Optimum Blending of Panels and Social Network Respondents

Steven Gittelman Ph.D.

Mktg, Inc.

Adam Portner

Research Now

The Opportunity

Social network growth statistics are staggering. What once was a phenomenon regaled only by the young now has broad demographic reach (Figure 1). Facebook has trampled age barriers working its way into almost every demographic corner.

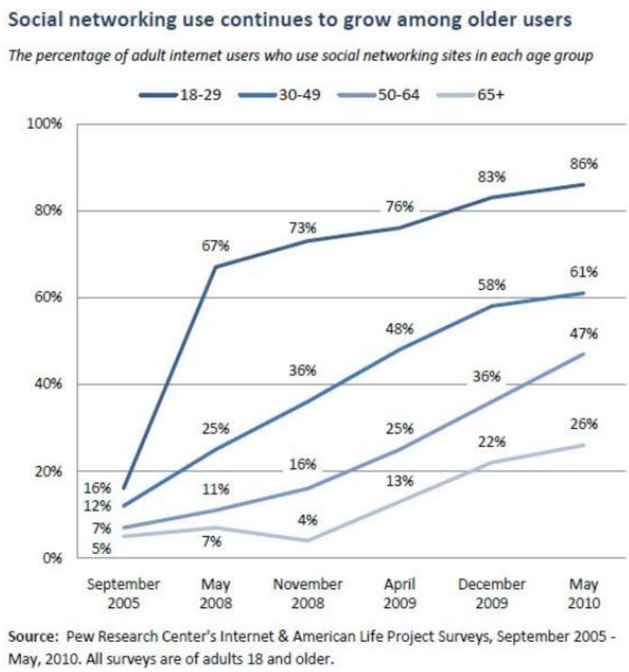

Figure 1 – Growth of Social Network by Age Category

Facebook alone has a half billion pre-profiled respondents in comparison to seven million panelized double opted-in respondents who constitute the core of the online research, (ARF 2009). This disparity highlights the critical shortage of respondents that exist in our online panels. Thus, where opportunities afford themselves, we must have methods that are tested for their inclusion.

The online panels appropriately seek to avoid overuse of their respondent base. The inclusion of social network respondents should relieve this pressure. In addition it allows the Market Research industry to involve people in research who might not participate in online panels. The result is a more comprehensive and inclusive sample frame.

In this paper we seek to determine the degree to which a social network population sourced from Peanut Labs respondents could be blended with an established panel, Research Now's American Valued Opinions Panel (VOP) while maintaining the original panel sample characteristics.

The Challenge: A Carpenter's Tale

There is an old adage in carpentry, "measure twice cut once." What appears as a simple axiom is a robust statement of the entire sphere of quality standards from ISO to Six Sigma. Our carpenter friend holds the key to quality: good measurement tools, precision, fit for purpose, metrics and record keeping. If he is sloppy at the tape measure and cuts prematurely, his craft will suffer, gaps will appear due to his lack of precision, a roof might fail to hold a snow load or in the case of a cabinet maker, the work will be 'shoddy' and loosen at the seams.

When we blend samples, we must rise to the standards of the fine craftsman. Clearly, the challenge behind combining sample sources is one of proper metrics, precision in measurement, properly crafted tools and an overriding sense of the purpose to which our samples will be employed.

Standards: Minimum Measurable Difference

In quantitative research we speak in terms of statistically significant difference when comparing populations. There is a threshold at which the difference is so slight that statistics fail to discriminate difference and we presume the two populations to be similar. We coin here the phrase, "minimum measurable difference" to be the smallest difference between two populations that we can discriminate through statistics.

More normally, we would express populations to be different by establishing an α (alpha) value associated with the precision, or likelihood that two samples are different. Thus, we might declare two populations different at an α value of ≤ 0.05. In situations where our measurements are less precise, we might settle for a shaky $\alpha \leq 0.1$. Often we make such compromises when compelled to work from samples that are either small or variable.

The *minimum measurable difference* is a means of determining the threshold at which we begin to detect statistical difference at an α value level so low that it represents a conservative measure of similarity. Anything below the minimum would be considered to be an undetectable difference that lends credence to the statement that in "as much as we fail to detect difference, we can declare the two populations similar in the metric that we are evaluating."

