Paid Media and the 2010 Census Communications Campaign: An Experiment to Assess Increased Spending[1]

Nancy Bates[1], Kristin McCue[2] and Mike Lotti[3]

[1]U.S. Census Bureau, Research and Methodology Directorate, Washington, D.C. 20233
[2]U.S. Census Bureau, Center for Economic Studies, Washington, D.C. 20233
[3]Accretive Insights, Lady Lake, FL. 32162

Abstract

As part of the 2010 Census integrated communications campaign (ICC), the Census Bureau contracted with an advertising agency to deliver a paid advertising campaign to increase census awareness and participation. This paper reports results from an experiment conducted to assess the impact of increased media spends in a small number of areas. Media outlets included television, radio, print advertisements, and digital advertising. This paper addresses two questions under the experiment. First, were increases in the level of paid media related to increases in census form mail returns? Second, were increases in the level of paid media related to increased census awareness and advertising recall? Results suggests that, for the most part, the increased media spends did not result in higher levels of advertising recall or higher mail return rates over and above the changes produced by the baseline integrated campaign. One explanation is that the baseline advertising levels were already at or near saturation.

Key Words: 2010 Census, communication campaign, experiment

1. Background

Paid advertising was used for the first time in a U.S. Decennial Census in 2000. During that campaign there were no controlled experiments carried out to vary the planned "dosage" of paid media administered in one geographic area versus another. It has been previously recommended that the Census Bureau undertake an in-market controlled test during the 2010 Census to assess the impact of paid media on census mail return. This recommendation was made in several Census 2000 evaluation reports, written recommendations from the Census Advisory Committees of Professional Associations, and in recommendations from an independent Academic Assessment Panel (see U.S. Census Bureau, 2009; Academic Assessment Panel, 2009; National Research Council 2004; Edwards and Wilson, 2004; Wolter et. al, 2002).

To act on these recommendations, the Census Bureau conducted a controlled experiment in the 2010 Census. Specifically, eight matched-pair sites were selected and one site from each pair was randomly selected to receive an increase in the amount of paid advertising "dosage" while the other half received the normal "dosage" applied to areas not included in the experiment. Such a design is often referred to as a "heavy-up" test. Our test is referred to as the paid advertising heavy up experiment or PAHUE.

1.1 Challenges implementing a heavy-up test

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

Given the 2010 Census ICC environment, there were challenges and risks to implementing a successful heavy-up test. First, the ICC was a nationwide, multi-audience, multi-component campaign that consisted not only of paid advertising but also earned media and large national and community-based Partnership Programs. Our experimental design only manipulated the paid media component and within that component, only one of the media plans (the Diverse Mass plan or DM). This imposed limitations on both the analyses and inferences in the test.

Second, there were risks that could undermine the test design. One was the risk that the additional heavy-up buys would not be executed exactly according to the heavy-up media plan. This can happen when a desired media buy at a given time slot in a particular medium or in a geographic area are not available. Another risk was that special interventions might occur in one or more of the treatment sites to correct for lower than anticipated mail response. This could contaminate the test to some extent. Finally, there was a risk that one or more of the treatment sites would be at or near saturation levels of paid advertising. Heavy-up tests have difficulty detecting an effect from increased media investment if the base plan is near saturation.

2. Methodology

Advertising response tests often use pairs or groups of markets that are matched on demographic and behavioral characteristics. Our matching was done using the geographic unit used to purchase broadcast media -- geographic areas known as designated market areas (DMAs). The experiment matched DMAs on several factors including population size, race and ethnic compositions, presence of media outlets, hard-to-count scores, and mail return rates from Census 2000 and the American Community Survey.

After carefully weighing the match criteria and consulting with the advertising vendor, we selected eight market pairs that would receive sufficiently similar baseline paid media treatments in sufficiently similar environments as to be considered evenly "matched". One DMA of each pair was randomly selected as the control site and the other was designated as the treatment site. The pairs are listed in Table 1 below.

Table 1: Experimental DMA Pairings

Control		**Treatment**
Columbia , SC	matched to →	Savannah, GA
Toledo, OH	matched to →	Flint-Saginaw-Bay City, MI
Jackson, MS	matched to →	Montgomery, AL
Augusta, GA	matched to →	Baton Rouge, LA
Tallahassee-Thomasville, FL	matched to →	Shreveport, LA
Lubbock, TX	matched to →	Odessa-Midland, TX
Joplin, MO	matched to →	Erie, PA
Little Rock-Pine Bluff, AR	matched to →	Jacksonville, FL

As part of the baseline advertising plan, four of the pairs were slated for inclusion in the Black Audience Plan (BAP). This was a targeted emphasis plan to reach DMAs with sizeable populations of Black adults 18 and older including African Americans, Afro-Caribbeans, Haitians and Black Africans. Pairs that received the BAP included: Columbia-Savannah, Jackson-Montgomery; Augusta-Baton Rouge; and Tallahassee-Shreveport. None of the other pairs fell into any other targeted emphasis plan.

