Media, Measurement & Methodology: 5 Keys for Success Today

Tony Incalcatera, Ipsos MediaCT U.S.

Brief History –

For thirty-eight years, prior to 2015, the Ipsos Affluent Survey USA (formerly the Mendelsohn Affluent Survey) had been conducted using a paper questionnaire mailed to potential respondents. Although there were changes to the questionnaire in subject matter and length over that time, the methodology was relatively unchanged. In practice, the paper questionnaire was mailed to *potentially* affluent households that had been drawn from address-based sample. In order to increase the efficiency of reaching suitable households (currently annual household income of US $100,000+), models were utilized to predict affluence. The survey covered **print** publication readership, television viewing, radio listening, and in recent years website visitation. In addition to collecting media behavior, the survey captured data on the ownership and purchasing of products and services, as well as attitudes and lifestyle choices appropriate to those high-income earning adults.

As users of the survey clamored for more and more items to be included in the survey, the questionnaire expanded to 28 pages of condensed type. This expansion obviously contributed to a decline in response rates – a familiar story to virtually every mail-based questionnaire. Unfortunately, the declining response rate also increased the likelihood that non-response bias was impacting study results.

In view of the dramatic changes that were taking place in how people consumed media, the constraints of a paper questionnaire were even more evident. The industry clamored for detailed cross-platform media consumption, which the paper survey simply could not accommodate.

The Challenge –

Ipsos felt that, to meet the industry need for a comprehensive cross-platform measurement, small changes to the survey methodology would not be enough. Expanding the number of pages in the questionnaire, in order to introduce cross-platform measures, would surely cause further erosion to the response rate. Our researchers also had previous experiences where the use of paper surveys to measure non-traditional media alongside traditional media caused confusion among respondents, yielding results that were not reasonable.

As part of our efforts to continually improve the survey instrument, we also wanted to:

- Improve the use of rotations and randomizations within the survey to reduce/eliminate positional biases
- Provide respondents with better instructions and visual cues to aid their recall
- Expand the in-tab to allow for better analyses of low incidence categories
- Improve our weighting structure to align closer to census data
- Introduce year-round fielding to eliminate seasonal bias

The Solution –

With more than 99% of all affluents who responded to the 2014 offline paper survey telling us that they use the Internet on a regular basis, migrating to online data collection made great sense. Internet usage was even high enough at the upper ends of our age cohorts (age 69+) assuring us that a pure online methodology would be able to collect information on all affluent adults.

Early testing of a new approach to data collection focused on using a hybrid methodology, incorporating both offline and online elements. Potential respondents were sent a package that included a copy of the paper questionnaire along with an invitation and instructions to complete the survey online as an alternative. The number of respondents who opted to complete the survey online was too small to even make rudimentary comparisons to the paper questionnaire.

Previous experience in two-stage recruit-to-web studies (via telephone or direct mail) produced low respondent cooperation, rendering the process extremely expensive while yielding poor overall response rates. If the study was going to migrate online, it would have to find an alternative approach. Throughout the testing phase and into the launch of the 2015 survey, we identified 5 key lessons that we feel should help others.

1. Test, test, and test again to build a viable sample.

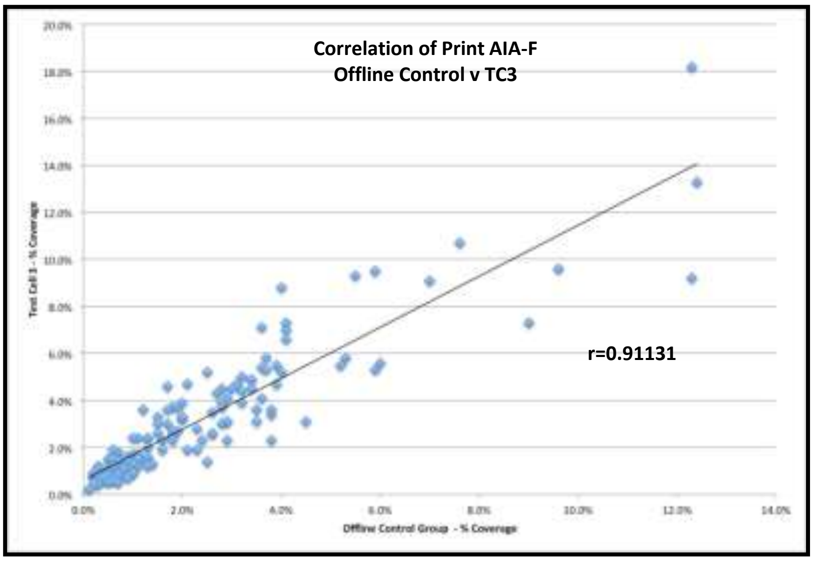
Building on the work of others that migrated media data collection from paper to online, Ipsos began the development of a comprehensive testing program in the fall of 2013. In all there were 9 online test cells covering more than 13,000 completed interviews. The first 4 were fielded concurrent with our 2014 offline mail questionnaire in order to gauge the impact of changing methodologies, while the subsequent 5 were utilized to fine-tune survey methodology.

