Research description for presentation
Using Lifestyle Segments to Enhance Response Propensity Models and Design Nonresponse Interventions

Mary H. Mulry, Nancy Bates and Matt Virgile[1]
U.S. Census Bureau, Washington, DC 20233

Abstract
As the 2020 U.S. Census approaches, the preparations include tests of new methodologies for enumeration that have the potential to reduce cost and improve quality. The 2015 Census Test in Savannah, GA included tests of Internet and mail response modes and of online delivery of social marketing communications focused on persuading the public to respond by Internet and mail. Merging data from the 2015 Census Test with third-party lifestyle segments and the Census Bureau's new Low Response Score produces a dataset suitable for studying relationships between census response, Low Response Scores, and lifestyle segments. This paper uses the merged dataset to examine whether lifestyle segments can provide insight to hard-to-survey populations, their response behavior, and interactions with social marketing communications. The paper also includes analyses with nationwide data that support the broader application of using segmentation variables in self-response propensity models and a discussion of potential applications of segment lifestyle information in tailored and targeted survey designs for hard-to-survey populations.

1. Introduction

Social marketing campaigns have been part of the U.S. Census since 1950 when the Census Bureau began a partnership with the Advertising Council, a pro bono group of advertising agencies that create marketing campaigns for non-profit causes. The 2000 Census was the first to invest hundreds of millions of dollars for a *paid advertising campaign*. The budgeted mail response rate in 2000 was 61 percent but, in fact, a mail response rate of 67 percent was achieved. Experts believe the paid campaign played a large role in this success. The 2000 campaign convinced managers to repeat such a paid campaign in the 2010 Census, and the 2020 Census will likewise leverage a communication program that will include paid advertising.

Unlike the 2010 Census, the 2020 Census will include the Internet as a mode of self-response. The 2020 Census also will allow for self-response without the requirement that households have access to a Census-assigned unique address identifier. This "Non-ID response" means that anyone with Internet access can self-respond on behalf of their household. In fact, the agency currently is planning for a 63.5 percent self-response rate of which 47 percent is predicted to come from Internet response, 11 percent by mail and 5 percent by telephone (Blumerman *et al.* 2015, p. 12).

The innovations of Internet response and non-ID response have implications for the communications campaign. With the expansion of the use of digital media in advertising in recent years, the census web

[1] This report is released to inform interested parties and encourage discussion of work in progress. The views expressed on statistical, methodological, and operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

response option means that digital ads can bring respondents directly to the Census form with one simple click. Combined with a non-ID response option, this click-to-complete functionality may help increase self-response among certain hard-to-survey populations. Recent studies suggest the smart phone ownership gap has closed between whites and racial and ethnic minorities in the U.S. In surveys conducted during 2016, the Pew Research Center (2017) found that 77 percent of whites owned smartphones compared to 72 percent of Blacks and 78 percent of Hispanics. Additionally, 12 percent reported their phone as their only access to the Internet, and this behavior was particularly prevalent among young adults, non-whites, low income, and low education populations with the highest being 27 percent for those who did not graduate from high school (Pew Research Center, 2017).

In this paper, we test the application of a commercial marketing database as a means for planning the 2020 communications campaign. For over 40 years, the Census Bureau has partnered with the mapping and spatial analytics software company ESRI to develop products for the Decennial, Economic, Demographic and Population divisions of the agency. Here, we use ESRI's geographic and lifestyle population segmentation known as Tapestry™, which is available at several geographic levels. This segmentation at the tract level is used in conjunction with a new Census Bureau metric (the Low Response Score or LRS) to retroactively analyze data from the 2015 Census Test in the Savannah, Georgia Direct Market Area (DMA). The 2015 test included a comprehensive social marketing campaign including the use of paid television, radio, print and digital advertising.

