 2016a; 2016b) to calculate the rates at which specific final dispositions occurred. To obtain each phone number's level of contactability, outcomes from each phone call for each telephone number on the transaction file were sorted into a hierarchical contactability index. Considering the cumulative history of each telephone number over the course of the CATI month, only the highest quality outcome that occurred for each phone number was used as that number's contactability. For example, consider the hypothetical case in which the first call for a particular telephone number resulted in contact, but contact was not made on any subsequent attempts. In that case, the phone number would be categorized as a verified telephone number (but not a sufficient interview). Contactability has six mutually exclusive categories, listed in order of descending outcome quality:

Contact:
1. Complete or sufficient interview
2. Verified telephone number (but not a sufficient interview)
3. Refusal (including immediate hangup)

Noncontact:
4. Not enough information
5. Not in service
6. Bad telephone number (ineligible or other)

A phone number outcome was categorized as "complete or sufficient interview" if at some point during the CATI month the interviewer obtained a complete or sufficient partial interview from a call attempt that involved the given number. An outcome of "verified telephone number" indicates that an interviewer made contact at the correct address, but did not complete an interview. An example of this is if the interviewer verified that the phone number is for the sampled address, but the respondent claimed to have already mailed in the paper questionnaire. A "refusal" outcome occurred if the respondent or the person who answered the associated telephone number refused to participate. This category included immediate hangups. These three groups were considered "contact," and are referred to as such throughout this report.

Outcomes classified as "not enough information" were those where not enough information was received to decipher the quality, such as a ring with no answer or voicemail. Numbers that were not in service had been disconnected or were otherwise not in service when dialed. Other "bad"

telephone numbers were those that were not able to be completed as dialed or resulted in contact with ineligible or out of scope households. These three categories were considered "noncontact," and are henceforth referred to as such in this report.

The above questions ask about the distribution of contactability. Research Question 4 focuses on comparing the test and control panels. The universe included every telephone number that was sent to CATI and called at least once, and compared contactability rates for the test and control groups. Research Question 5 involves an additional analysis of second-best telephone numbers within the test and control groups, to compare the added value of making cellular numbers eligible for the second-best numbers. In both the test and control groups, second-best numbers were included for about 10 percent of cases, but in each case, a second-best number was only called if the first number was closed out before using a case's totality of call attempts. In the control group, only landlines were considered when sending second-best numbers. For the test group, the second-best number could be a cellular or landline number. Finally, Research Question 6 compares the contactability among aggregated landline and cellular telephone numbers. The landline numbers included those from both the test and control groups. In general, past research has found somewhat higher nonresponse and refusal rates for cellular telephones (Brick et al., 2007). On the other hand, efforts have been made to restrict the ACS CATI workload to cases with an elevated quality score. From February 2016 onwards, the CATI operation used a model to determine which telephone numbers were most likely to reach the correct household. Each telephone number was given a quality score and rank within the workload. The same scoring model was used for both cellular and landline telephone numbers, enabling us to compare their respective relationships with contactability outcomes.

Additionally, the contactability data in each research question were visually explored by grouping the numbers into ranked score groups, and calculating the distribution of contactability across score groups. It is important to note that these figures are not meant to statistically compare contactability by score decile, but to show general contactability trends within the groups compared in each research question.

- Research Question 7. For cellular and landline telephone numbers, how is CATI contactability distributed across number of call attempts? Specifically, what is the distribution of call attempt totals within CATI outcomes? What is the distribution of final CATI case outcomes by the number of call attempts?

This analysis compared the relationships between the number of call attempts and call outcomes for landline and cellular telephone numbers. These results give us insight into the general trends of phone number quality among cellular telephone numbers, and reveal potential differences that exist with landline numbers in terms of achieving certain case dispositions. For example, it might be that cellular telephone numbers result in more refusals than landlines, but that refusals from cellular numbers achieve their final disposition with fewer call attempts.

Data about the call attempts came from the CATI transaction file. The analysis included all telephone numbers sent to CATI that were called at least once, and that did not culminate with a late self-response during the CATI month. To focus on outcomes by telephone number type within cases, the analysis was restricted to the best telephone number for each case, and all call

details from second-best phone numbers were excluded. Additionally, the analysis was unweighted and without statistical tests for significance, and was thus meant to be exploratory, since the use of cellular numbers and landline numbers for the test treatment were not assigned randomly (they were assigned based on the model score). The number of callback attempts associated with each telephone number was the total number of callback attempts found in the CATI transaction file, or the first callback attempt in which the case received at least a sufficient partial interview.

- Research Question 8. Do call attempts from cellular and landline telephones make contact at different rates at different times-of-day? Are these rates statistically different for landline and cellular numbers?

This analysis aimed to find out if type if call attempts made from cellular or landline telephones have different rates of making contact, at various times-of-day or days-of-the-week. For example, it might be that call attempts made to landline numbers are more successful at making contact in the evening versus attempts from cellular numbers. Call attempts were sorted by weekdays and weekends, and categorized into four time periods: 1) morning (9:00am – 11:59am), 2) early afternoon (12:00pm – 2:59pm), 3) late afternoon (3:00pm – 5:59pm), and 4) evening (6:00pm – 9:00pm). Calls were categorized into a time-of-day/day-of-the-week combination based on when the call was placed in the sample unit's time zone.

Each time-of-day/day-of-the-week combination had an associated "contact rate" for both cellular and landline telephones. Each denominator included every call attempt that was placed or received by interviewers during the CATI month, as found on the CATI transaction file. The numerator included call attempts during the given time-of-day/day-of-the-week that made contact, which could include a completed interview, any verified contact, or a refusal. The data from this analysis were unweighted. Differences between cellular and landline contact rates by times of day were tested for statistical significance using a two-tailed test for differences.

[8.1] Time/Day *x* Contact Rate $= \frac{\text{\# of call attempts during time/day } x \text{ that make contact}}{\text{\# of call attempts during time/day } x} * 100$

- Research Question 9. Do demographic characteristics differ between households that completed an interview with a cellular phone and those who completed an interview on a landline phone in terms of age, race, Hispanic origin, family type, tenure, and household income?

Past research indicates that occupants of households without a landline telephone are more likely to be younger, Hispanic, and have more children (Hall, Carlson, & Cybulski, 2011). However, as landline household coverage has continued to erode over the past decade (Blumberg & Luke, 2016), it is possible that households that only have a cellular telephone are more representative of the general population or more diverse those with landlines. Thus, introducing cellular telephone numbers to the CATI workload might change the characteristics of the respondents.

Response data in this analysis were weighted and came from the ACS data capture file (DCF), which consists of response data from all modes that have not gone through edits or imputations. This analysis included three household-level and three person-level variables, including age, race, ethnicity, family type, tenure, and household income. For all variables except household income, nonresponses and refusals were removed from the denominator. Household income included a household in the denominator only if all eligible individuals within the household had a valid total income response. Total household income was calculated by aggregating the total income for all eligible individuals. If any eligible individuals had a nonresponse or refusal for personal total income, the entire household was removed from the denominator.

All cases from the control group were included as landline numbers, along with interviews completed with a number flagged as a landline from the test group. Interviews from the test group completed with a number flagged as a cellular number were included as cellular numbers in this analysis. Since some cases in the test group had both landline and cellular telephone numbers, the telephone type used during the first call where a case received a disposition of at least a sufficient partial interview was used to classify the household.

- Research Question 10. How often do people report safety or privacy issues? At what rate did interviewers utilize a new software feature enabling the addition of a new telephone number? Are these rates statistically different for landline and cellular numbers?

With the introduction of cellular telephone numbers, the ACS modified the survey introduction to ensure the safety and privacy of respondents. If the respondent was not able to speak on the phone due to safety or privacy issues (e.g., driving or in a public space where they could not speak privately), the interviewer could make note of this, which was then encoded on the CATI transaction file. To give respondents more flexibility in completing a CATI interview, interviewers were able to ask for a different number to call in the future if the respondent prefers. When an interviewer added a new number, this action was noted on the CATI Transaction file. The next outgoing call for that case would then have been made with a number that was likely to not be on the input file. For this analysis, the tabulation showed the unweighted frequency that this occurred for both landline and cellular telephone numbers. One limitation of this analysis to note was that some interviewers may have encountered a respondent with a safety or privacy issue and then also made an appointment for a future callback. In these cases, the outcome of the call attempt would have indicated a future appointment rather than a safety concern.