During the spring of 2014, we fielded a test to compare results from our mail-based questionnaire to an optimized online format (referred to as test cell 3 – TC3). Knowing that a recruit-to-web strategy would yield low levels of cooperation, we designed an intricate quota system, utilizing a variety of online sample providers, in order to match, as closely as possible, the address-based sample employed in our offline group.

The results showed that the optimized online survey would produce readership that was higher than the paper questionnaire, including a significant increase in the average number of publication titles read (+33.9%). We attributed these changes to better survey design including:

- Increased respondent instruction
- Introduction of screening questions
- Logo images in combination with text, replacing plain text
- Publications grouped by genre instead of publishing interval
- Virtually unlimited randomization online instead of 4 printed versions of the paper questionnaire

Despite the increase in overall readership, we saw a reasonable correlation (r=0.911) for print average-issue audience rankings between the online survey and offline paper questionnaire.

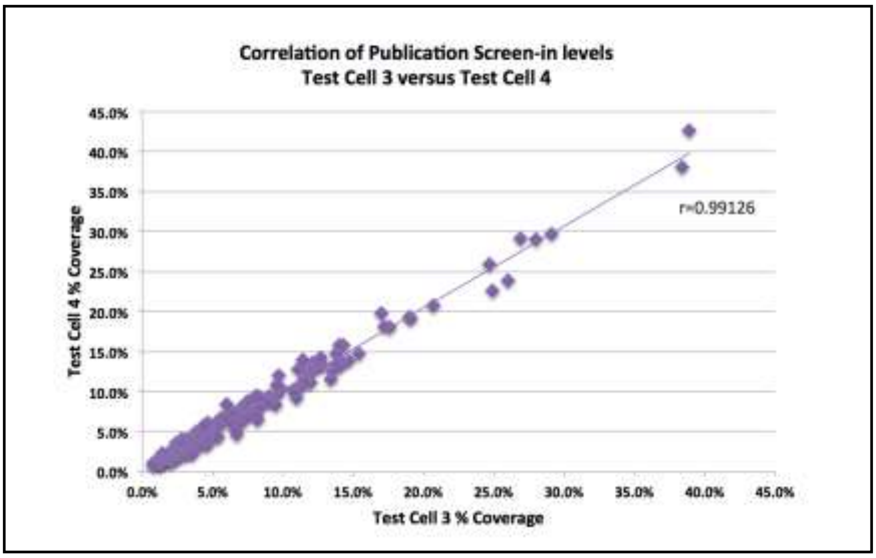


Since the survey design and methodology were so different, we were not surprised to see differences among titles (both increases and decreases in AIA). We were pleased, though, that the relative rankings within competitive sets remained fairly consistent. The chart below shows an example of the ranking order of 3 competitive genres when comparing the average-issue-audience from the paper survey to the online survey. Although there was some movement, the general stability of the competitive set ranking remained.