Our first research question is the following: *Can lifestyle segments provide insight into hard-to-survey populations, their response behavior, and interactions with social marketing communications?* A secondary research question is: *Can lifestyle segments improve self-response propensity models?*

2. Data and Segmentation

Our strategy for obtaining answers to the research questions involves partitioning the 2015 Savannah Test Census site by the Tapestry segments and then merging the Test Census self-response results and LRSs. With this dataset, we are able to explore whether the segments reveal a relationship between propensity to self-respond and hard-to-count status as indicated by the LRS. This relationship is of interest because LRS is available nationwide and the Savannah Census Test is the only current source of census self-response data in a setting similar to plans for the 2020 Census in that the Internet response option is available in the presence of a communications campaign.

In addition, partitioning the 2015 Savannah Census Test data by segment enables us to examine whether the segments display patterns in the respondents' choices for self-response. The unit of analysis in our study is the segment, which is a level that a communications campaign would find useful for planning messages and media purchases. This section discusses how we organize the data by segment. Section 3 discusses the implementation of the 2015 Savannah Census Test.

2.1 2010 Audience Segmentation and ESRI Tapestry Segments

For the 2010 Census communications campaign, the Census Bureau developed a geographic audience segmentation strategy (Bates and Mulry 2011). Eight population segments were identified, each according to propensity for self-response. Segments were developed using cluster analysis and twelve socio-economic, demographic, and housing variables known to correlate with hard-to-count areas. When developing the segments, the agency relied on internal data sources from the 2000 Census and American Community Surveys conducted after the 2000 Census.

The ESRI Tapestry segmentation is an external third-party database used to market goods and services. The segmentation integrates geographic and demographic data from multiple Big Data sources to classify

neighborhoods in a manner that provides insight about consumer purchasing patterns, housing characteristics, and socioeconomic characteristics. The methodology for forming the segments uses cluster analysis and data mining and is updated annually. Our research used the 2015 version to correspond with the timing of the 2015 Census Test. The data sources for recent versions of the Tapestry segmentation include the 2010 Census, the American Community Survey, Experian's ConsumerView database (2017), and the Survey of the American Consumer from GfK MRI (ESRI 2015).

While Tapestry segments are primarily used for consumer marketing purposes, their distinguishing characteristics seem intuitively applicable to the Census social marketing campaign directed at encouraging census response. For example, each segment is scored along a race and ethnicity diversity index, income and net worth, age by sex distributions, and neighborhood and socioeconomic traits. These descriptions include household compositions, vacancy rates, housing structures, employment indices, government program participation rates, financial well-being, media consumption habits, and Internet usage. Additionally, the segments are distinguished along more psychographical constructs including attitudes regarding household gender roles and reliance on faith-based organizations – important variables when crafting communication messaging.

The ESRI segmentation has three nested levels, which provide flexibility to the user. The 67 Tapestry segments aggregate to 14 LifeMode groups based on lifestyle and life stage. Further aggregation into six urbanization groups reflects area features such as population density and size of city (ESRI 2015). Detailed descriptions of the Tapestry segments, including the LifeMode to which they belong, may be found at http://downloads.esri.com/ESRI_CONTENT_DOC/DBL/US/TAPESTRY/TAPESTRY_FLIERS_ALL_0914.PDF. Tapestry segments are available at different levels of geography including DMA, county, tract and block-groups.

For purposes of our paper, we use segments assigned at the Census tract level and concentrate primarily on the lowest segmentation level (the 67 segments). Housing units in the 2015 Savannah Census Test were assigned to segments based on the tract of their address. With this approach, all the housing units in a tract are in the same segment.

Although the exact formulation of the ESRI Tapestry segments is proprietary, ESRI's chief demographer Lynn Wombold said in a personal communication (05/04/2017) that Census and survey response rates *did not* figure into the development of the Tapestry segments. Therefore, confounding should not be a problem when using the ESRI segments to study census response. The formation of the segments does use some variables that previous studies have shown to be correlated with response, such as tenure, income, and education, which had the potential of being an advantage in our research (Bates and Mulry, 2011).