Here we choose to set our α value levels at one standard error for a sample size of 1500. Examination of sample sizes among online research studies conducted by Mktg, Inc. showed that less than 5% of studies we have performed employ samples of more than 1500. This is a conservative standard.

Metrics

Our days are ruled by measurement. Intuitively, we understand metrics to gauge temperature, humidity, pressure, automotive velocity, our blood cholesterol, the calorie content of our food and so on. The science behind measurement is at times so exact, that it is no wonder that statistical significance in the hard sciences often begins at $\alpha \leq 0.01$ and beyond.

We measure human behaviors, online behaviors. Be it, buying behavior, media preference and socio-graphics, they are quite variable. Establishing metrics that reflect populations and their behaviors is a difficult task. Speak to the next Six Sigma Black Belt with whom you become entangled and you will be given a hard lesson in the need for measurement and associated standards. Difficult or not, we need metrics to establish and maintain Quality.

Standards

The word *representative* drives fear into the hearts of many members of the market research profession. In fact, we wither at any question of what our sample might represent. In a phrase, "We really don't know exactly."

Here we reject demography as the only suitable **stand-alone** standard for online market research samples. When we attempt to calibrate behavior by demography alone, we assume that a proper distribution of demography assures us of a reliable sampling of behavior. We have found that highly nested demographic samples of different online sources yield significant and meaningful behavioral differences between populations (Gittelman and Trimarchi, Esomar 2010).

Our standards have to relate to the measures that we seek to represent. The need for these standards is at the crux of blending online samples. We must blend to a relevant target. In market research we measure behavior. Often the purchasing patterns, buying behavior and other predilections of our target audience are the most germane subjects of our interest. Thus, in creating our metrics, we employ segmentations based on buying behavior, purchasing intent, media preference and socio-graphic behaviors. The metrics we use at Mktg, Inc. are the result of highly refined segmentations, collected in thirty five countries and tested over a four year span with over two hundred online panels.

We seek to determine the degree to which social network respondents emanating from Peanut Labs can be blended with panel respondents who belong to Research Now's American Valued Opinions Panel (VOP). Our analysis included 4,009 US VOP respondents (9/14/2010 – 11/1/2010) and 3,871 US Peanut Labs respondents (9/14/2010 – 12/19/2010), using identically nested Sex x Age x Income.

The distribution of behavior segments represented by highly balanced samples of VOP, acts as our standard. We might use other standards but here we seek to sustain consistency of the VOP sample as we add respondents from Peanut Labs. We use an iterative model to determine how many social network respondents, originating from Peanut Labs, can be added before we detect a *minimum measurable difference* in the blended combination. Our purpose here is to achieve a consistent blend to eliminate changes in survey data that might otherwise be created by changes in the underlying sample frame.

Understand the Differences

Social Network respondents are different from those now drawn from double opt-in panelized respondents. These differences are inherent in their reasons for being online. Those who are using the social network arena to communicate with others, obtain news, or be entertained are likely to be different from others who are participating only for online purchases, doing their banking, or searching for the best deal available for an airline ticket. Those who embrace the potential of the Internet for social interaction are systematically different from those who see it merely as a means to expediting their offline lives. As many of our online panels are sourced from a combination of commerce sites such as frequent flier, various reward programs or simply special interest groups they are likely to be different from someone seeking social contacts or the latest viral treat on YouTube.

At first, we are challenged to understand the differences and to establish methods for blending this new wave of respondents into existing panels, while maintaining consistent results. Users of these panels must be assured that the addition of any new source, including Social Networks, will not introduce instability to the samples and increase the variability in their data.

While we attempt to control for differences in our respondents via demographic quotas, it's clear~~,~~ that individuals from Social Networks are considerably different. When examining education (Figure 2), among social networked individuals, with identical Sex x Age x Income distributions, we find a far less educated population than the ones derived from a typical online panel. But while these differences are suggestive of underlying problems, demographics do not tell the whole tale.