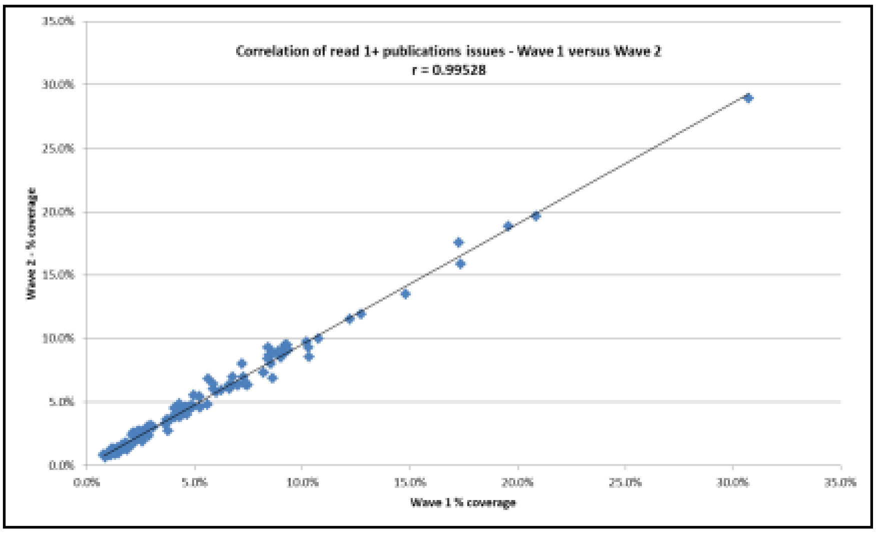
Category A AIA Rank order		Category B AIA Rank Order		Category C AIA Rank order	
Offline	TC3	Offline	TC3	Offline	TC3
1	1	1	1	1	1
2	4	2	2	2	2
3	3	3	3	3	3
4	2	4	4		
5	6	5	8		
6	5	6	5		
7	8	7	7		
8	9	8	6		
9	7	9	11		
		10	9		
		11	10		

2. Make sure your success is repeatable.

The question remained as to whether this relative consistency in audience estimates was coincidence or would be replicable wave after wave. To that end, we repeated the test cell survey with completely fresh sample one month later. The correlation of publication screen-in levels between the two surveys was 0.991, giving us comfort that the methodology would yield consistent results over time, subject to changes in the marketplace.



Based on the results of our testing program, we launched the first wave of the newly revised Ipsos Affluent Survey USA on January 16, 2015. In order to keep on schedule to release results to subscribers in September 2015, we decided to field in two waves. Wave 1 (n=7,176) fielded between January and March for the initial surveys, while Wave 2 (n= 8,091) fielded from April to June. The two waves were combined, and will be released in late September. Now with two completed waves (combined n= 15,267) the correlation of read 1 or more issues of the measured publications between waves was r=0.995.



Thankfully, we did see changes that are likely caused by seasonal media choices. Among other things, we saw distinct shifts in viewing specific sports television networks as the year progressed. Comparing results from the first wave of fielding (mid-January to end of March) with those of our second wave (fielded April through end of June), we saw relatively steady results for networks that cover multiple sports, while those that cater to single sports saw changes that coincided with events happening in those sports.

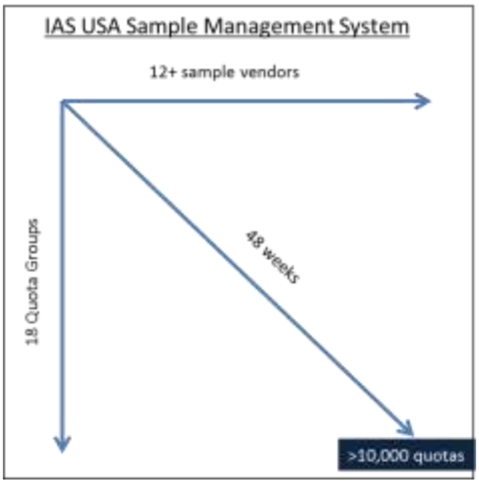
		Comparison of April - June fielding compared to January to March fielding	
	Index		
CBS Sports	89		
ESPN	98		
Fox Sports 1	102		
NBC Sports	104		
MLB Network	229	Opening Day	5-Apr-2015
NBA TV	123	Finals	4-Jun-2015
NFL Network	68	Super Bowl	7-Feb-2015
NHL Network	107	Stanley Cup Playoffs	3-Jun-2015

3. Complexity is okay, overly complicated is not.

Moving to year-round fielding was a critical element of our methodological evolution. We wanted to be able to ensure that the media behavior and purchasing patterns reported were based on real world activities and not skewed by an unbalanced proportion of interviews taking place across the field period. The key to reduce/eliminate this bias through a steady stream of interviews over time was tight control of the quota structure.

To accomplish that, we set up a weekly quota system of 18 cells – gender (2) x age (3) x household income (3). In addition, we established a secondary layer of soft quotas to keep 9 geographic regions, 4 race classifications, and Hispanic/non-Hispanic proportions in balance with the U.S. census. With plans to be in field 48 weeks of the year, we would be establishing 864 hard quotas over the course of the year.