On average, the DMAs represented areas with a population aged 18 and older of around 500,000. We acknowledge the selected pairs do not represent a random sample of all DMAs. Consequently, the inferences in this report cannot be generalized more widely to all DMAs.

The media strategy for the treatment sites was to use local media buys to increase the media level in the mass communication base plan by 100 percent. Industry media level tests typically use a 50 percent - 100 percent increase in media level. Lodish, et al. (1995) found an average increase of 85 percent for established brands over the 141 tests they analyzed. Our media buys were designed to replicate the timing, media mix, and mix of specific messages from the national plan in the form of a "heavy-up" media schedule. The goal was to effectively double the media intensity in the treatment DMAs as compared to the control DMAs. It is important to note that the extra media purchases were restricted only to advertising targets of the *DM campaign* – that is anyone who consumes English language media. The additional buys occurred in local television and radio broadcasts, print media, and online media.

A critical part of the 2010 ICC involved an advertising budget reserve to make last minute ad buys in underachieving markets (referred to as the Rapid Response Program). These markets were identified by consulting daily mail participation rates. In mid-April we were notified that incremental Rapid Response radio and television media buys had been placed in five of the eight PAHUE pairs of sites. In two of the pairs (the Augusta-Baton Rouge and Jacksonville-Little Rock pairs) the extra media buys were purchased in the *treatment* sites, therefore not compromising the integrity of the experiment. In the remaining three pairs (Flint-Toledo, Shreveport-Tallahassee, and Lubbock-Odessa), the extra buys occurred in the *control sites*, causing an imbalance in the heavy-up test. At the request of the Census Bureau, the advertising agency made additional last-minute radio and television buys in the *treatment sites* for these three pairs to try to recalibrate back to a doubling of media spend. In two of the sites, we were able to buy at least half the extra spends that occurred in the control (in Flint-Saginaw and Shreveport). However, due to market saturation, the agency was not able to purchase enough extra advertising in the remaining treatment site (Odessa) to completely rebalance that pair.

Due to the timing of the buys, localized nature of the buys, and inability to completely rebalance the pairs, the Rapid Response interventions undoubtedly introduced some element of contamination to the larger test.

2.1 Media spend audit

Prior to addressing the major research questions of interest, we sought to understand how successful we were in increasing the paid advertising in the treatment areas. Because the experiment was limited to increased advertising as part of the DM audience plan and relied on local media buys to deliver the extra dosages, restrictions were placed on the experiment.

One way to assess the level of exposure to advertising in a market is through the rating points delivered during the campaign. The *rating* of an advertisement is the percentage of the target audience which had an opportunity to see or hear a given ad (i.e., tuned to the TV program, read the magazine, or listened to the radio at that moment). Each *rating point* corresponds to 1 percent of the target population having the opportunity to see an ad. The *Gross Rating Points* (GRPs) figure for a campaign is the sum of all of the ratings for all of the ads in the campaign. In practice, GRPs are measured only for television and radio.

One way of interpreting the overall impact from a given level of GRPs on the target audience is in terms of the *reach and frequency* of the paid advertising campaign. The *reach* of a campaign is the percentage of the total target audience who were in the audience rating for at least one ad. The *frequency* of a campaign is the average number of times a target audience member who was reached by the campaign had the opportunity to be exposed to an advertisement. Frequency can be calculated by dividing the total GRPs by the percent of the population reached. Across all DMAs, the average DM GRP deliveries for the Awareness, Motivation, and Rapid Response phases of the campaign were 1,428, 2,004, and 354, respectively, for a total of 3,786 GRPs. The estimated reach in the Awareness and Motivation phases was 97 percent. The Rapid Response phase was coincident with the Motivation phase and contributed to its

97 percent reach. This works out to the average person having about 39 potential exposures to the DM advertising during these two time periods.

Table 2: Diverse Mass Plan Gross Rating Points for Matched Pair DMAs

Control Site	Total GRPs	Treatment Site	Total GRPs	Percent Increase in Treatment DMA
Columbia	3,480	Savannah	6,128	76%
Toledo	3,504	Flint	6,663	90%
Jackson	3,597	Montgomery	6,091	69%
Augusta	3,663	Baton Rouge	6,158	68%
Tallahassee	3,601	Shreveport	5,850	62%
Lubbock	3,372	Odessa	5,937	76%
Joplin	3,442	Erie	6,131	78%
Little Rock	4,018	Jacksonville	7,203	79%

Source: DraftFCB database.

Table 2 contains the measured GRPs from the *Diverse Mass Plan* for adults 18 and older delivered by television and radio in the eight matched pairs. Despite the Rapid Response interventions and other differences between planned and actual spends, the experiment came close to achieving the goal of doubling the DM paid media GRP dosage in most of the treatment sites for the measured media of television and radio. For the DM plan, we see that GRP percentage increases in the treatment sites range from 62 to 90 percent.