2.2 Planning Database and LRS

Recently, the Census Bureau developed a hard-to-survey metric known as the Low Response Score, or LRS (Erdman and Bates, 2017). The score is simply a geographical area's fitted value from a regression model predicting the area's mail return rate from the 2010 Census. In all, 25 predictor variables were used as input to develop the LRS[2]. A high LRS suggests that an area will have a low level of self-response and may require additional resources by Census Bureau Partnership Specialists, local officials, community advocates, and city planners. The LRS is publicly available at both the Census tract and block-group level on the Census Bureau's Planning Database (https://www.census.gov/research/data/planning_database/).

[2] The 25 variables included % age 5-17; % age 18-24; % not High School graduate; % below poverty; % female headed households, no husband; mean number persons per household; % related child under age 6; % moved in last year; % renter occupied; % vacant houses; % Black; % married family households; median household income; median house value; % Hispanic; population density; % non-Hispanic White; % age 65+; % males; % college graduates; % moved in last 5 years; % single family units; % single person households; % age 25-44; % age 45-64.

For our study, housing units were assigned the LRS associated with the tract of their address. For each segment, we calculated the mean LRS over all the addresses comprising the segment. In effect, the mean LRS is the weighted average of the LRS for the tracts in the segment where the weight for a tract is determined by the proportion of the segment's housing unit population that is in the tract.

2.3 2015 Census Test Participation Rates

Our primary outcome variable of interest is the participation rate for segments. Simply stated, the participation rate is the number of completed test questionnaires (mail, Internet and telephone response combined) divided by the number of households receiving Census mail pieces minus the number of addresses indicated as Undeliverable As Addressed (UAA) after the first mailing. The definition of the participation rate follows:

$$participation\ rate = \frac{number\ of\ addresses\ that\ self\ respond}{number\ of\ addresses - number\ of\ UAAs}.$$

The calculation of the participation rate for a segment pooled all the addresses in the segment. Similar to the mean LRS, the participation rate may be viewed as a weighted sum of the participation rates for the tracts in the segment where the weights are the proportion of the segment's population in the tract.

The U.S. Postal Service (USPS) indicated that the first mail piece received a UAA for 20.9%[3] of the housing units in the Savannah Test Site. Often a UAA means the housing unit is vacant, but that is not necessarily the case as found in the 2015 Census Test in Maricopa, AZ, which included a Nonresponse Followup and a subsequent Evaluation Followup (Mulry *et al.* 2016; see Table 7). There are other reasons for the assignment of a UAA, such as the address on the mail piece has an error; the person at the address refuses the mail, or the mailbox is full. The 2015 Savannah Test did not have a Nonresponse Followup operation that would have determined whether these housing units were occupied on Census Day.

3 The Savannah Census Test

The 2015 Census Test in the Savannah, Georgia DMA varied the exposure to different communications campaign elements by household with the goal to garner insights on effective digital advertising. The test included a variety of communication elements to mimic a census communications environment. These elements included partnership activities, an earned media campaign, and a traditional paid media campaign (i.e., broadcast and cable television, radio, print, and out-of-home). In addition, an extensive digital advertising effort explored using different spending levels and targeted digital advertising to reach traditionally hard-to-count populations. Census Day for the 2015 Census Test was April 1, 2015 and self-responses were accepted by Internet, mail, and telephone so that the conditions were as similar as possible to the 2020 Census, including the Census Day of April 1, 2020. Since the focus was on self-response, the test did not include Nonresponse Followup or some of the other operations that will be part of the 2020 Census.

3.1 Two goals: Non-ID response and Digital Advertising

The 2015 Census Test in Savannah marked several firsts. It was the first time the Census Bureau used communications and paid advertising to not only raise awareness but also drive direct response to the online data collection instrument, by which respondents could complete the questionnaire with or without having

[3] The two segments with the highest percentages of housing units with UAAs on the first mailing were City Lights and Rural Resort Dwellers with 73.8% and 66.6%, respectively. The range for the other segments was 2.7% for Home Improvement to 38.1% for Rural Bypasses.

a preassigned Census ID (i.e., non-ID response). This test census was also the first opportunity—since the Census Bureau began introducing the mailout/mailback approach in the 1960s, which it more fully adopted in the 1970s—for some households to participate without receiving any mailed materials by responding to a digital advertisement (Vines et al. 2017).