Social network respondents are different from panel respondents and the degree of that difference dictates the number of social network respondents that can be added to an existing sample frame without changing the behavior represented by the original panel. The issue is further complicated as we drill into different demographic groups. We find that the differences between groups are not consistent. As a rule, older respondents are more different than are younger ones.

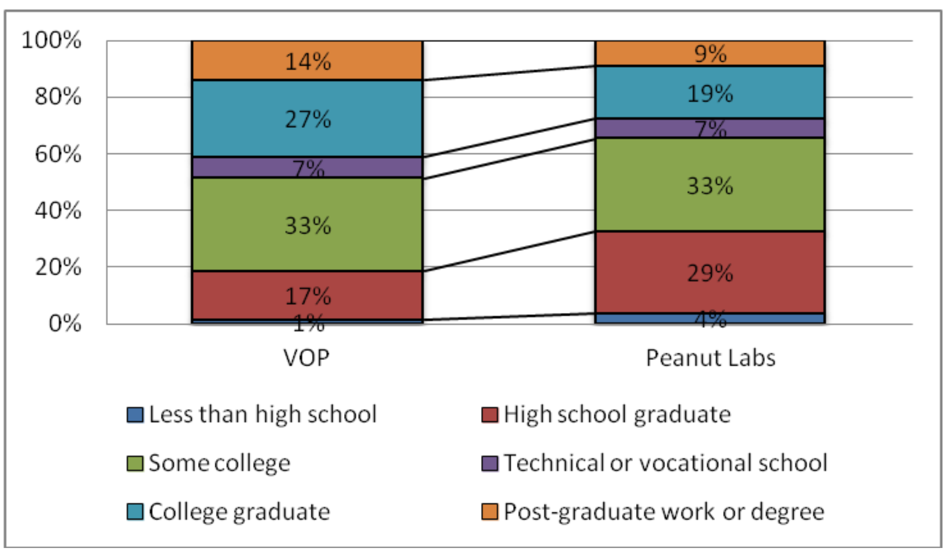

Figure 2: Education Distribution of the Samples

1 Structural Segments

Identifying Structural Segments

Individual consumers have different motivations and habits, with different factors influencing their adoption or purchase of a particular product or service. With demography insufficient to independently represent each of these consumer 'segments', it is important to form a typology through which they might be identified to ensure a behaviorally consistent sampling frame. The process of identifying structural segments can be thought of as having four steps, going from the selection of the variables through identifying segments and developing a regression model.

Select Variables ⇒ Cluster Analysis ⇒ Logit Regression Model ⇒ Test Results

This task is done with a substantial set of data, within a single country in order to provide a stable structure. The parameter estimates from the resulting regression model are then used to assign segments for all other datasets, creating an internally consistent set of distinct respondent groups. The requirements for an acceptable structural segmentation scheme are formidable in that the resulting scheme must consist of highly distinct groups whose differences are reliable across samples. The resulting model must provide clear assignments of respondents to segments, which may require several iterations of this process until the ideal group of variables is identified.

Respondents completed a 17 minute standard questionnaire covering media, technology usage, lifestyle and purchasing intent. These questions were utilized to create a standard battery of three structural segments: (1) Buying Behavior, describing generalized purchasing behavior and involving 37 questions, (2) Socio-graphic behavior, describing lifestyle choices, with 31 questions and (3) Media usage, describing general modes of media consumption, with 31 questions. Each respondent was assigned to one segment in each of the three segmentation schemes, with each scheme consisting of three or four segments. For example, the average young male would be classified as a "purchaser" in the Buyer Behavior segmentation, as being "Social Networked" in the Socio-graphic segmentation, and as a regular "Internet" user in the Media segmentation. The composition of each segment is displayed in the following sections.

1.1 Buyer Behavior Segments

The Buyer Behavior segments capture major differences in respondent purchasing behavior. Figure 3 shows the standardized profile of the segments based on the questions included in the executed questionnaire. These cover both frequency of use and of purchases and attitudes. These profiles show the degree of impact the variables have in determining a respondents' behavioral classification. Deviations from zero indicate the impact on the respective segment in either a positive or negative direction.