Data for Research Question 10 came from the CATI transaction file. This was an exploratory analysis for operational purposes, thus data were unweighted. This analysis was meant to shed some light on general trends. The denominators for both the safety/privacy issue rate and the new number given rate included all telephone numbers that were called at least once during the CATI month. For the safety/privacy issue rate, a telephone number was included in the numerator if the interviewer indicated that the respondent cited safety or privacy issues and had to terminate the call on at least one call attempt to the specified phone number.

$$\text{[10.1] Safety/Privacy Issue Rate} = \frac{\text{\# of phone numbers with at least one call attempt outcome of reporting a safety/privacy issue}}{\text{\# of phone numbers called at least once}} * 100$$

$$\text{[10.2] New Number Added Rate} = \frac{\text{\# of phone numbers with at least one call attempt outcome of interviewer adding a telephone number}}{\text{\# of phone numbers called at least once}} * 100$$

- Research Question 11. What are the main findings from the contact center debriefing with the supervisors and monitors regarding the cellular phone training and operational questions?

Towards the end of the CATI month in January 2017, ACS managers and interviewers at all three contact centers were debriefed to gain insight into the effectiveness of the cellular telephone trainings and interview instrument changes that were implemented, and to give staff an opportunity to voice their concerns about the new changes. At each contact center, employees were debriefed with a standardized, semi-structured questionnaire, and their responses were noted by observers. At the conclusion of all three debriefings, notes were aggregated and common themes were discussed. This report documents the most prevalent themes that were found across the contact centers.

5. LIMITATIONS

The main limitation concerning this study is that it is bound by the scoring model in terms of judging the use of including cellular telephone numbers. Importantly, the model at this point has not been optimized to account for differences between landline and cellular numbers in terms of contactability and interview propensity. So while the results of this analysis may suggest that incorporating cellular telephone numbers increases or decreases certain productivity or efficiency metrics, these results could change if there are updates to the cellular telephone number scoring methods. At the same time, it is reasonable to expect that such model changes would be likely to improve the capacity of the ACS CATI operation to incorporate cellular telephone numbers, perhaps leaving this report more at risk of downplaying the potential gains from the use of cellular telephones.

6. RESULTS

Research Question 1. What are the unweighted CATI contact, completion, and completion-given-contact rates for the control and test groups? Are the rates statistically different between the control and test groups?

The first research question considered CATI contact, completion, and completion-given-contact rates. Results are shown in Table 1. In the control group, which includes only landline telephone numbers, interviewers made contact with the sample unit in 26.3 percent of cases. In the test group, which included both landline and cellular telephone numbers, contact was made in 34.1 percent of cases. The 7.8 percentage point difference in contact rates was a significant difference. The control group achieved a 10.4 percent interview completion rate, compared to 12.6 percent in the test group. This 2.2 percentage point difference was also statistically significant. Lastly, completion rates were compared among cases with which CATI interviewers made contact. In the control group, an interview was obtained from 39.7 percent of contacted cases, compared to 36.9 percent in the test group. This difference was also statistically significant. In summary, the test group had higher contact and interview completion rates than the control group. However, contact was more likely to lead to a completed interview in the control group. That is, contact was more of an impetus to a completed interview among the landline-only control group than in the test group, even though the test group had both higher completion rates and contact rates.

Table 1. Comparison of Contact, Completion, and Completion Given Contact Rates

Measure	Control Group N	Control Group	Test Group N	Test Group	Difference	Significant?
Contact Rate	57,115	26.3 (0.3)	18,732	34.1 (0.5)	7.8 (0.6)	Yes
Completion Rate	57,115	10.4 (0.2)	18,732	12.6 (0.4)	2.2 (0.5)	Yes
Completion-Given-Contact Rate	15,014	39.7 (0.7)	6,387	36.9 (1.0)	-2.8 (1.2)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey Computer-Assisted Telephone Interview Paradata.
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Research Question 2. What are the weighted CATI response rates of the control and test groups? Are the rates statistically different between the control and test groups?

Although similar, CATI response rates are distinct from completion rates in that they excluded cases that were later deemed to be ineligible due to being a business or vacant housing unit, were weighted, and accounted for CAPI subsampling by removing cases that were excluded from CAPI and adding a subsampling factor weight to CAPI cases. Table 2 shows the results from the comparison of response rates. The CATI response rates were not significantly different between the control (24.1 percent) and test (23.4 percent) groups. This finding was due to the test group having fewer ineligible cases, which were excluded from the final denominator when calculating weighted CATI response rates. Indeed, prior to removing ineligibles, the test group had a significantly higher response rate than the control group. This suggests that landline numbers replaced by cellular numbers were more likely to be ineligible cases.

Table 2. Comparison of CATI Response Rates

Measure	Control Group (N=40,264)	Test Group (N=15,642)	Difference	Significant?
Response Rate	24.1 (0.7)	23.4 (1.0)	-0.7 (1.3)	No

Source: U.S. Census Bureau, January 2017 American Community Survey Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Research Question 3. How do unweighted late self-response rates compare between the control and test group, among cases that were called at least once? Are the rates statistically different between the control and test groups?

Table 3 shows the comparison of late self-response rates. For the control group, 11.4 percent of cases that were called at least once self-responded during the CATI month. In the test group, the late self-response rate was 12.8 percent, or 1.4 percentage points higher than the control group. This difference was found to be statistically significant, meaning that adding cellular telephones to the CATI workload led to increased self-response during the CATI month. This is a notable finding, as it shows the potential for higher quality telephone numbers to prompt self-response, perhaps by acting as an additional reminder. Higher self-response in the CATI month also has the potential to save operational resources, or enable the Census Bureau to reallocate resources towards other initiatives.

Table 3. Comparison of CATI Late Self-Response Rates

Measure	Control Group Self-Response Rate (N=64,469)	Test Group Self-Response Rate (N=21,474)	Difference	Significant?
Late Self-Response Rate	11.4 (0.2)	12.8 (0.4)	1.4 (0.4)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Difference is statistically significant at the 0.10 level using a two-tailed t-test.

Research Question 4. What is the distribution of contactability in the control and test groups? Are the contactability rates statistically different between the control and test groups? How does contactability vary by model score?

Research Questions 4-6 examined measures of telephone number contactability to gauge variation in the general level of telephone number quality. For each research question, overall contactability is first shown for the specific comparison, followed by contactability distributed across model score groups. As explained above, a match-score was assigned to each telephone number that reflected the predicted probability of an address match. Thus, contactability was predicted to increase as model scores increased. Parsing the telephone number scores into groups allowed us to visualize the score-contactability relationship. Score groups were made by splitting the telephone number workload into score groups. For ease of interpretation, the same score group ranges were used in all contactability visualization figures. All analyses for Research Questions 4 - 6 were performed at the telephone number level.

Table 4 compares contactability of the control and test groups. Overall, numbers from the test group had higher levels of completes, verified incomplete telephone numbers, refusals, and those with not enough information to classify than the control group. These differences were found to be statistically significant. In contrast, the number of telephone numbers classified as either bad or not in service was significantly higher in the control than the test group.

Table 4. Comparison of Overall Contactability

Contactability	Control Group (N=67,997)	Test Group (N=22,301)	Difference	Significant?
Complete	8.7 (0.2)	10.4 (0.4)	1.7 (0.4)	Yes
Verified Number, incomplete	14.2 (0.2)	19.0 (0.4)	4.7 (0.5)	Yes
Refusal	4.6 (0.1)	6.1 (0.3)	1.4 (0.3)	Yes
Not Enough Information	23.7 (0.3)	26.1 (0.5)	2.4 (0.6)	Yes
Not in Service	25.3 (0.3)	17.4 (0.4)	-7.9 (0.5)	Yes
Bad, other	23.4 (0.3)	21.1 (0.4)	-2.4 (0.5)	Yes