3.2 The Mailout Panels

The 2015 Census Test included 439,918 housing units in the Savannah DMA. The DMA included 17 counties in Georgia and 3 in South Carolina. A sample of 90,000 addresses received mailings that encouraged self-response to the census and were partitioned into three panels, each a subsample that received a different mail strategy (Mathews and Rothhaas, 2015). The sample selection procedure initially divided the tracts, and therefore the addresses, into nine strata defined by residential Internet penetration based on 2013 Federal Communications Commission (FCC) broadband data and the Census Bureau's LRS. The selection of the 90,000 addresses used a systematic sampling scheme with a random start within each stratum. Each stratum was sorted by tract, then block group, block, and assigned address number. The sampling rate within the nine strata were chosen to assure that the distribution of the nine strata within the 90,000 addresses was the same as the distribution across the U.S. Next, the random selection of addresses for the three panels also assured each had a distribution across the nine strata that was comparable to the distribution across the 90,000 addresses. The process assigned the 1st address to the 1st panel, the 2nd address to the 2nd panel, the 3rd address to the 3rd panel and repeating the assignments throughout the file. The comparability of the three panels was important for the comparison of the effectiveness of the mail strategies. There was no sampling of group quarters. For our retrospective analyses, we have combined all three panels since panel assignment was random across the 90,000 addresses. Essentially, we are assuming that differences in self-response due to the mail strategies would affect all subgroups and in particular, all ESRI Tapestry segments in a comparable manner. The preservation the U.S. distribution within the 40 Tapestry segments found in 90,000 addresses was not part of the design, which may be considered a limitation of the study.

Each panel of addresses was assigned to one of three self-response contact strategies described below. Table 1 has the mailing contact strategies, the sample sizes for each treatment, and the dates of communications sent to the sample addresses.

1. **"Notify Me" Pre-registration Mailing Panel.** The addresses in this panel first received a pre-notice postcard inviting them to visit the pre-registration website to sign up to receive a notification when it was time to respond to the test census. Pre-registrants could choose to be notified by text or email. The Census Bureau sent pre-registered households as many as three notifications about the test census via their chosen medium. If there was no self-response after the third notification, a paper questionnaire was sent. Sample households in the Pre-registration mailing panel that did not preregister were sent ID panel mail materials, which are described later.

2. **ID Mailing Panel.** The first contact to the addresses in this panel was a letter along with an instruction card on how to respond online using the assigned Census ID. This strategy of initially directing respondents to the Internet is known as the Internet First. After the initial contact letter, sample addresses received as many as two additional reminder postcards in the mail. If there was no self-response from the addresses, a paper questionnaire was sent. Recipients still had the option of responding without the preassigned Census ID.

3. **Non-ID Mailing Panel.** This panel was designed to test the Census Bureau's automated matching and geocoding processes for cases without a preassigned, unique Census ID. The addresses in this mailing panel received standard Internet Push mail materials, but they were not provided a Census ID. Instead, respondents provided their address as part of the self-response process. After the initial contact letter,

sample addresses received as many as two more reminder postcards by mail. If there was no self-response, a paper questionnaire was sent (Vines et al. 2017).

Table 1: 2015 Census Test Mail Contact Panels

Mail Panel	*Sample Size*	*Pre-Notice (Feb 23)*	*Contact 1 (March 23)*	*Contact 2 (April 1)*	*Contact 3 (April 8)*	*Contact 4 (April 15/16)*
"Notify Me" preregistration	30,000	Postcard invite to register	Choice of email or text*	Choice of email or text*	Choice of email or text*	Mail questionnaire
ID	30,000	N/A	Letter	Postcard	Postcard	Mail questionnaire
Non-ID	30,000	N/A	Letter (no ID)	Postcard (no ID)	Postcard (no ID)	Mail questionnaire