However, when we consider the *total combined audience GRPs* (which include DM rating points plus those associated with the targeted audience plans) differences between treatment and control GRPs are greatly reduced in the pairs affected by targeted audience plans (see Table 3). The "watering-down" of the heavy-up dosage is very evident in these pairs (i.e., Columbia-Savannah; Jackson-Montgomery; Augusta-Baton Rouge; and Tallahassee-Shreveport).

Table 3: Total Combined Audience Gross Rating Points for Matched Pair DMAs

Control Site	Total GRPs	Treatment Site	Total GRPs	Percent Increase in Treatment DMA
Columbia	5,664	Savannah	7,684	36%
Toledo	3,994	Flint	7,171	80%
Jackson	6,590	Montgomery	8,895	35%
Augusta	5,856	Baton Rouge	8,013	37%
Tallahassee	5,448	Shreveport	7,718	42%
Lubbock	4,092	Odessa	7,285	78%
Joplin	3,613	Erie	6,301	74%
Little Rock	4,729	Jacksonville	8,490	80%

Source: DraftFCB database.

Using the total GRPs combined across audience plans, we can observe the frequency with which the DMA audience had the opportunity to see Census Bureau paid advertising. For example, in Table 3 the total combined audience GRPs in the control site Toledo is 3994. So, the estimated average frequency of exposure to a Census television or radio ad between January 11 and April 18 for an adult in Toledo aged 18 and older was 41.18 times (3,994 GRPs / 97 percent reach = 41.18).

This compares to 73.93 times in the paired treatment site of Flint (7,171 GRPs / 97 percent reach = 73.93).

These numbers give a sense of the breadth and depth of the campaign. Even in the control sites, the media spend levels were among the top five heaviest advertising campaigns during the time period. Between January and June 2010, the Census Bureau ranked fourth out of the top six advertisers behind McDonalds, Wal-Mart and Geico but ahead of Budweiser and Nike. Knowing this provides context as to whether the control sites may have been saturated – given such a large control dosage, more advertising may have little effect.

In summary, the audit revealed our experiment fell short of delivering double media dosages in the test areas and that baseline dosage levels were already quite high in many places.

3. Results

3.1 Mail Return Rates

The primary research question under the test is whether increased media spends were associated with increases in mail returns. To explore this, we used operational data from the Census for all housing units in the PAHUE DMAs that were eligible to mail back a census form. This provides information on the primary outcome of interest—mail return rates—along with factors that might influence rates of return such as household composition, housing characteristics, and whether a replacement form was sent. These variables, along with the level of paid advertising GRPs were used to analyze the determinants of the share of households that completed and mailed back a census form by the cutoff date for sending cases to personal visit follow-up (April 19, 2010).

The combined mail return rate was 75.3 percent across all control sites and 74.9 percent in the treatment sites (see Table 4). Results across pairs were mixed. For three of the pairs, the rates are in the expected direction with the treatment sites having higher mail returns. For another three pairs, the control sites had higher returns and for two pairs (Tallahassee-Shreveport and Little Rock-Jacksonville), the difference was negligible (less then one-half of a percent).

Table 4: Mail Return Rates for the Eight PAHUE DMA Pairs

Control Site	Mail Return Rate	Treatment Site	Mail Return Rate	T-C Difference
1. Columbia	78.3%	Savannah	74.4%	-3.9
2. Toledo	78.6%	Flint-Saginaw	83.4%	4.8
3. Jackson	72.4%	Montgomery	73.1%	0.7
4. Augusta	76.0%	Baton Rouge	69.2%	-6.8
5. Tallahassee	72.6%	Shreveport	72.4%	-0.2
6. Lubbock	74.9%	Odessa	72.0%	-2.9
7. Joplin	76.3%	Erie	82.0%	5.7
8. Little Rock	73.1%	Jacksonville	72.8%	-0.3

Source: 2010 Mail Response/Return Assessment data for PAHUE areas.

Table 5: Mail Return Rates by Audience Segment, Treatment Versus Control

Audience Segmentation Cluster	Control MRR	Treatment MRR	Treatment-Control Difference
All Around Average (homeowner skew)	75.9%	75.1%	-0.8
All Around Average (renter skew)	72.0%	71.2%	-0.8
Economically Disadvantaged (homeowner skew)	70.5%	69.3%	-1.2
Economically Disadvantaged (renter skew)	63.0%	64.8%	1.8
Ethnic Enclave (homeowner skew)	72.2%	70.2%	-2.0
Ethnic Enclave (renter skew)*	n/a	n/a	n/a
Mobile singles	65.8%	66.8%	1.0
Advantaged Homeowners	81.2%	81.6%	0.4

* Too few tracts in the PAHUE fell into this cluster to be represented here.
Source: 2010 Mail Response/Return Assessment data for PAHUE areas.