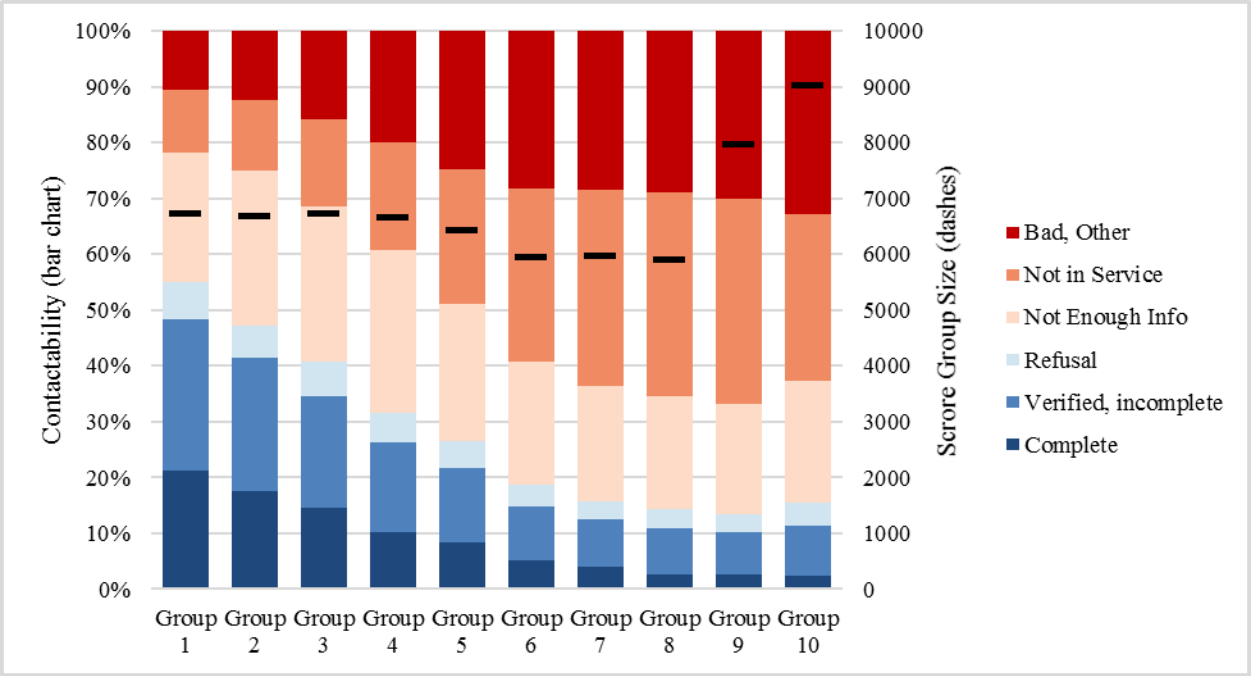
Source: U.S. Census Bureau, January 2017 American Community Survey Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

In Figures 1 – 6, contactability trends are shown across telephone number score groups. To create score groups, the entire distribution of telephone numbers from the combined test and control groups were sorted into one of 10 equal-sized ranked score groups. Thus, the score groups across all figures contain a consistent range of telephone number scores.[12] For each figure, the right y-axis and black dashes display the size of each score group in terms of telephone numbers. The left y-axis shows the level of contactability associated with each score group. It is important to note that some charts do not have numbers in every score group.

Figures 1 and 2 show how contactability varies across model score group in the control and test groups. The colored bars show the distribution of contactability. Between 13.4 and 15.7 percent of the lowest four score groups (Groups 7 through 10) in the control group that made contact culminated in either a complete interview, a verified contact, or a refusal (these groups are the three shades of blue in Figure 1). In the test group, seen in Figure 2, these score groups had similar outcomes for between 26.6 and 28.7 percent of telephone numbers. Additionally, in the test group there were no telephone numbers in the lowest score group (Group 10), as the test group has an overall higher distribution of scores than the control group.

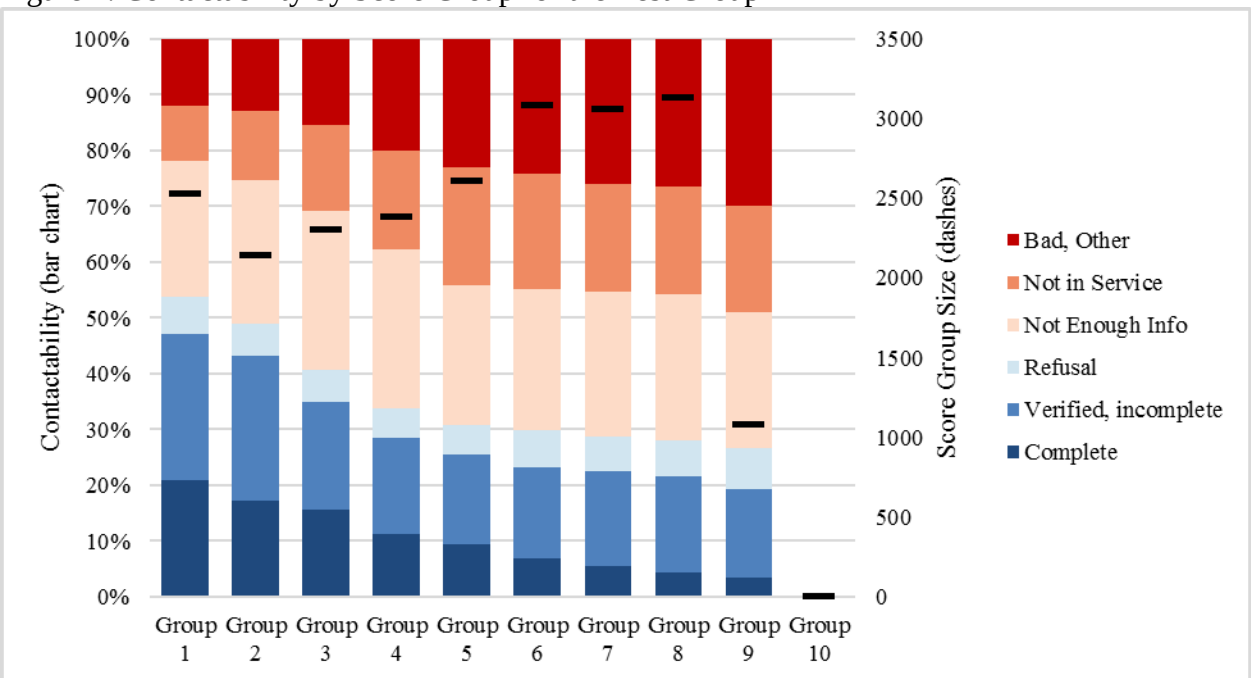
[12] Score group ranges were as follows in Figures 1 - 6: Group 1, 29.78 – 61.55; Group 2, 18.67 – 29.76; Group 3, 9.81 – 18.63; Group 4, 5.00 – 9.80; Group 5, 2.42 – 5.00; Group 6, 1.41 – 2.42; Group 7, 0.92 – 1.41; Group 8, 0.58 – 0.92; Group 9, 0.28 – 0.58; Group 10, 0.10 – 0.28.

Figure 1. Contactability by Score Group for the Control Group



Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata.
Note: Black dash indicates the size of the score group in the figure.

Figure 2. Contactability by Score Group for the Test Group



Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Black dash indicates the size of the score group in the figure.

Research Question 5. What is the distribution of contactability for second-best numbers in the control and test groups? Are the contactability rates statistically different between the control and test groups? How does contactability vary by model score?

Research Question 5 focused on second-best telephone numbers. As described in Section 2, two telephone numbers were sent for roughly 10 percent of cases in the CATI workload. Table 5 compares the contactability for the second-best numbers in the test and control groups. As with the general comparison in Table 4, second-best numbers in the test group had higher contactability in general, with significantly higher proportions of numbers with a complete interview, verified numbers, and refusals. Additionally, the test group had a much lower proportion of numbers that were not in service.