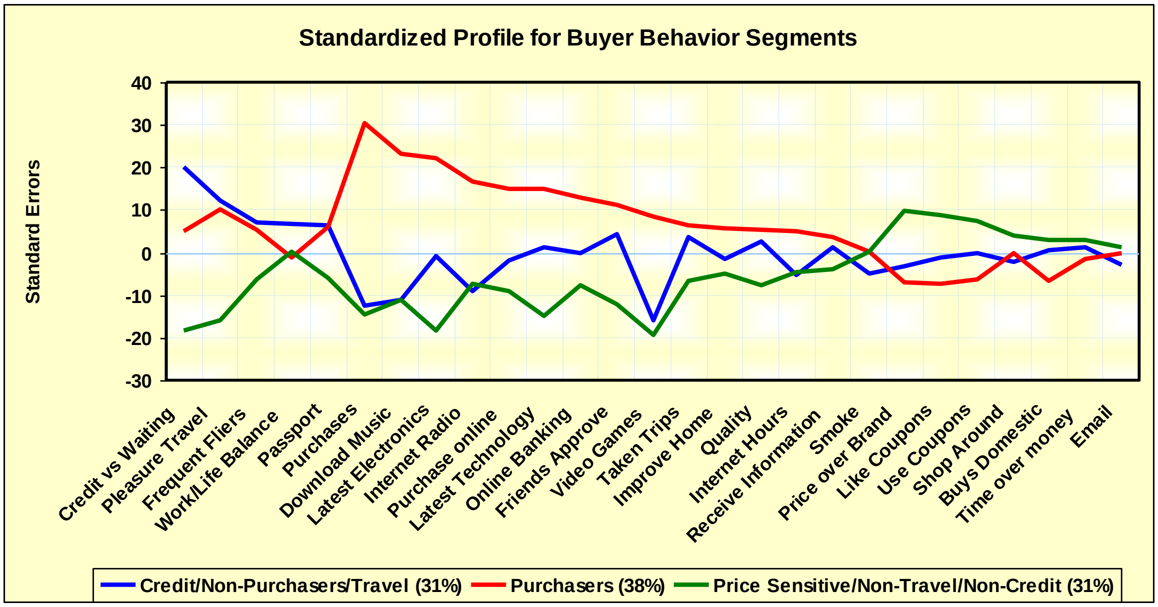

Figure 3, Standardized Profile of the Buyer Behavior Segments

Figure 4 shows the distribution of these segments between the host, VOP, and the alternative source, Peanut Labs. Note that these are significantly different overall. Differences, of course, would be expected to vary in sub-groups of respondents within these sources.

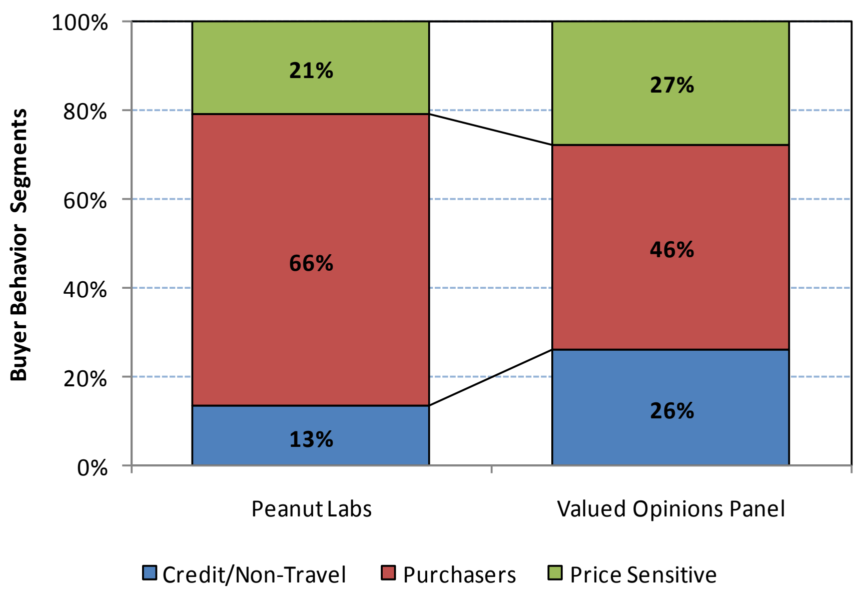

Figure 4, The Buyer Behavior Segment Distributions

1.2 Socio-graphic Segments

The socio-graphic segments capture population differences in attitudes, behavior and to some extent, lifestyle. Figure 5 shows the importance of the various questions used to formulate this segmentation scheme. As in the last section, these are standardized profiles.