Table 5 presents the mail return rates by audience segments by treatment and control. These segments were developed as part of the ICC audience segmentation. In all, eight clusters were identified with each Census tract assigned to one cluster. The clusters include: Advantaged Homeowners, All Around Average (owner skewed), All Around Average (renter skewed), Economically Disadvantaged (owner skewed), Economically Disadvantaged (renter skewed); Ethnic Enclave (owner skewed); Ethnic Enclave (renter skewed) and Mobile Singles. For more information, see Bates and Mulry, (2011).

The difference in mail return rates between treatment and control are small for most segments – usually less than one percent. The Economically Disadvantaged renter skewed households had slightly higher mail returns in the treatment areas while the Ethnic Enclave homeowner skewed had slightly lower.

One of the limitations of the PAHUE and most in-market media tests is the small number of test pairs. As such, the analysis is sensitive to DMA-to-DMA differences that would exist in the absence of any experimental treatment. Some across-DMA variation in return rates may be due to factors we can measure—for example, the size of the foreign-born population or activities and outreach of the local 2010 Partnership Program. If large differences exist, it can reduce the ability to detect a significant difference in mail return associated with the treatment. One way to explore this is to examine difference in MRRs between treatment and control DMAs for narrower groups.

Pursuing such a strategy, we examined mail return rates for each DMA by race/ethnic group (data not shown). The variation in mail return rates between the control sites is large for some groups. For example, mail return for the Black population ranged from a low of 61.3 percent in Little Rock to as high as 73.8 percent in Columbia. This suggests that other covariates are probably in play and adding noise to our ability to detect the heavy-up's impact on differences in mail return associated with the treatment dosages. But with many potentially inter-related factors, a strategy of examining one or two factors at a time is unlikely to make clear whether the differences we know exist in these areas are confounding the pattern we expected to find. With this in mind, we use multivariate statistical models to control for an extensive set of characteristics that may influence the likelihood of a mail return.

Our goal here is to determine if there is a significant difference in response rates across treatment and control sites when we compare households with similar characteristics. We carry out this analysis using logistic models, in which the dependent variable equals 1 if the household returned a form by April 19, 2010, and 0 if not. In the results presented here, we use the logarithm of the GRPs associated with a

householder's race/ethnic group in the DMA they live in as our measure of the experimental treatment. We have also estimated models that parallel those presented here but use (i) an indicator variable for treatment sites, and (ii) use GRPs without taking the logarithm. The three alternatives we investigated gave quite similar results, and we chose to present the log GRP results because the assumption underlying this functional form is more consistent with the idea that there are decreasing returns to increasing the level of advertising. The log GRP form implies, for example, that increasing GRPs by 100 points has a much larger effect on mail return rates when moving from 500 to 600 GRPs than it does when moving from 5000 to 5100 GRPs.

We include indicator variables for the relevant DMA pair for each household in order to account for mean differences in response rates across pairs. This leaves only within-pair MRR differences to be picked up by the log GRP variable. We include a large number of household characteristics, such as whether the household rents or owns its residence, whether it is headed by a married couple, householder age, householder race/ethnicity, the number of household residents, and how many of them are children. We also include several controls for differences in how an address was treated during survey operations: whether the household was scheduled to receive a replacement form, whether it received a bilingual form, and how forms were delivered to the address.

Finally, we include a measure of the level of partnership activity in the county in which each household lives. Census partnered with a wide range of groups to increase awareness of the census and encourage households to return their forms. While precisely measuring differences across areas in the levels of this activity would be very difficult, the Census Bureau did maintain a database that included county-level information on activity such as the number of active partners, the dollar value of grants awarded for partnership support, and counts of local partnership staff. Here we used a scale measure based on these data that was developed by NORC in consultation with the Census Bureau's Field Division (see Datta, et al., 2012 for details). It is constructed so that a one unit increase in the scale represents a one standard deviation increase in the level of measured partnership activity in the household's county of residence.

Table 6 presents odds ratios for the effect of each of the variables included as controls in our model. The first line gives the odds ratio and confidence interval around it for log GRP. An odds ratio above 1 means that a characteristic is associated with a higher value of the dependent variable, or in this case, with a higher likelihood that a household returns a form prior to NRFU. If the confidence interval around the odds ratio includes 1 in its range, then the effect of that control is not significant. Here the odds ratio on the log GRP is slightly above 1, but is not significantly different from 1.

A number of characteristics are associated with significant differences. For example, replacement forms were associated with a significant increase in mail returns, while bilingual forms were not (though the PAHUE DMAs may not have included enough Spanish-speaking households to provide a good sample to identify their effects). Several household demographic characteristics had significant effects in the expected direction. For example, married couple households had relatively high rates of return, renters had low rates relative to owners, and the age of the householder was consistently positively associated with the likelihood of a mail return. The partnership activity scale had no significant relationship to MRRs, though this may be due to the difficulty of quantifying geographic variation in the effectiveness of partnership activity rather than a lack of effect.