Table 5. Comparison of Contactability of Second Best Numbers

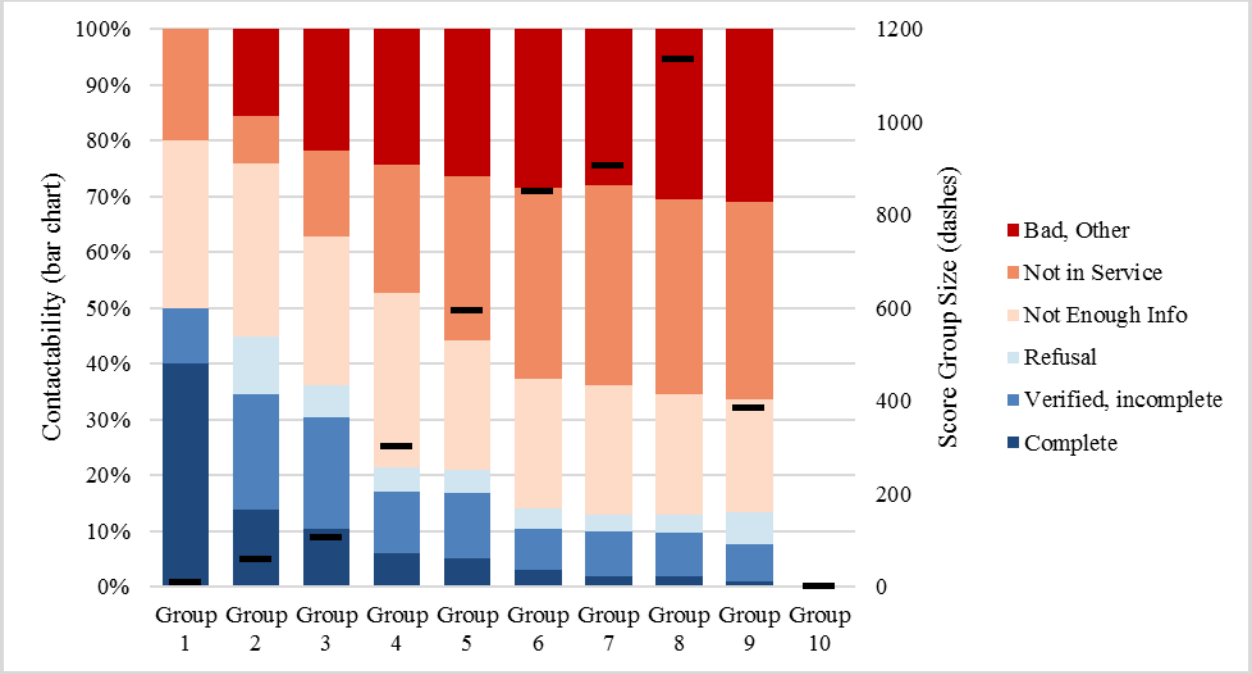
Contactability	Control Group (N=4,339)	Test Group (N=1,011)	Difference	Significant?
Complete	3.2 (0.4)	6.0 (1.3)	2.8 (1.3)	Yes
Verified Number, incomplete	8.8 (0.7)	16.6 (2.0)	7.8 (2.2)	Yes
Refusal	3.8 (0.5)	5.6 (1.2)	1.8 (1.3)	Yes
Not Enough Information	23.3 (0.9)	27.3 (2.5)	4.0 (2.5)	Yes
Not in Service	32.6 (1.2)	18.4 (2.0)	-14.2 (2.3)	Yes
Bad, other	28.2 (1.1)	26.0 (2.1)	-2.2 (2.3)	No

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

This relationship was further explored in Figures 3 and 4. Figure 3 shows that among the lowest-scored four groups, on average between 10 and 15 percent of the telephone numbers called made contact.[13] In contrast, all score groups in the test group had an average of at least 20 percent making contact. There was also evidence of the improved score distribution by introducing cellular numbers, as the test group had no numbers in the lowest four score groups.

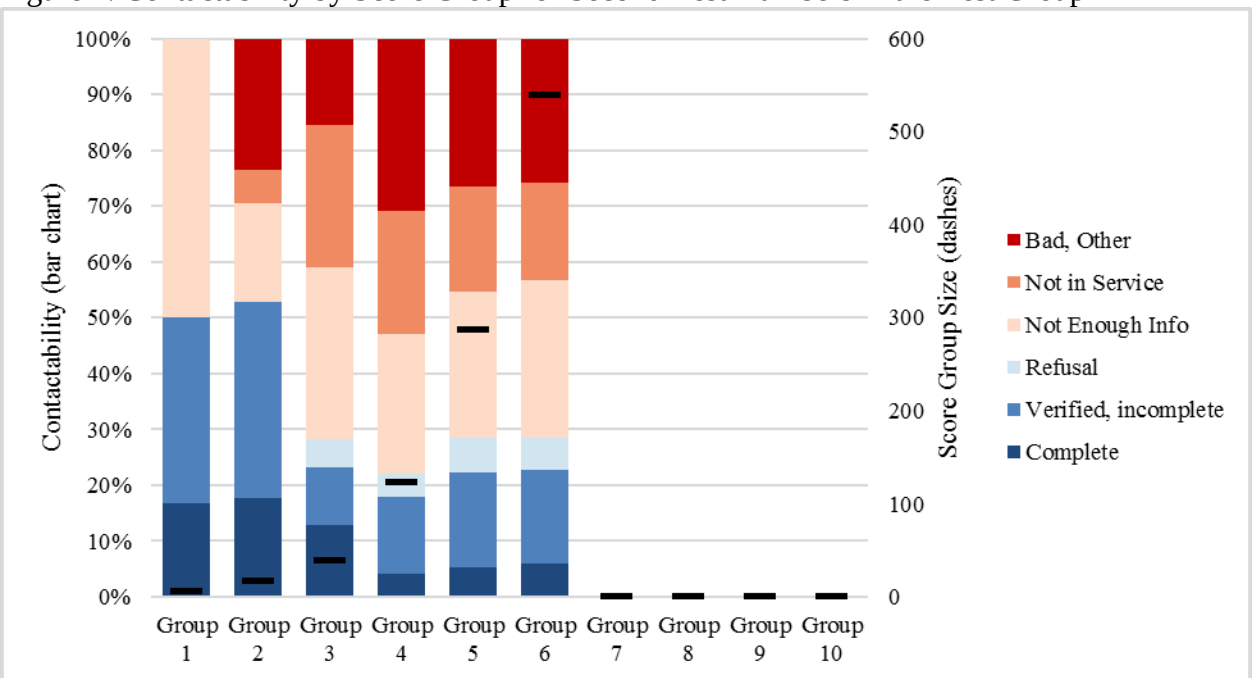
[13] As explained in the earlier section, contact combines the "complete," "verified, incomplete," and "refusal" outcome categories.

Figure 3. Contactability by Score Group for Second Best Numbers in the Control Group



Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Black dash indicates the size of the score group in the figure.

Figure 4. Contactability by Score Group for Second Best Numbers in the Test Group



Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Black dash indicates the size of the score group in the figure.

Research Question 6. What is the distribution of contactability outcomes among landline and cellular telephone numbers? Are the contactability rates statistically different between landline and cellular telephone numbers? How does contactability vary by model score?

Finally, contactability among landline and cellular telephone numbers were compared. Table 6 shows the results of this analysis. Overall, landline numbers had higher rates of complete interviews. However, cellular numbers had higher rates of telephone numbers that make verified contact with an incomplete interview, higher rates of refusals, and lower rates of numbers not in service. These findings suggest that cellular numbers make more contact, but are less successful at converting contact into a completed interview, which aligns with findings from comparing the control and test groups in Research Question 1.

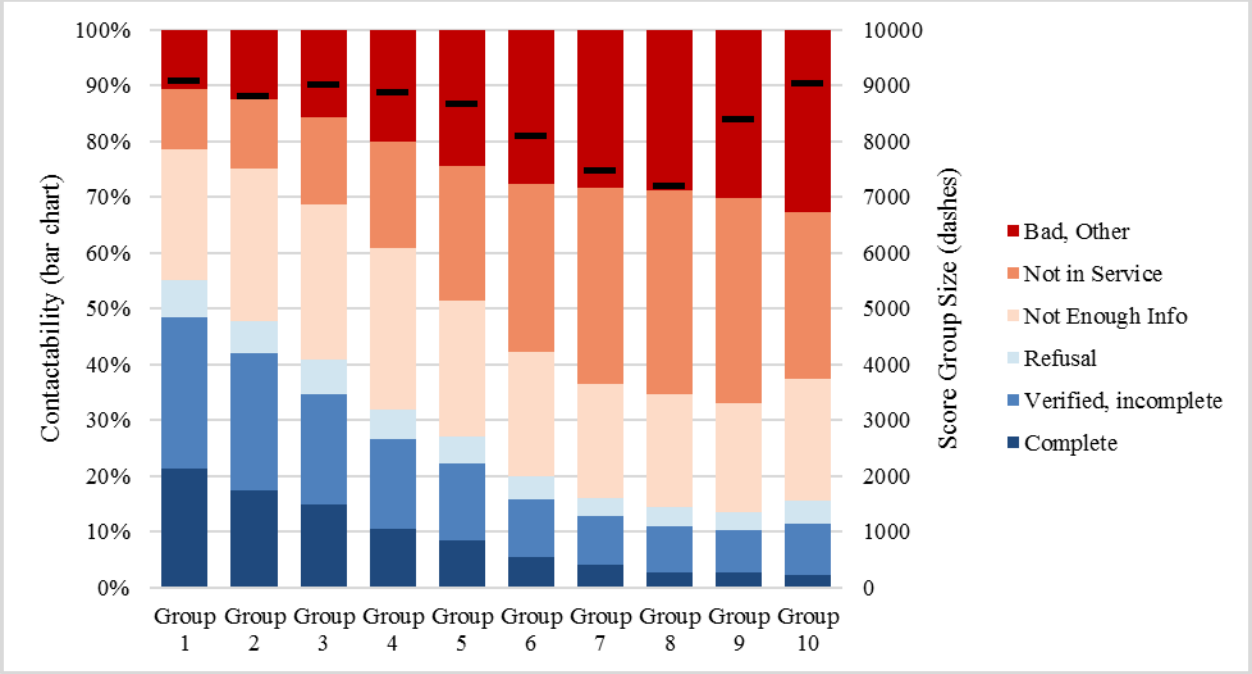
Table 6. Comparison of Contactability by Telephone Type