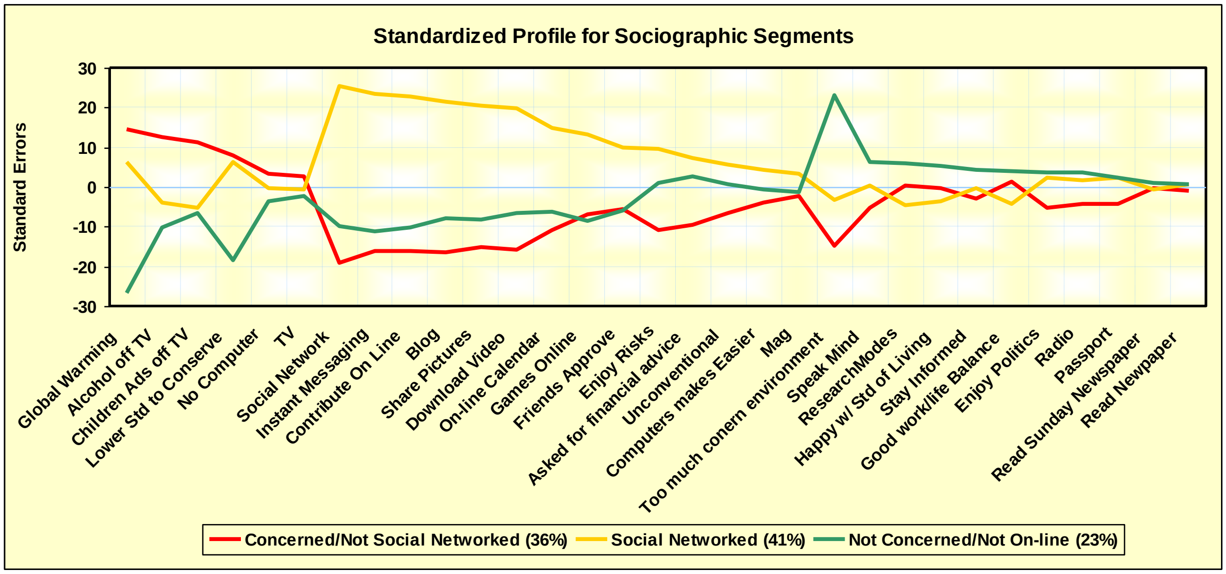

Figure 5, Standardized Profile for the Socio-graphic Segments

Figure 6 shows the distribution of socio-graphic segments between the host and the alternative source. Once again there are very large differences.