As a measure of affluence in the U.S., our survey requires that respondents have a minimum household income of US $100,000 – an incidence of about 25% among U.S. adults 18+. In order to achieve our sample targets, we needed to engage with more than a dozen sample partners to ensure that we would have enough respondents qualify to complete the survey. This meant that we would be maintaining more than 10,000 quotas over the course of a year.



We learned during our testing phase that the only way to succeed in our mission of evenly distributing interviews over time was to closely monitor and scrutinize fielding. Online sample providers tend to have a fairly simple economic model: 1) the more completes they deliver, the more they get paid; and 2) the faster they deliver those completes, the faster they can move on to the next project. Adherence to our plan required educating each member of the internal sample management team, as well as the online sample providers, about the need for consistency in fieldwork.

After 3 months, we realized that the burden of overseeing this complex quota management system was overtaxing our sample partners, and revised the sample management from weekly quotas to bi-weekly quotas. This system has settled nicely into place and continues. As of the end of August 2015, we had already been in field for 30 weeks with interviews completing within ± 5% of the bi-weekly target.

We continue to be ever-vigilant in monitoring results through our 24/7 Ipsos Cortex® survey system that regulates the opening and closing of quota groups. It also means that we are educating new hires (internal and external), as well as reeducating and retraining existing personnel whenever we see items that give us cause for concern.

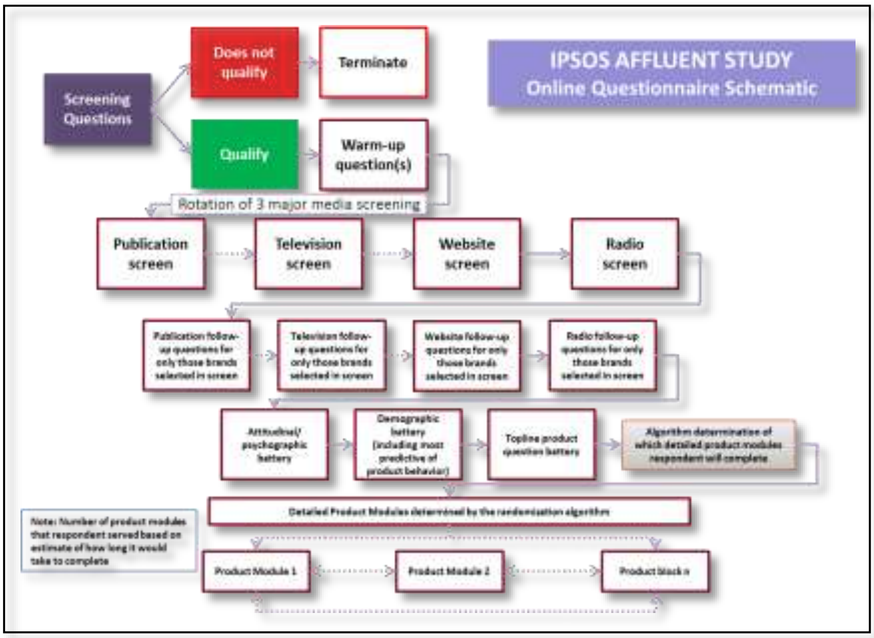
4. Respecting respondents isn't always about shortening the survey.

When we started the process, the Ipsos Affluent Survey USA was longer than we would have liked, and with the addition of cross-platform utilization questions, became longer. To avoid the uneven impact on media behavior due to position effect, we designed the survey with multiple levels of randomization. Traditionally, we had kept the order in which we measured media constant: 1) publications, 2) television networks, 3) radio networks, and lastly, 4) websites. However, during the second phase of testing we compared the fixed order to a randomization of the 3 major media formats (publications, television, and websites). Not surprisingly, the results showed us the need to incorporate a further layer of programming so that the order in which the media screening occurred was randomized and each major media format had an equal chance of appearing first, second, and third.

	Average number of Media Brands Screened into		
	Control Cell (Fixed Order)	TC5 (Random Order)	TC5A (Random Order)
Publications	10.5	9.3	9.4
TV Networks	18.9	19.1	19.3
Websites	16.1	18.2	17.6
Radio (Always fixed in 4th position)	1.8	2.0	1.9

Detailed follow-up questions were not revealed to respondents until after they were screened for all of the media brands.

Two more layers of randomization were incorporated to ensure that within media platform, brands had the same probability of appearing in every position. Finally, after all of the screening took place, the respondent saw the first of the follow-up questions. This avoided the bias associated with response avoidance after learning that a positive response required further information from the respondent.