Thus, while we find that there are many factors that significantly affect mail return rates, the estimated PAHUE effect is still not significantly different from zero. For the overall sample, differences in the characteristics we can measure cannot explain the lack of evidence for a positive treatment effect. Comparing households that are similar along the dimensions included in the model, there is no statistical evidence that the households receiving the heavy-up treatment had higher mail return rates.

Table 6: Logistic Model of Mail Return Rate, All Eligible Households

Dependent var = Valid mail return by 4/19/10	Odds Ratio		95% Confidence Interval
Log (GRP)	1.013		[0.957,1.072]
Type of enumeration area			
Update/leave enumeration area	0.710	***	[0.644,0.783]
Military enumeration area	1.035		[0.869,1.233]
Urban update/leave enumeration area	0.619	***	[0.554,0.691]
Outcome of 2000 collection at this address			
Address not in 2000 census	0.824	***	[0.800,0.849]
Enumerator completed return in 2000	0.635	***	[0.623,0.648]
Other outcome in 2000	0.719	***	[0.705,0.734]
Type of housing unit			
Multi-unit building	0.857	***	[0.836,0.878]
Trailer/mobile home/boat/RV, etc	0.737	***	[0.714,0.761]
Married couple	1.608	***	[1.587,1.630]
Household rents or occupies without paying rent	0.627	***	[0.613,0.641]
DMA pair for PAHUE			
Savannah/Columbia SC	1.244	*	[1.010,1.534]
Flint-Saginaw/Toledo	1.433	*	[1.044,1.967]
Selma/Jackson MS	1.043		[0.976,1.114]
Baton Rouge/Augusta	0.998		[0.832,1.198]
Shreveport/Tallahassee	1.098	*	[1.005,1.199]
Odessa-Midland/Lubbock	1.192		[0.858,1.654]
Erie/Joplin-Pittsburg	1.298		[0.992,1.700]
Household of unrelated people	0.717	***	[0.688,0.747]
Partnership activity scale (#1)	0.999		[0.991,1.007]
Age of householder			
Householder age <=25	0.744	***	[0.720,0.769]
Householder age 41-64	1.596	***	[1.561,1.632]
Householder age 65+	2.991	***	[2.847,3.141]
Number of children age<18 in household			
1 child	0.959	**	[0.940,0.979]
2 children	1.008		[0.994,1.022]
3-4 children	0.978		[0.956,1.000]
5-8 children	0.870	***	[0.826,0.917]
9+ children	0.745		[0.470,1.182]
Number of residents in household			
2 people	1.181	***	[1.171,1.191]
3 people	1.033	**	[1.014,1.052]
4 people	0.939	***	[0.921,0.958]
5 people	0.820	***	[0.801,0.839]

Table 6: Logistic Model of Mail Return Rate, All Eligible Households (continued)

Dependent var = Valid mail return by 4/19/10	Odds Ratio		95% Confidence Interval
9-12 people	0.885	**	[0.827,0.947]
13+ people	0.31	***	[0.222,0.434]
In area blanketed by replacement forms	1.215	***	[1.146,1.289]
In area where non-respondents received replacement form	1.128	***	[1.075,1.184]
Address is not residential	0.035	***	[0.026,0.047]
Non-Hispanic Black	0.748	***	[0.690,0.811]
Hispanic	0.607	***	[0.526,0.700]
Received bilingual form (English/Spanish)	1.045		[0.957,1.141]

Note: Omitted categories: Mail-out/mail-back enumeration area, address had respondent return in 2000, owner, single-family house, Jacksonville/Little Rock DMA pair, single person household, non-Hispanic White, householder aged 26-40.

p-values: *p<0.05, **p<0.01, ***p<0.001
Source: 2010 Mail Response/Return Assessment data for PAHUE areas.

Next, we examined different racial/ethnic groups, to see if we find any evidence of significant effects for some sub-populations. In doing so we use essentially the same model for each group.
We examined the estimated effects for three groups that are defined by the race and ethnicity of the householder: non-Hispanic Whites, non-Hispanic Blacks, and Hispanics. In general, the odds ratios estimated for the demographic, housing and survey variables do not vary substantially across groups, so we present only the estimated treatment effects for these specifications. Again, we find no significant difference in MRRs associated with higher GRPs (table 7).

Table 7. Logistic Model of Mail Return Rate by Race/Ethnicity of Householder

Group	Odds Ratio for Log(GRP)	95% Confidence Interval
Non-Hispanic Whites	1.003	[0.971,1.036]
Non-Hispanic Blacks	0.956	[0.900,1.015]
Hispanics	0.903	[0.815,1.001]

Notes: The estimates presented here come from models with the same set of controls listed in Table 6.
Source: 2010 Mail Response/Return Assessment data for PAHUE areas.