Contactability	Landline Numbers (N=84,537)	Cellular Numbers (N=5,761)	Difference	Significant?
Complete	9.3 (0.2)	6.6 (0.5)	-2.7 (0.5)	Yes
Verified Number, incomplete	14.8 (0.2)	23.9 (0.7)	9.1 (0.7)	Yes
Refusal	4.7 (0.1)	9.0 (0.6)	4.3 (0.6)	Yes
Not Enough Information	23.9 (0.2)	30.5 (0.9)	6.6 (0.9)	Yes
Not in Service	24.5 (0.3)	6.4 (0.6)	-18.1 (0.7)	Yes
Bad, other	22.8 (0.3)	23.6 (0.8)	0.8 (1.0)	No

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

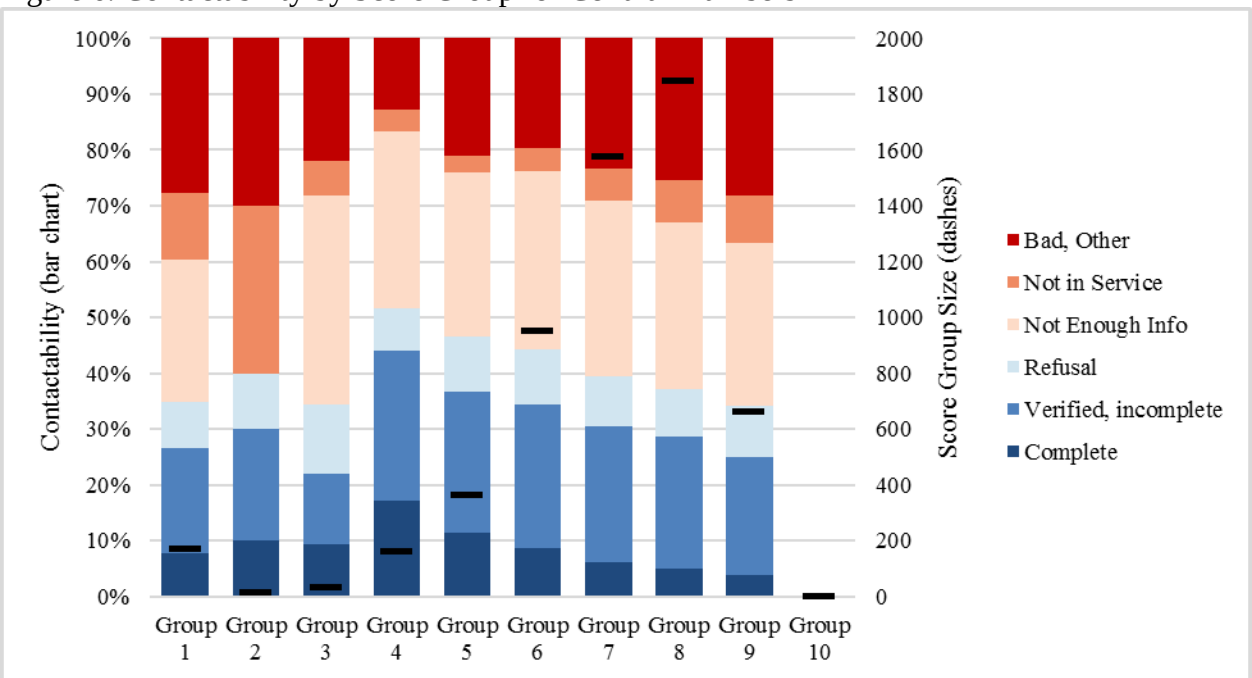
Figures 5 and 6 explore the relationships between model score and contactability within telephone type. There are similarities with Figures 1 and 2, which looked at contactability among the test and control groups. In the lower score groups (Groups 7 through 10), contact rates hovered between 13.5 and 16.0 percent for landlines, and between 34.2 and 39.5 percent for cellular telephones.

Figure 5. Contactability by Score Group for Landline Numbers



Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata.
Note: Black dash indicates the size of the score group in the figure.

Figure 6. Contactability by Score Group for Cellular Numbers



Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata.
Note: Black dash indicates the size of the score group in the figure.

Research Question 7. For cellular and landline telephone numbers, how is CATI contactability distributed across number of call attempts? Specifically, what is the distribution of call attempt totals within CATI outcomes? What is the distribution of final CATI case outcomes by the number of call attempts?

In Tables 7 and 8, the relationship between CATI outcomes and call attempts was examined, comparing landlines and cellular numbers. As described in the methods section, the analysis of Research Question 7 was exploratory and thus was not weighted or tested for statistical significance. To more conclusively separate landline and cellular telephone numbers, only calls to the best number for each case were included, since some cases have both telephone types. Cases that received a late self-response were also excluded from the analysis. Table 7 shows the cumulative distribution of call attempts within CATI outcomes. Overall, landlines reached a final disposition with fewer call attempts. For example, after seven call attempts, 67.8 percent of landlines were closed out, compared to 47.2 percent of cellular numbers. Additionally, almost 12 percent of cellular numbers received a 15th call attempt, compared to less than 5 percent of landlines.

Table 8 shows the outcome distribution for cases that closed after each call attempt, separated by landline and cellular telephones. For example, for cases that closed after one call attempt, about 90 percent were noninterviews for both telephone types. Landlines outpaced cellular phones in most rounds of calling for interviews, while cellular numbers tended to outpace landlines in refusals.

Table 7. Distribution of Call Attempts within CATI Case Outcomes by Telephone Type

Call #	Int., Land	Int., Cell	Ref. Max, Land	Ref. Max, Cell	Call Max, Land	Call Max, Cell	Other non-int., Land	Other non-int., Cell	Total, Land	Total, Cell
1	20.9	16.1					41.0	22.3	29.5	12.5
2	41.8	33.5	5.2	4.4			47.6	33.9	36.7	20.5
3	54.8	48.1	12.8	10.2			67.7	46.3	52.2	28.9
4	64.0	54.9	20.1	16.3			73.4	54.5	57.8	34.7
5	71.6	63.5	26.5	22.6			76.0	61.0	60.9	39.8
6	76.6	68.2	32.0	26.3			80.9	66.8	65.4	43.7
7	81.2	75.5	37.4	29.5			83.0	71.4	67.8	47.2
8	85.0	79.9	42.5	33.8			85.9	74.6	70.6	49.9
9	88.5	82.5	47.5	38.2			89.5	77.5	73.9	52.3
10	91.2	85.6	52.5	41.9			92.0	80.6	76.4	54.8
11	93.5	90.5	58.0	46.8			94.1	83.5	78.6	57.5
12	95.5	94.1	64.0	52.2	95.2	97.3	95.1	85.2	92.1	82.9
13	96.9	96.7	70.3	58.4	99.2	99.8	95.8	86.9	93.8	85.7
14	98.2	97.5	77.2	66.3	99.7	99.9	96.6	88.8	95.2	88.1
15	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata

Note: Int.=Interview, Ref.=Refusal, non-int.=Noninterview. Landline N = 69,804; Cellular N = 5,102.