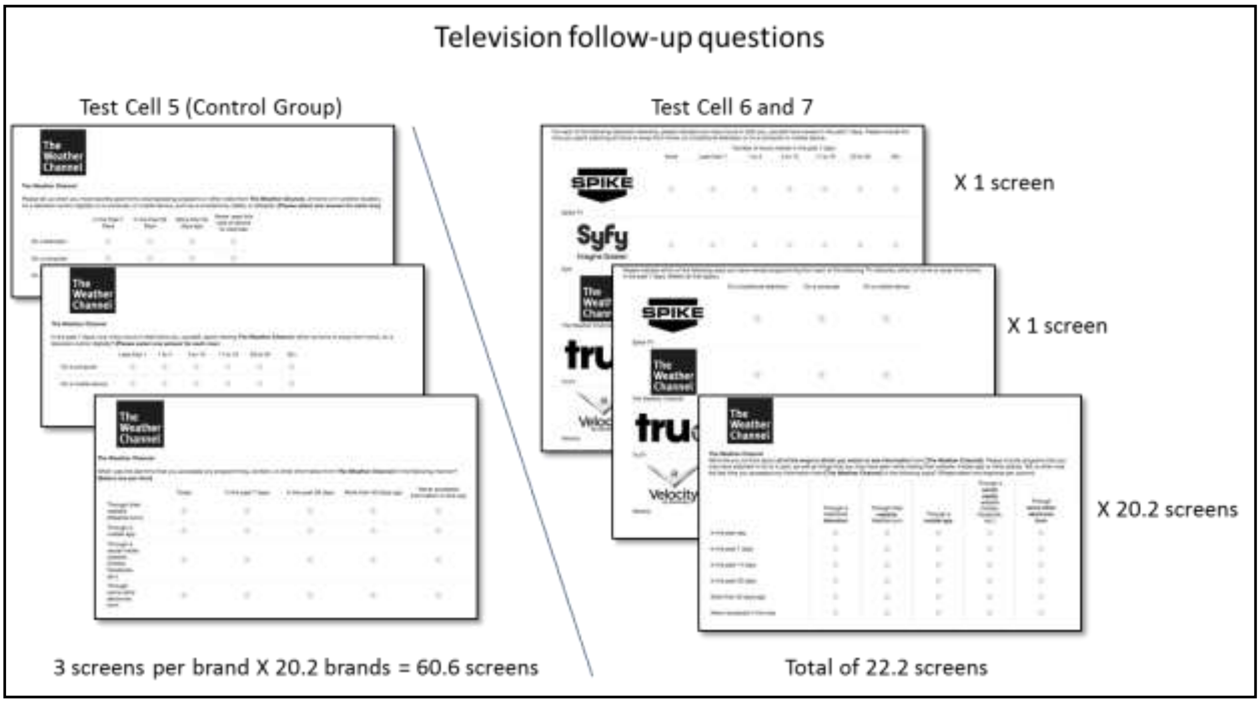
In total, 376 media brands were measured as part of the initial testing:

- 155 publications
- 99 television networks
- 112 non-media affiliated websites
- 10 radio networks

The screening process was designed to collect potential exposure to any form of content from the media brands to allow us to only ask follow-up questions of those who had likely exposure. For every brand that the respondent screened into, they faced 1, 2, or 3 additional computer screens of follow-up questions ascertaining exposure to multiple formats of the brands' content. On average that meant that a respondent had to go through 121 screens just to get through this section.

	# brands screened into	# of follow-up question screens per brand	Total number of screens required
Publications	12.4	2	24.8
Television networks	20.2	3	60.6
Non-media affiliated websites	16.9	2	33.8
Radio Networks	1.8	1	1.8
Total	51.3		121.0

In analyzing "quit reports", (showing where incomplete interviews were being abandoned), we saw that this sequence of follow-up question screens was accounting for a larger than desirable portion of drop outs. Deciding that the best solution was to condense the number of screens respondents would need to go through, we created two test cells where we modified the follow-up questions to incorporate multiple brands at a single time. As a control, we created Test Cell 5 (replicating Test Cells 3 and 4 with some slight modifications that didn't affect media questions). Two versions of the modified follow-up questions (Test Cells 6 and 7) were launched simultaneously with Test Cell 5. The effect of this redesign was to reduce the number of screens a respondent would need to go through by 49%. It seemed like the perfect solution, particularly for the television media questions.