In Table 8, we apply the logistic model to sub-populations defined by the primary audience segment for each tract. Here we do find a significant effect for one sub-population: Mobile singles. This group has roughly a 66 percent mail return rate on average. The point estimate from the model implies that, starting from 66 percent as a base, a 50 percent increase in GRPs would increase the mail return rate by roughly one percentage point. The effect for a second segment, Advantaged Homeowners, is significant at the 10 percent level. Both of these segments would be expected a priori to be more likely affected by media advertising than by partnership activities.

Table 8. Logistic Model of Mail Return Rate by Audience Segment

Segment	Odds Ratio for Log(GRP)	95% Confidence Interval
All around average owner/renter	0.997	[0.944,1.053]
Economically disadvantaged owner/renter	0.971	[0.877,1.076]
Mobile singles	1.240*****	[1.025,1.500]
Advantaged home-owners	1.082	[0.984,1.189]

Notes: Audience segment is determined by the characteristics of a household's tract of residence, as measured in 2000. The estimates presented here come from models with the same set of controls listed in Table 6.
p-values: *p<0.05, ** p<0.01, *** p<0.001
Source: 2010 Mail Response/Return Assessment data for PAHUE areas.

3.2 Self-reported ad exposure and recall

A second question under the test was whether residents recalled different levels of advertising exposure and recall between the test and control sites. To assess this, survey data were analyzed. The National Opinion Research Center (NORC) conducted a set of three household surveys based on a nationally representative sample and oversampling of minority populations. This data collection is known as the 2010 Census Integrated Communication Program Evaluation (CICPE). The survey items included measures of self-reported Census ICC exposure, Census awareness, knowledge, and attitudes and self-reported intent to participate. The first wave was conducted prior to the paid advertising to assess baseline levels (mid-September to mid-January, 2010). The second wave took place January 19 through March 18, 2010 during the peak of the paid media campaign and the third wave was conducted mid-April through mid-July 2010 when door-to-door enumerators conducted nonresponse followup interviews.

Approximately 2,000 completed interviews were collected in PAHUE DMAs in both the wave 1 and wave 3 field periods[2]. This provided pre and post heavy-up intervention interviews. Wave 1 had a 68.3% response rate and wave 2 had 70.8%[3]. Readers can find more information about CICPE data collection in Datta et al., 2012.

We know from the ratings data (GRPs) that there were substantial increases in actual exposure to the ads. The CICPE survey allows us to measure perceived exposure by examining treatment-control differences in self reports. In both waves 1 and 3, respondents were asked a series of questions about whether they recalled hearing or seeing information about the census. The survey asked separately about exposure via different media outlets, including paid advertising, Census's community partners, and various sources of news coverage.

We note, however, that self-reports of media exposure are prone to measurement error and far from a perfect indicator. For example, in their investigation of error in self-reported exposure to 2010 Census paid advertising, NORC found confirmed awareness to be low relative to self-reported recall. That is, the majority of respondents who claimed to see a particular ad failed to recall meaningful details to confirm

[2] Data were not collected in the PAHUE DMAs during wave 2 of the CICPE.
[3] Response rates were calculated using the AAPOR RR2 formula (see AAPOR, 2011).

awareness. Further, the degree of confirmed awareness was found to vary by age, education, and media outlet (Datta et al, 2011). This is a limitation to keep mind.

Table 9 present measures of self-reported exposure to paid advertising. The measures are based on the following questions from the CICPE survey, where question 18 was asked only if the respondent said yes to one of the parts of question 17:

> Q17. Have you heard or seen advertisements about the census...a. on television? b. on the radio? c. in magazines? d. in newspapers? e. on the Internet? f. in other places such as coffee cups, billboards, or park benches?
>
> Q18. Thinking about all of the advertisements you heard or saw in the **past 30 days** about the census, how many different times in the **past 30 days** would you say you saw or heard something about the census? [Note that the wave 3 questionnaire asked about exposure in the **past 90 days**.][4]

Table 9 presents estimates of whether respondents reported having seen any ads (column on left) and of the number of times respondents reported seeing or hearing a census ad (column on right). Three tests of differences are presented. The first test is the difference between the wave 1 treatment and wave 1 control. This provides a benchmark of how close the two measures were prior to advertising. The second is the difference between wave 1 and wave 3 for the control cases. This is a good indicator to measure change over time because of *normal* campaign levels. It also serves as an indicator of potential saturation as a result of normal campaign dosages – that is, if wave 3 levels for the control are very high, we might conclude the selected markets were at or near the saturation point. The final test is whether the *difference* between the treatment and control wave 1-wave 3 differences is significant – this is the test of most interest because it tells us whether differences between the treatment and control were large enough to be statistically meaningful. Each of the tests presented in the table is based on a t-test of the hypothesis that differences equal zero.[5]

[4] Categories: None / 1-2 times / 3-5 times (W1) / 3-15 times (W3) / 6-10 times (W1) / 16-30 times (W3) /11 times or more (W1) / 31 times of more (W3)

[5] The standard error used to calculate the t-statistic is estimated using Taylor series (or linearization) methods that adjust for the complex design of the CICPE survey