Table 8. Distribution of CATI Outcomes within Call Attempt Totals by Telephone Type

Call #	Int., Land	Int., Cell	Ref. Max, Land	Ref. Max, Cell	Call Max, Land	Call Max, Cell	Other non-int., Land	Other non-int., Cell
1	7.9	9.7					92.1	90.3
2	32.2	16.5	7.1	9.9			60.6	73.6
3	9.3	13.0	4.8	12.3			85.9	74.7
4	18.5	8.9	13.1	19.1			68.5	72.0
5	24.5	12.7	20.0	22.3			55.5	65.0
6	14.2	9.0	12.3	16.9			73.5	74.1
7	21.1	15.9	22.0	16.5			56.9	67.6
8	14.8	12.1	18.0	27.9			67.2	60.0
9	11.6	8.1	15.0	32.3			73.3	59.7
10	11.8	9.4	20.1	26.8			68.1	63.8
11	11.8	13.7	24.7	32.4			63.5	54.0
12	1.6	1.1	4.4	3.8	89.3	91.6	4.7	3.5
13	8.9	7.0	36.2	39.9	29.8	21.7	25.1	31.5
14	9.8	2.4	48.3	57.6	4.9	0.8	37.0	39.2
15	4.5	1.5	47.6	50.8	0.6	0.2	47.3	47.5
Total	11.1	7.5	9.9	17.9	12.7	23.9	66.3	50.7

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Int.=Interview, Ref.=Refusal, non-int.=Noninterview. Landline N = 69,804; Cellular N = 5,102.

Research Question 8. Do call attempts from cellular and landline telephones make contact at different rates at different times-of-day? Are these rates statistically different for landline and cellular numbers?

This analysis examined variation in the call attempt contact rate by the time-of-day and day-of-the week a call was placed, according to the sample unit's time zone. Results are shown in Table 9. Weekday call attempt contact rates across phone type varied from 11.8 percent for landlines in early afternoon to 13.0 percent for landlines in the morning. The only time period on weekdays where the difference between landline and cellular telephone contact rates was statistically significant is in the morning, where landlines had a contact rate of 13.0 percent compared to 11.9 percent for cellular telephones. On weekends, the contact rate varied from 8.1 percent for cellular numbers during the morning, to 13.8 percent for landlines in the evenings. During the weekend, the contact rate was lower for cellular numbers in every time interval compared to landlines except late afternoon (between 3PM and 5:59PM), when there was no statistically significant difference.

Table 9. Call Attempt Contact Rate by Telephone Type, Time, and Day

Time and Day of Placed Call	Landline Calls	Landline Contact Rate	Cellular Calls	Cellular Contact Rate	Difference	Significant?
Weekdays						
9AM – 11:59AM	43,006	13.0 (0.3)	4,608	11.9 (0.8)	-1.0 (0.9)	Yes
12PM – 2:59PM	44,220	11.8 (0.3)	4,461	12.2 (0.7)	0.4 (0.7)	No
3PM – 5:59PM	92,180	12.3 (0.2)	9,661	12.3 (0.6)	0.1 (0.7)	No
6PM – 9PM	145,771	12.0 (0.2)	13,394	12.3 (0.5)	0.3 (0.5)	No
Weekends						
9AM – 11:59AM	31,039	12.0 (0.3)	3,552	8.1 (0.8)	-3.9 (0.9)	Yes
12PM – 2:59PM	28,521	10.1 (0.3)	2,973	8.7 (1.0)	-1.5 (0.9)	Yes
3PM – 5:59PM	31,315	10.9 (0.3)	3,102	11.8 (0.9)	1.0 (1.0)	No
6PM – 9PM	20,339	13.8 (0.5)	2,075	10.8 (1.2)	-2.9 (1.3)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Research Question 9. Do the characteristics differ between households that completed an interview with a cellular phone and those who completed an interview on a landline phone on the basis of age, race, Hispanic origin, household type, homeownership, and household income?

This research question examined basic household and person-level characteristics among interviews completed via landline and cellular telephones. The first three tables look at person-level characteristics, while the latter three tables consider household-level characteristics. As described above, an interview was classified by telephone type according to the first instance that the case achieved a disposition of at least a sufficient partial interview in the CATI mode.

Table 10 looks at the age distribution for person-level data obtained with a landline or cellular telephone interview. A clear trend was that landline CATI interviews tended to cover older adults, with 36.4 percent at least 65 years old. In contrast, only 10.5 percent of individuals from cellular telephone interviews were 65 or older. Similarly, 26.1 percent of individuals from cellular interviews were under 18 years old, while for landlines this figure was 15.1 percent.

Tables 11 and 12 look at responses to the race and ethnicity questions. Household members from landline CATI interviews were less likely to have selected White alone than those from cellular numbers, less likely to have selected American Indian or Alaskan Native alone, and more likely to have selected Some Other Race alone. Additionally, household members from cellular numbers were more likely to have selected Hispanic ethnicity, as shown in Table 2.

Table 10. CATI Age Response, by Interview Telephone Type

Age Category	Landline Numbers (N=17,311)	Cellular Numbers (N=1,091)	Difference	Significant?
Less than 18 years	15.1 (0.6)	26.1 (2.0)	11.0 (2.1)	Yes
Less than 5 years	2.4 (0.2)	6.4 (1.4)	4.0 (1.5)	Yes
5 to 17 years	12.7 (0.5)	19.7 (2.1)	7.0 (2.1)	Yes
18 to 65 years	48.5 (0.8)	63.4 (2.3)	14.9 (2.5)	Yes
18 to 24 years	5.0 (0.4)	6.3 (1.3)	1.3 (1.3)	No
25 to 44 years	12.1 (0.5)	27.3 (2.3)	15.2 (2.4)	Yes
25 to 34 years	5.3 (0.4)	10.0 (1.8)	4.7 (1.8)	Yes
35 to 44 years	6.9 (0.4)	17.3 (1.9)	10.4 (2.0)	Yes
45 to 64 years	31.3 (0.7)	29.7 (3.0)	-1.6 (3.2)	No
45 to 54 years	13.2 (0.6)	14.9 (2.2)	1.7 (2.2)	No
55 to 64 years	18.1 (0.7)	14.9 (2.4)	-3.2 (2.6)	Yes
65 years or older	36.4 (0.9)	10.5 (1.9)	-25.9 (2.2)	Yes
65 to 74 years	18.3 (0.6)	6.7 (1.5)	-11.5 (1.7)	Yes
75 to 84 years	12.4 (0.6)	3.3 (1.3)	-9.1 (1.4)	Yes
85 years or older	5.8 (0.3)	0.5 (0.4)	-5.3 (0.4)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Table 11. CATI Race Response, by Interview Telephone Type

Race Category	Landline Numbers (N=17,175)	Cellular Numbers (N=1,077)	Difference	Significant?
White alone	74.4 (0.9)	66.4 (5.0)	-8.0 (5.1)	Yes
Black alone	15.4 (0.8)	13.2 (4.0)	-2.2 (4.1)	No
American Indian and Alaska Native alone	0.9 (0.2)	0.1 (0.2)	-0.7 (0.3)	Yes
Asian alone	3.1 (0.4)	5.1 (2.5)	2.1 (2.5)	No
Native Hawaiian and Other Pacific Islander alone	0.2 (0.1)	1.1 (1.0)	0.9 (1.0)	No
Some Other Race alone	3.6 (0.5)	10.1 (3.1)	6.5 (3.1)	Yes
Two or More Races	2.4 (0.3)	6.4 (3.1)	1.5 (1.7)	No

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level.

Table 12. CATI Ethnicity Response, by Interview Telephone Type

Ethnicity Category	Landline Numbers (N=17,305)	Cellular Numbers (N=1,090)	Difference	Significant?
Hispanic	12.5 (1.0)	23.9 (3.8)	11.4 (3.8)	Yes
Non-Hispanic	87.5 (1.0)	76.1 (3.8)	-11.4 (3.8)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Tables 13, 14, and 15 compare household characteristics of sampled units that responded via a landline or cellular telephone. First, Table 13 shows that married-couple households were the largest category for both telephone types. However, cellular interviews resulted in more interviews from female-headed households without a husband present than landline numbers. In contrast, landline interviews resulted in more interviews from nonfamily households than cellular numbers, particularly single-person household. Table 14 shows that households interviewed via cellular numbers were more likely to be renter-occupied compared to those interviewed via a landline number. Finally, the distribution of household income by interview telephone type is shown in Table 15. While most income categories were not statistically different across type of telephone, landline responders were significantly more likely to belong to a household in the highest earning category ($200,000 or more) or the lower income categories ($10,000 - $14,999 and $15,000 - $24,999).