The reaction to the shorter survey length was counterintuitive. Despite a reduction in the number of screens required to complete this sections, the number of respondents in test cells 6 and 7 who abandoned the survey rose by more than 60% compared to the control group. In analyzing the test data, we were able to diagnose the problem as the repetitive sequencing of the cross-platform media question. In the control group, the respondent received 3 different questions about a brand before moving on to the next brand, whereas in the test cells, they received two longer entry questions, before facing more than 20 screens with the same question for different brands. The monotony of this survey sequence caused the large increase in abandonment.

In the end, it appeared that respondents preferred a longer survey task to one that seemed to repeat itself over and over. We ran two more tests to confirm our findings before finalizing the questionnaire prior to full survey launch.

5. Rethink ascription approaches.

As stated earlier, the Ipsos Affluent Survey is quite long. During testing, we determined that it was too difficult for respondents to complete the entire survey in a single interview. We knew that we would need to break the survey into multiple parts, but getting respondents to return and complete an additional survey would mean that we would have a large amount of missing data for respondents who did not complete the follow-up survey. To reduce the impact of this missing data, we instituted a dual strategy that first:

1) Collected top-level product behavior from all respondents during the initial survey
2) Split the remaining product behavior survey questions into 17 modules
3) Developed 62 aggregated blocks of those 17 modules to ensure that we would have data collected on every module individually, as well as overlapping with every other module
4) Programmed the survey so that respondents received between 2 and 5 of those modules when they completed the initial survey
5) Recontacted respondents to the initial survey to complete as many of the remaining modules as possible within time constraints

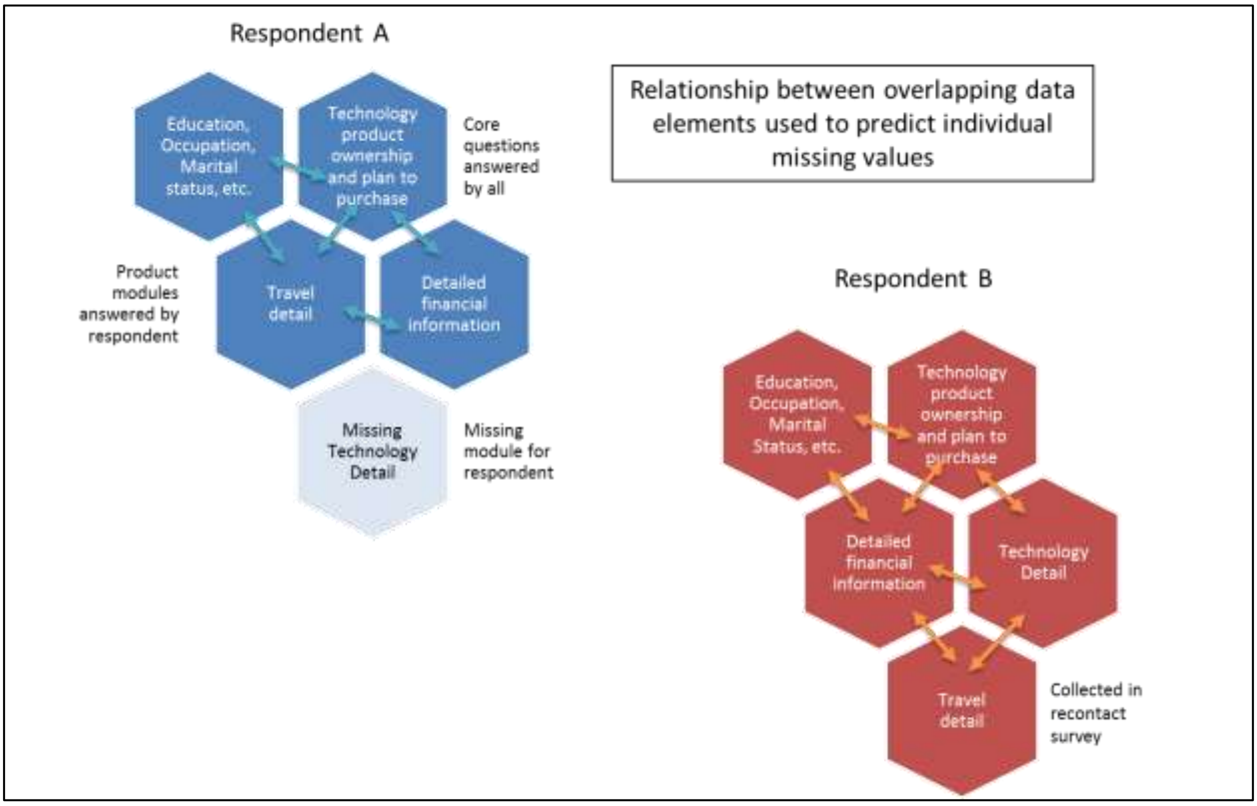
In order to make the modeling possible, it required that we have data on all of the overlapping relationships. The 62 module blocks were designed to ensure that we not only collected product data relating to specific top level questions, but we would also know the relationship between different product categories. This would allow us to more accurately predict missing values for a respondent using the probabilities developed from overlapping behaviors.