**Pre-registrants were sent their choice of email or text messages. Otherwise, addresses in the Pre-registration Panel that did not preregister were sent mail materials identical to the ID panel.*

The overall participation rate for the three panels in the 2015 Savannah Census Test site was 61.0 percent. The Non-ID panel had a 58.1 percent participation rate while the ID panel was 61.2 percent and the Pre-registration panel was 60.2 percent (Vines et al., 2017 p. 50-51). The participation rate of the Non-ID panel was 3.1 percent (S.E.=0.4) lower than the participation rate in the ID panel with a p-value < 0.05, which was statistically significant. On the other hand, the difference between the participation rates in the ID panel and the Pre-registration panel, which was 1.0 percent (S.E.=0.4) had a p-value < 0.05 and therefore, was not statistically significant.

References

Bates, N. and Mulry, M. H. 2011. Using a Geographic Segmentation to Understand, Predict, and Plan for Census and Survey Mail Nonresponse. *Journal of Official Statistics.* 27(4). 601–618. http://www.jos.nu/articles/abstract.asp?article=274601 (accessed July 2017).

Blumerman, L., Bishop, D., Moffett, E., Bentley, M., Boone, T., and Chapin, M. 2015. The 2020 Census Operational Plan. 2020 Census Program Management Review. October 6, 2015. Washington, DC:U.S. Census Bureau. https://www2.census.gov/census_2020/pmr_materials/2015-10-06/3%202020%20census%20op%20plan%2010_6_2015%20v1%20final.pdf (accessed July 2017).

Erdman, C. and Bates, N. 2017. The Low Response Score: A Metric to Locate, Predict, and Manage Hard-to-Survey Populations. *Public Opinion Quarterly.* 81 (1). 144-156. DOI: https://doi.org/10.1093/poq/nfw040

ESRI. 2015. Tapestry Segmentation: Methodology. Redlands, CA:ESRI. Available at: http://downloads.esri.com/esri_content_doc/dbl/us/J9941Tapestry_Segmentation_Methodology.pdf (accessed July 2017).

Experian Marketing Services. 2017. ConsumerView. New York:New York. http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiRr8C-gc3VAhWCJCYKHZVkAGwQFggmMAA&url=http%3A%2F%2Fwww.experian.com%2Fmarketing-services%2Ftargeting%2Fdata-driven-marketing%2Fconsumer-view-data.html&usg=AFQjCNGCVMgCwQPZYxVBxIM-bg_xwkNlPg (accessed August 2017).

Matthews, K. and Rothhaas, C. 2015. Sample Design Specifications for the 2015 Census Savannah Site Test. DSSD 2020 Decennial Census R&T Memorandum Series #E-05. Washington, DC:U.S. Census Bureau.

Mulry, M. H., Mule, T., and Clark, B. 2016. Using the 2015 Census Test Evaluation Follow-up to Compare the Nonresponse Follow-up with Administrative Record. In *JSM Proceedings,* Survey Research Methods Section, American Statistical Association. Alexandria, VA: American Statistical Association. 503-516. http://ww2.amstat.org/sections/srms/Proceedings/y2016/files/389527.pdf (accessed July 2017).

Pew Research Center. 2017. Mobile Fact Sheet. Dated January 12, 2017. Washington, DC:Pew Research Center. Available at: http://www.pewinternet.org/fact-sheet/mobile/ (accessed July 2017).

U.S. Census Bureau. 2017. Savannah Area Test Site. Washington, DC: U.S Census Bureau. https://www.census.gov/programs-surveys/decennial-census/2020-census/research-testing/testing-activities/2015-census-tests/savannah.html (accessed July 2017).

Vines, M., Bates, N., Virgile, M., Walejko, G., Hagedorn, S., McCaffrey, K., and Otmany, J. 2017. 2020 Census Research and Testing: 2015 Test of Digital Advertising and Other Communications in the Savannah DMA. 2020 Census Program Internal Memorandum Series: 2017.14.8. May 17, 2017.