Table 9: Exposure to Paid Advertising About Census, Treatment Versus Control

	Percent Heard or Saw at Least One Census Ad		Number of Times Heard or Saw Paid Ad in Last…	
	Wave 1	Wave 3	Wave 1 (Last 30 days)	Wave 3 (Last 90 days)
Treatment	30.4	76.0	0.81	10.37
	(2.6)	(2.0)	(.08)	(0.97)
Control	37.7	76.5	1.18	9.50
	(3.5)	(4.4)	(0.11)	(0.80)

Test	Estimate	Estimate
Treatment minus control in wave 1	-7.3 (5.5)	-0.37** (0.13)
Change for control (wave 3 minus wave 1)	38.7*** (3.2)	8.32*** (0.73)
Change in treatment minus change in control	6.9 (5.4)	1.24 (1.32)

p-values: *<.10, **<.05, ***<.01

Base: Weighted estimates using all respondents. Respondents who reported they had never heard of the census were coded as having heard or seen no ads. Responses to question 18 were given in intervals (0, 1-2,etc). Each respondent was assigned the mid-point of his or her interval response (e.g., 1.5 for the category 1-2 times). For wave 1 we assigned a value of 15 for the category 11 or more times and for wave 3 we assigned a value of 40 for the category 31 or more times. This is consistent with methods used in Datta, Yan, et al (2011)

Source: Waves 1 and 3 of CICPE survey.

Evidence from the survey indicates that reported exposure to census ads increased between waves 1 and 3. In Table 9, there are large and significant increases in reported media exposure and frequency between waves 1 and 3, and the treatment/control differences in changes are in the expected direction, but not significantly so.

One possible reason we find no evidence of a response to the heavy-up treatment here is that census advertising in non-treated areas was so high that the heavy-up increase occurred in a range of diminishing returns. In other words, the control areas were already saturated with ads to the point that the extra dosage delivered by the heavy-up was impossible to detect using measures of self-reported ad exposure. A final comment on why we failed to detect higher self-reported exposure in the treatment areas may be related to the small differences in GRP delivery noted earlier. Recall that once the *total audience* GRPs were taken into account, the treatment dosages fell far short of a doubling heavy-up.

Another portion of the CICPE survey questionnaire was devoted to assessing the respondents' memory of and reaction to specific ads in the campaign. Respondents were screened early in the questionnaire by asking whether or not they had seen or heard anything about the Census recently. Those who answered positively were probed about four specific ads, a battery of overall reactions to all of the ads in the campaign, and their recall of specific ads using audio and visual prompts. These questions appeared in wave 3 only.

Because the DM campaign was targeted to all English speaking residents, it is likely that most members of the PAHUE sample were exposed to one or more of the DM ads. Consequently, all PAHUE respondents were asked about the first ad, titled "Frank," which was the single ad used most heavily during the Motivation phase of the DM campaign. Respondents were prompted with a short, general description of the ad, referenced below as a "verbal prompt." The purpose was to trigger the respondent's memory of the ad while not providing details of the characters, visuals, or copy in the ad. For those

claiming to have recently seen the ad based on the short description, a follow-up open-ended question probed for any relevant details which the respondent recalled. The claim to have seen the ad was then classified as "confirmed" if the recalled details were sufficiently accurate. If the respondent indicated that they had recently seen the ad, they were asked to estimate how often they had seen it in the past 90 days, asked whether they thought the ad had "grabbed their attention," and whether they thought the ad gave them good reasons to mail back their census form. The exact wording of the questions varied slightly for each ad probed.

The results of these probes should not be taken as literal estimates of the size of the population exposed to the advertising. Advertising uses visual and auditory stimulation to gain attention and create memory. It is likely that many who had seen the ad did not recall it from the simple, verbal prompt. Asking respondents to recall one or more of a set list of specific details in the ad may also raise the bar for "confirmed" recall (for more on a discussion of confirmed recall in the CICPE, see Datta, Hepburn, Yan and Evans, 2011).

Table 10: Ad Recall Measures for "Frank" Comparing Treatment and Control

	Recalled seeing the ad (percent yes)	Recalled details from the ad (percent with at least one detail coded)	Grabbed attention (percent yes)	Gave good reasons to mail census form (percent yes)	Recalled seeing the ad 3+ times (percent yes)
Treatment	17.6 *** (1.76)	7.6 ** (1.79)	14.0 *** (1.23)	12.8 ** (1.32)	11.4 ** (1.44)
Control	11.4 (1.14)	2.5 (0.53)	7.9 (1.57)	8.1 (1.17)	6.6 (0.41)

p-values: *<.10, **<.05, ***<.01, for test of hypothesis that treatment/control difference is zero.
Base: Respondents who reported having seen/heard something about the Census recently. Results are weighted.
Source: Wave 3 of CICPE survey.