Because of the varying length of time it took respondents to complete the modules, we utilized the design below during the initial survey, randomly assigning respondents to complete one of the 62 variations.

Modules	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1		M35	M52	M2	M5	M52 M53	M7	M9	M17	M53	M49	M49	M34	M52	M19	M49	M53
2	M35		M51	M4	M6	M51	M8	M10	M18	M51	M54	M54	M37	M54	M42	M37	M42
3	M52	M51		M41	M50	M51 M52 M55 M57 M58	M55	M48	M56	M43 M48 M51	M43	M55 M56	M58	M50 M52 M57	M57	M41 M48 M50	M43 M56 M58
4	M2	M4	M41		M40	M46	M11	M46	M20	M1	M3	M38	M38	M46	M22	M41	M40
5	M5	M6	M50	M40		M39	M12	M45	M21	M13	M14	M45	M39	M50	M23	M50	M40 M45
6	M52 M53	M51	M51 M52 M55 M57 M58	M46	M39		M55	M46 M47 M61	M61	M51 M53	M47	M55	M39 M58	M46 M52 M57 M61	M57	M47 M61	M53 M58
7	M7	M8	M55	M11	M12	M55		M62	M28	M15	M16	M55	M31	M62	M28	M62	M62
8	M9	M10	M48	M46	M45	M46 M47 M61	M62		M61	M48	M47	M45 M60	M32	M46 M61 M62	M60	M47 M48 M61 M62	M45 M60 M62
9	M17	M18	M56	M20	M21	M61	M28	M61		M26	M27	M56	M33	M61	M30	M61	MM56
10	M53	M51	M43 M48 M51	M1	M13	M51 M53	M15	M48	M26		M43	M44	M44	M44	M24	M48	M43 M53
11	M49	M54	M43	M3	M14	M47	M16	M47	M27	M43		M49 M54 M59	M59	M54 M59	M25	M47 M49	M43
12	M49	M54	M55 M56	M38	M45	M55	M55	M45 M60	M56	M44	M49 M54 M59		M38 M44 M59	M44 M54 M59	M60	M49	M45 M56 M60
13	M34	M37	M58	M38	M39	M39 M58	M31	M32	M33	M44	M59	M38 M44 M59		M44 M59	M36	M36 M37	M58
14	M52	M54	M50 M52 M57	M46	M50	M46 M52 M57 M61	M62	M46 M61 M62	M61	M44	M54 M59	M44 M54 M59	M44 M59		M57	M50 M61 M62	M62
15	M19	M42	M57	M22	M23	M57	M29	M60	M30	M24	M25	M60	M36	M57		M36	M42 M60
16	M49	M37	M41 M48 M50	M41	M50	M47 M61	M62	M47 M48 M61 M62	M61	M48	M47 M49	M49	M36 M37	M50 M61 M62	M36		M62
17	M53	M42	M43 M56 M58	M40	M40 M45	M53 M58	M62	M45 M60 M62	M56	M43 M53	M43	M45 M56 M60	M58	M62	M42 M60	M62	