Table 13. CATI Household Type Response, by Interview Telephone Type

Family Type	Landline Numbers (N=7,742)	Cellular Numbers (N=407)	Difference	Significant?
Married-couple family	48.8 (1.0)	53.1 (4.8)	4.3 (4.8)	No
Other family households	15.1 (0.8)	20.2 (3.6)	5.1 (3.6)	Yes
Male householder, no wife present	4.1 (0.4)	5.6 (2.0)	1.4 (2.0)	No
Female household, no husband present	11.0 (0.6)	14.7 (3.2)	3.7 (3.3)	Yes
Nonfamily household	36.1 (0.9)	26.7 (4.0)	-9.4 (4.1)	Yes
Householder living alone	33.4 (0.9)	21.4 (3.7)	-12.0 (3.9)	Yes
Householder not living alone	2.6 (0.3)	5.2 (2.2)	2.6 (2.2)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Table 14. CATI Housing Tenure Response, by Interview Telephone Type

Tenure	Landline Numbers (N=7,721)	Cellular Numbers (N=405)	Difference	Significant?
Owner-occupied	86.3 (0.7)	73.4 (3.6)	-12.9 (3.7)	Yes
Renter-occupied	13.7 (0.7)	26.6 (3.6)	12.9 (3.7)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Table 15. CATI Household Income, by Interview Telephone Type

Household Income	Landline Numbers (N=5,067)	Cellular Numbers (N=272)	Difference	Significant?
Less than $10,000	6.6 (0.6)	4.5 (2.3)	-2.1 (2.4)	No
$10,000 - $14,999	7.5 (0.8)	3.3 (1.7)	-4.2 (1.9)	Yes
$15,000 - $24,999	15.0 (1.1)	7.7 (2.9)	-7.3 (3.0)	Yes
$25,000 - $34,999	11.6 (0.8)	11.8 (3.7)	0.3 (3.8)	No
$35,000 - $49,999	12.7 (1.0)	14.3 (3.7)	1.7 (3.6)	No
$50,000 - $74,999	15.3 (0.9)	18.6 (3.9)	3.3 (3.9)	No
$75,000 - $99,999	9.9 (0.8)	16.2 (4.5)	6.2 (4.4)	Yes
$100,000 - $149,999	11.6 (0.9)	13.3 (4.1)	1.6 (4.3)	No
$150,000 - $199,999	4.9 (0.6)	7.5 (3.0)	2.6 (3.1)	No
$200,000 or More	4.8 (0.6)	2.8 (1.8)	-2.0 (1.9)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Research Question 10. How often do people report safety or privacy issues? At what rate did interviewers utilize a new software feature enabling the addition of a new telephone number? Are these rates statistically different for landline and cellular numbers?

With the cellular telephone test, the ACS added new interviewer capabilities of reporting safety/privacy issues and adding a new, respondent-preferred telephone number. Table 16 explores the frequency at which these features were utilized, comparing cellular and landline numbers. Safety concerns were cited at least once for 3.5 percent of cellular numbers called during the CATI month, compared to 0.2 percent of landline numbers. This difference was statistically significant and somewhat expected, as people are probably more likely to be engaged in another activity, such as driving, when answering a cellular call.

There was a limitation to calculating the frequency at which a new number is added, because this might have also included setting up an interview appointment. When an interview appointment

was set up, the documented outcome was for the new appointment rather than for the new number. However, cellular number interviews had a higher rate of adding a new number, being an outcome at least once for 0.6 percent of cellular numbers called compared to 0.3 percent of landlines. This difference was also statistically significant.

Table 16. Occurrence of Safety Concerns and New Number Added by Telephone Type

Telephone Type	Landline Numbers (N=84,215)	Cellular Numbers (N=6,083)	Difference	Significant?
Safety Concerns	0.2 (0.0)	3.5 (0.4)	3.4 (0.4)	Yes
New Number Added	0.3 (0.0)	0.6 (0.2)	0.4 (0.2)	Yes

Source: U.S. Census Bureau, January 2017 American Community Survey, Computer-Assisted Telephone Interview Paradata
Note: Differences are statistically significant at the 0.10 level using a two-tailed t-test.

Research Question 11. What were the main findings from the contact center debriefing with the supervisors and monitors regarding the cellular phone training and operational questions?

In early 2017, ACS staff traveled to the Census Bureau's three contact centers to debrief the interviewers and managers on the CATI changes and the training leading up to the changes. In the debriefings, both managers and interviewers were asked questions about the training process, which included a job aid self-study, a flow chart that displayed the pathing of how an interviewer was to code a call outcome, as well as a walk-through practice session. Both the interviewers and managers noted that the flow chart showing the pathing changes in the instrument were easy to understand and helpful. It was also noted that more hand-on practice, such as a walk-through, would have been beneficial.

Some of the interviewers seemed to indicate that adding cellular numbers was an overall positive undertaking, but that some of the operational changes needed tweaking. First, the earlier part of the interview where the interviewer asked the respondent if they were safe received a lot of negative feedback. Interviewers felt this statement resulted in general confusion, suspicion, or even interview breakoffs and hangups. Interviewers indicated that respondents did not understand what was meant by asking about general "safety," and the felt the question could be improved by asking more directly if the respondent is driving.

Several other concerns were noted. Interviewers also wanted the flexibility to be able to ask for a better time to call back if the respondent indicated that they were driving, and felt this would improve response. Currently, the instrument asks the interviewer to immediately stop the interview, which they felt was unnecessary or sometimes even rude. Another question that got a negative reaction from the interviewers was asking if the respondent was on a cellular phone or landline. They reported that this question felt invasive for some respondents, as there was no immediately noted reason for the question, and the answer did nothing to change the course of the interview. Finally, the interviewers noted that the changes to the introduction of the interview were good overall (i.e., where they ask about "census forms" rather than a "survey"), but that there was no place to introduce yourself.

7. CONCLUSION

The purpose of this research was to test the effect of including cellular telephone numbers in the ACS CATI workload. This was accomplished by partitioning 25 percent of the production panel into a test group that used the new methodology of including cellular numbers. The remaining 75 percent of the production sample was a control group using the current methodology.

This report contains several important findings. First, the test group, which included cellular numbers, had a higher contact rate (31.4 vs. 26.3 percent) and interview completion rate (12.6 vs. 10.4 percent). This means that the test group proportionally completed more CATI interviews, and more frequently made contact with the sampled housing unit. However, the control group was more successful at converting contact into a completed interview. This was because the contactability issues for landlines was the result of out of service numbers, rather than failing to convert a contact into an interview. Cellular numbers, on the other hand, had higher contactability but were somewhat less successful at converting the contact into a completed interview.

Additionally, cases that were called at least once in the test group were more likely to have a late self-response (12.8 vs. 11.4 percent). This was likely due to a higher contact rate, which was shown to be correlated with higher late self-response rates as the telephone contact served as an additional reminder to respond. However, after accounting for CAPI subsampling and cases later found to be ineligible, differences in the CATI response rate between the test and control were not statistically significant. This means that even though the test group had a higher interview completion rate, the overall effect on the CATI response rate was not significant. However, this was an artifact of the test group having fewer ineligible cases, which were excluded from the final denominator when calculating weighted CATI response rates.

The test group had higher levels of contactability and lower rates of bad numbers. This was seen when considering all telephone numbers aggregated, as well as when focusing on the second-best number specifically. When comparing landline and cellular numbers directly, landlines had a higher interview completion rate, but they also had lower levels of other forms of contact and a much higher proportion of out-of-service numbers. Overall, general contactability improved with the addition of cellular telephone numbers.

Lastly, adding cellular numbers appears to have reached a different demographic population compared to only including landline numbers. This test found that landlines reached people who were older, lived alone and owned their home; while cellular numbers appeared to include people who were somewhat more racially and ethnically diverse, and who rented their home.

8. REFERENCES

AAPOR Cell Phone Task Force (2010). *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers.* Retrieved from: *https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/2010AAPORCellPhoneTFReport.pdf*

Blumberg, S.J. & Luke, J.V. (2016). *Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January-June 2016.* Retrieved from Center of Disease Control and Prevention Website: https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201612.pdf.