As shown in Table 10, the treatment cell has a significant increase in both the recall of "Frank" and in recall of relevant details from "Frank" when compared to the control cell. Table 10 also demonstrates that significantly more households in the treatment cell found that "Frank" grabbed their attention and gave good reasons to mail back the census form than in the control cell. Respondents in the treatment cell also reported that they had seen "Frank" at higher frequencies that those in the control cell. These findings are consistent with higher exposure to advertising in treatment sites.

4. Discussion

The PAHUE was the first test of its kind implemented with a Census Bureau paid media campaign. Census Advisory Committees, academic experts, and program evaluation experts have recommended that this type of controlled in-market test be undertaken to better understand the impact of the ICC,. The test did identify some areas and hypotheses for further exploration.

A close audit of the experimental media spends and GRPs revealed two facts: (1) because the test manipulated only the DM media plan, it fell short of the goal to double media dosages in the test sites and (2) the 2010 ICC base plan advertising levels were likely at or above saturation levels. As a result, the ability to detect a "signal" in the experiment was difficult. Going forward, we recommend the Census

Bureau continue in-market testing of variables subject to Census Bureau control in the communication strategy, but that a heavy-up only design not be repeated.

For the most part, our analysis of mail return rates failed to uncover an association between extra media spends and an increase in mail returns. This could be due to saturation or the presence of unmeasured, uncontrolled covariates affecting return behavior such as the success of Partnership Program efforts in small areas. That our models failed to establish a link between the heavy-up and mail response does not mean the advertising had no effect. It is plausible the baseline levels very much had an effect but that advertising beyond the baseline did not. The ideal future experiment would manipulate dosages in the other direction, that is, decrease media spends in some areas (or even cut them out completely). However, we recognize that conducting such a test may be unlikely given political sensitivities and operational difficulties.

Our analysis also failed to uncover a link between extra media spends and respondent recall of general campaign exposure to advertising. This is not too surprising given potential measurement error associated with self-reports. However, when we asked about a specific ad that was only in the DM campaign (and therefore less prone to saturation), we did find some evidence of enhanced positive communication in the test sites. For more detailed recommendations from the PAHUE, see Bates, McCue and Lotti (2012).

Acknowledgements

We acknowledge Earl Letourneau, Dennis Stoudt and his staff, Burton Reist, Frank Vitrano, Laura Sewell, Alexandra Figueroa, Mike Bentley, Lucia Foster, Tasha Boone, Adrienne Oneto, Fred Meier, and Jason Machowski for their assistance in reviewing the report and making the test possible.

References

Academic Assessment Panel. 2009. "Academic Assessment Panel Final Report". A report submitted to the U.S. Census Bureau, September 8, 2009. http://2010.census.gov/news/releases/operations/independent-panel-commends-the-2010-census-paid-media-plan.html.

American Association for Public Opinion Research. 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th edition. AAPOR.

Bates, N., K. McCue, and M. Lotti. 2012. "2010 Census Paid Advertising Heavy-Up Experiment Evaluation." 2010 Census Planning Memoranda Series, No. 191, April 30, 2012. Available at: http://2010.census.gov/2010census/about/cpex.php.

Bates, N. and M. Mulry. 2011. "Using a Geographic Segmentation to Understand, Predict, and Plan for Census and Survey Mail Nonresponse." *Journal of Official Statistics*, 27, 601-618.

Datta, R., T. Yan, D. Evans, S. Pedlow, B. Spencer, and R. Bautista. 2011. "Final Report: 2010 Census Integrated Communications Program Evaluation (CICPE). National Opinion Research Center at the University of Chicago. March 13, 2012.

Datta, R., P. Hepburn, T. Yan, T. and D. Evans. 2011. "Have you Really Seen this Ad?: Using Confirmed Awareness to Correct Estimates of Self-Reported Exposure." A paper presented at the Joint Statistical Meetings of the American Statistical Association, Miami, FL, July 30-August 4.

Edward, S. and M. Wilson. 2004. Census 2000 Testing, Experimentation and Evaluation Program Topic Report No. 6, TR-6, Evaluation of the Census 2000 Partnership and Marketing Program.

Lodish, L., M. Abraham, S. Kalmenson, J. Livelsberger, B. Lebetkin, B. Richardson, and M. Stevens, 1995. "How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments." Journal of Marketing Research, Vol XXXII (May 1995), 125-139.

National Research Council. 2004. The 2000 Census: Counting Under Adversity. Panel to Review the 2000 Census. Constance F. Citro, Daniel L. Cork and Janet L. Norwood, eds. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.

U.S. Census Bureau. 2009. "Recommendations made by the Census Advisory Committee of Professional Associations (CACPA) at the meetings held April 16-17, 2009." Internal U.S. Census Bureau Memorandum from Jeri Green, Chief Advisory Committee Office, June 19, 2009.

Wolter, K., B. Calder, E. Malthouse, S. Murphy, S. Pedlow, and J. Porras. 2002. Partnership and Marketing Program Evaluation. Final Report. Census 2000 Evaluation, July 17, 2002.