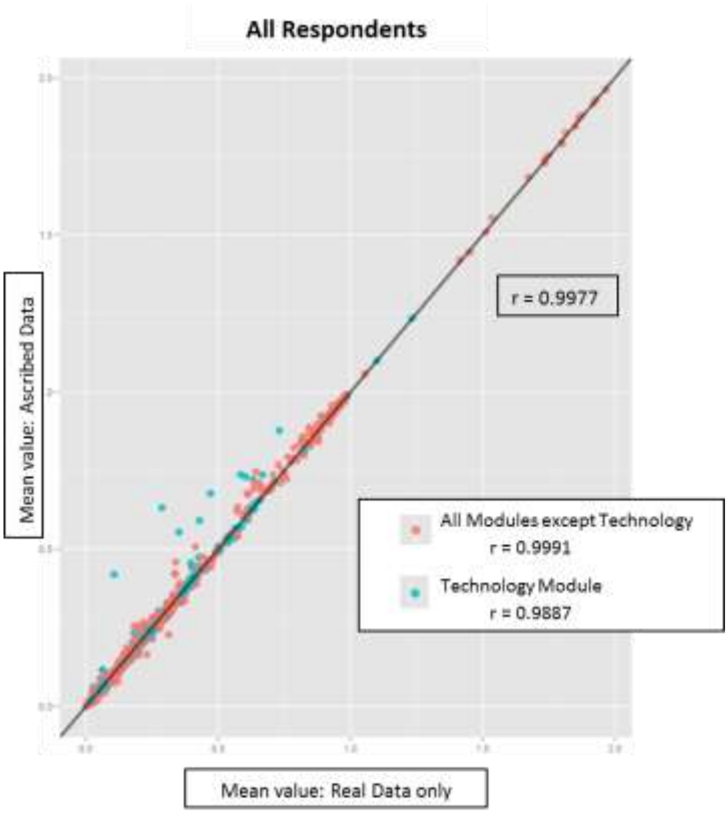
The ascription algorithm is a modified hot-deck imputation, utilizing 20 demographic variables, as well as specific product variables for each of the product modules being ascribed. The matching algorithm uses these attributes to pair donor and recipients, while the value prediction algorithm uses an ensemble of trees to predict the missing values on a question-by-question basis. Given the large number of columns in the Affluent data, the donor-recipient matching method is preferred for most of the modules. The value prediction method is only used where it would be impossible to find matches that fit the survey logic in the matching method.

In the simplified example below, we were missing the detailed technology product information for Respondent A, but knew the relationship for a host of other factors. Utilizing data from Respondent B, where we had the detailed technology brand information and usage along with the other elements, allowed us to develop individual propensities to predict Respondent A's technology preferences.



It should be noted that this type of ascription modelling requires a substantial amount of computing time due to the complex pattern of responses and the sheer number of variables. In order to complete the process on a little over 15 thousand records, the program ran continuously for 194 hours.

Once complete, we check to see the agreement between the resulting full data set (with ascription) and the partial data set containing only real respondent answers. The chart below shows the level of agreement for the 2015 survey results (r = .998).



Conclusions and next steps –
We feel that we have successfully completed our migration from an outdated methodology to one that is responsive to the needs of the marketplace. As an online survey, we have transcended the limitations of our printed questionnaires. We have developed a sample frame that allows us to reach respondents across all demographics in the affluent markets without utilizing unnecessarily high projection weights. As a result, we are able to project to a population that is more in line with U.S. census estimates using the online methodology than we could with the mailed questionnaire.

With the new methodology we are able to provide the industry with single-source cross-platform media data for all of the media brands measured (including the use of mobile apps, social media, and other electronic forms of connection). We have coupled those measures with further insight into how affluents use specific media formats. The study has evolved from a somewhat limited view of media behavior to a truly multidimensional one that allows the industry to target consumers in a way that mimics how they actually consume media today.

We incorporated virtually unlimited rotations/randomizations to reduce/eliminate positional biases in the survey, and provided visuals to aid respondent recall. Our replicability testing has shown that this new methodology will provide wave-to-wave results that will be reasonably consistent, and with the carefully monitored field procedures, will reflect impacts that are occurring in the marketplace.

We continue to explore ways of adjusting our survey so that we can collect the desired information without overburdening respondents. We already fuse our data with two other syndicated sources to provide users with greater detail and search for other alliances that would potentially allow us to integrate passively collected data to replace recall-based questions. Through survey software improvements, we look to simplify and visually improve the questionnaire.

As we continue to fine-tune our methodology, our marketing scientists will continue to work on increasing the precision of our predictive modeling. While the goal of having no missing data is some time off, our predictive modeling imputation gives us comfort that we have improved the accuracy of the imputation process.

It was a major step for us to abandon 38 years of methodology, but we feel the rewards of doing so have enabled us to meet the challenges of providing clients with the information and insight they need to market to affluents in the U.S. for the foreseeable future.