Brick, J.M., Brick, P.D., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2007). Cell Phone Survey Feasibility in the U.S.: Sampling and Calling Cell Numbers Versus Landline Numbers, *Public Opinion Quarterly*, *71*(1), 23-29.

Dutwin, D. & Lavrakas, P.J. (2016), Trends in Telephone Outcomes, 2008-2015, *Survey Practice*, *9*(2).

Griffin, D.H., Fischer, D. P., & Morgan, M.T. (2001). *Testing an Internet Response Option for the American Community Survey*. Paper Presented at the annual meeting of the American Association for Public Opinion Research, Montreal, Quebec.

Groves, R.M. & Couper, M.P. (2012). *Nonresponse in household interview surveys*. New York: John Wiley & Sons.

Hall, J., Carlson, B.L., & CyBulski, K. (2011). *RDD Unplugged: Findings from a Household Survey Using a Cell Overlap Design* (Mathematica Policy Research, Working Paper 1); Alexandria, VA: American Statistical Association.

McGreeney, K. (2016). *Pew Research Center Will Call 75% Cellphones for Surveys in 2016.* Retrieved from Pew Research Center Website: *http://www.pewresearch.org/fact-tank/2016/01/05/pew-research-center-will-call-75-cellphones-for-surveys-in-2016.*

Mills, G. (2016a). *Evaluation of Transitioning Telephone Number Sources for the American Community Survey* (ACS16-RER-12). Suitland, MD; U.S. Census Bureau.

Mills, G. (2016b). *Simulated Effects of Changing Calling Parameters and Workload Size on Computer Assisted Telephone Interview Productivity in the American Community Survey* (ACS16-RER-22). Suitland, MD; U.S. Census Bureau.

Nichols, E., Horwitz, R., & Tancreto, J. G. (2013). *Response Mode Choice and the Hard-to-Interview in the American Community Survey* (RSM2013/01). Suitland, MD; U.S. Census Bureau.

U.S. Census Bureau. (2014). *American Community Survey Design and Methodology (January 2014).* Retrieved August 14, 2017 from http://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

Appendix A: ACS 2017 Cell Phone Debriefing – Interviewers

In January we tested including cellular telephone numbers, in addition to landline numbers, for part of the sample. A job aid/flow chart was provided to implement this change, with changes to the instrument expected in May. We are interested in your experience during January with this change.

Which Contact Center is this?

Hagerstown Contact Center
Jeffersonville Contact Center
Tucson Contact Center

About You and Your Experience

1. How many years of experience do you have working on ACS NRFU production?
2. Do you also work on TPOPS? If so, how long?

Cell Phone Training

3. Did you have a chance to review the job aid as a self-study?
   Yes
   No
   3a. If yes, was it useful? If not, why not?

4. Did you have a chance to attend training in a classroom setting with a supervisor?
   Yes
   No
   4a. If yes, was it useful? If not, why not?

5. Did you have a chance to use the training instrument and have a practice session in WebCATI?
   Yes
   No
   5a. If yes, was it useful? If not, why not?

6. Was there anything in the training that wasn't clear?
7. Was there anything in the training that you would change or improve?
8. Did you receive additional instruction from your supervisor about the procedure (at any time during January) that was not covered in the self-study or classroom training? If so, what?

Operational Notes and Issues

9. Were the 2 old job aids (Alternate Wording for Introduction, Removal of the Verified Number Job Aid) removed from your workstation at the start of the month?
   Yes
   No

10. Was the new job aid (ACS Nonresponse Followup Cell Phone Test Flow Chart) available to you at your workstation at the start of the month?
    Yes
    No
11. Were you able to navigate the instrument while using the new job aid? Is there anything about this processes that can be improved (it will be fully automated in May)?
12. You asked the following question: "We want to ensure your safety and privacy since we reached you on a cell phone. Are you able to safely talk at this time?" Did this question make sense to the respondent? If not, what was the confusion? How did you clarify?
13. If provided, what were some of the situations that caused respondents to say they had privacy or security issues? (for example, they told me they were driving)
14. If respondents indicated that they were not able to safely talk, how did you end the call (what did you say, what did you do)?
15. Were there respondents who indicated that they wanted to continue the call while driving? If so, what did you do?
16. Did you find out later in the interview that the respondent was driving? How often did this happen? What did you do?
17. Did you get any incoming calls from a respondent who was driving? If so, what did you do?
18. Did you receive any comments or concerns from respondents about calling them on their cell phone? If so, what were they? (Followup if necessary. Did anyone raise concerns about using their minutes for the call or paying for the call? If so, what did you do?)
19. Did you receive any complaints or indication from the respondent that we were calling them outside of the 9am-9pm local calling time?
20. Did you ever reach a child 14 or under on a cell phone? What did you do?
21. If respondents told you they were on a cell phone, were there any questions that were harder to answer because, for example, the respondent was not at home when we called? (Followup if necessary. Did respondents indicate that they would like to look up records but weren't able to because you reached them on a cell phone away from home?)

Other

22. Are there any other concerns or issues about the use of cellular telephones that you would like us to know about?

Appendix B: ACS 2017 Cell Phone Debriefing – Supervisors and Monitors

In January we tested including cellular telephone numbers, in addition to landline numbers, for part of the sample. A job aid/flow chart was provided to implement this change, with changes to the instrument expected in May. We are interested in your experience during January with this change.

Which Contact Center is this?

Hagerstown Contact Center
Jeffersonville Contact Center
Tucson Contact Center

About You and Your Experience

1. What is your job title and how many years of experience do you have working on ACS NRFU production?
2. Do you also work on TPOPS? If so, how long?

Cell Phone Training

3. Did you have a chance to review the job aid as a self-study?
   Yes
   No
   3a. If yes, was it useful? If not, why not?

4. Did you have a chance to attend training in a classroom setting with a supervisor?
   Yes
   No
   4a. If yes, was it useful? If not, why not?

5. Did you have a chance to use the training instrument and have a practice session in WebCATI?
   Yes
   No
   5a. If yes, was it useful? If not, why not?

6. Was there anything in the training that wasn't clear?
7. Was there anything in the training that you would change or improve?
8. Did you provide additional instruction about the procedure (at any time during January) that was not covered in the self-study or classroom training? If so, what?
9. Was there any additional feedback you received from the Interviewers about the training?

Operational Notes and Issues

Please consider any feedback you heard from Interviewers as well as your own experiences when answering these questions.

10. Were the 2 old job aids (Alternate Wording for Introduction, Removal of the Verified Number Job Aid) removed from your workstation at the start of the month?

    Yes
    No

11. Was the new job aid (ACS Nonresponse Followup Cell Phone Test Flow Chart) available to you at your workstation at the start of the month?

    Yes
    No

12. Were you able to navigate the instrument while using the new job aid? Is there anything about this processes that can be improved (it will be fully automated in May)?
13. ** You asked the following question: "We want to ensure your safety and privacy since we reached you on a cell phone. Are you able to safely talk at this time?" Did this question make sense to the respondent? If not, what was the confusion? How did you clarify?
14. If provided, what were some of the situations that caused respondents to say they had privacy or security issues? (for example, they told me they were driving)
15. If respondents indicated that they were not able to safely talk, how did you end the call (what did you say, what did you do)?
16. Were there respondents who indicated that they wanted to continue the call while driving? If so, what did you do?
17. Did you find out later in the interview that the respondent was driving? How often did this happen? What did you do?
18. Did you get any incoming calls from a respondent who was driving? If so, what did you do?
19. Did you receive any comments or concerns from respondents about calling them on their cell phone? If so, what were they? (Followup if necessary. Did anyone raise concerns about using their minutes for the call or paying for the call? If so, what did you do?)
20. Did you receive any complaints or indication from the respondent that we were calling them outside of the 9am-9pm local calling time?
21. Did you ever reach a child 14 or under on a cell phone? What did you do?
22. If respondents told you they were on a cell phone, were there any questions that were harder to answer because, for example, the respondent was not at home when we called? (Followup if necessary, did respondents indicate that they would like to look up records but weren't able to because you reached them on a cell phone away from home?)

Other

23. Are there any other concerns or issues about the use of cellular telephones that you would like us to know about?

For Supervisors/Monitors Only

24. How did monitoring go for the two cell phone questions (Are you speaking on a cell phone? We want to ensure your safety and privacy since we reached you on a cell phone. Are you able to safely talk at this time?)?
25. What issues you were you identifying?
26. Did you get any pushback from interviewers on how you scored these items?