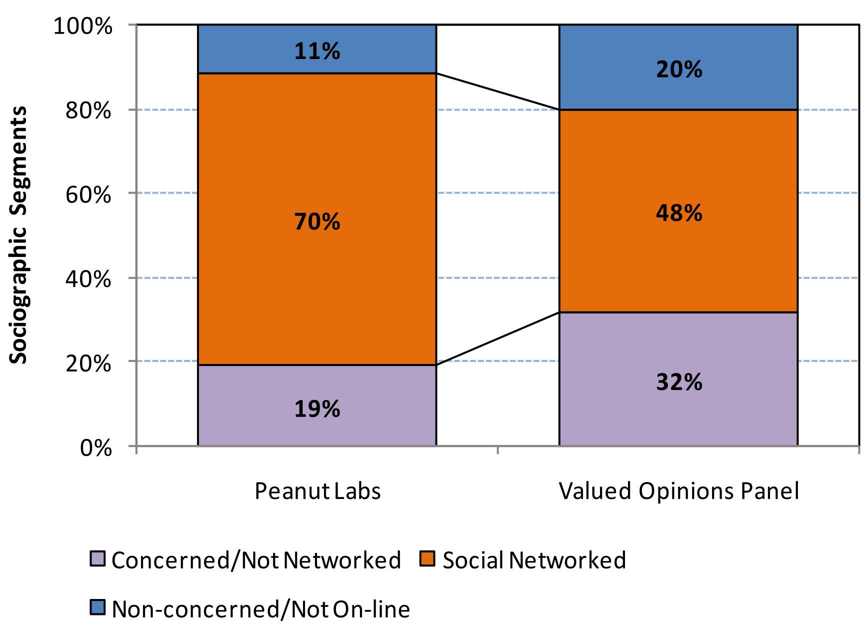

Figure 6, The Socio-graphic Segment Distribution

1.3 Media Usage Segments

The Media Usage Segments capture the sources of information and use of media by respondents. As before, the following figure, figure 7, shows the relative importance of the various responses from the questionnaire to forming the segments. It should be noted that the segment on the Internet usage, is expected to be dependent on the sources of respondents.

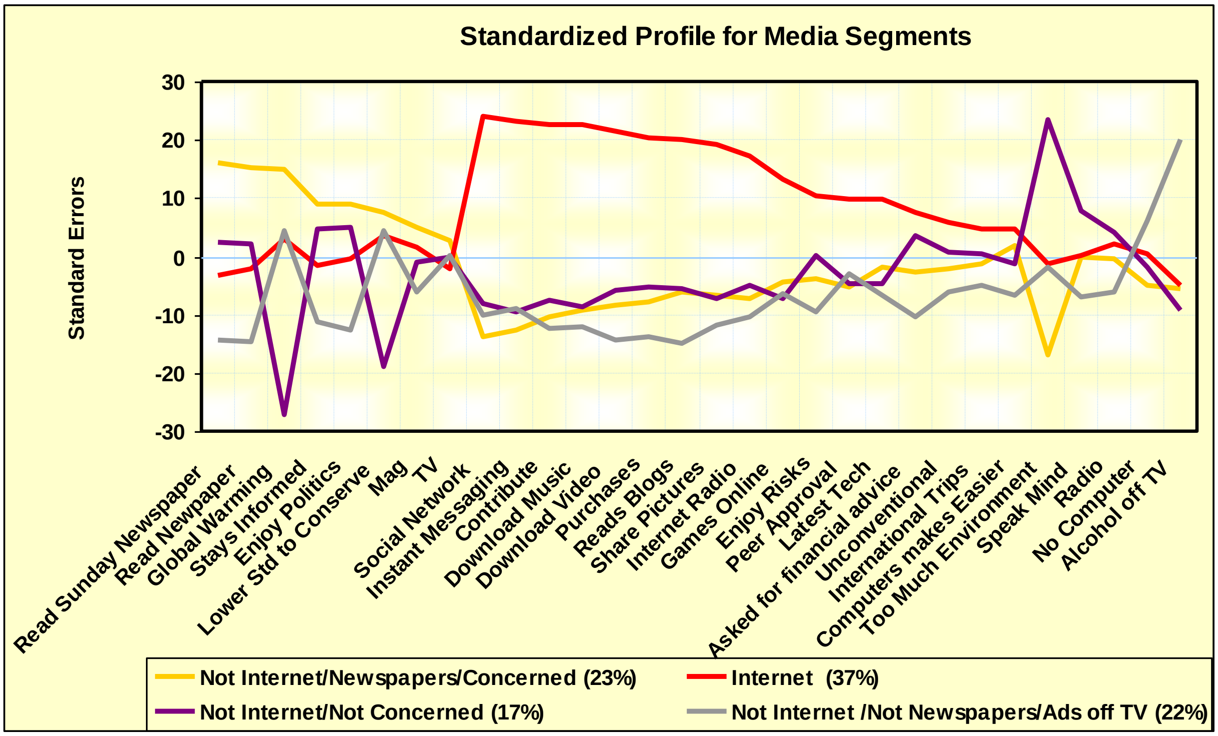

Figure 7, Standardized Profiles of the Media Usage Segments

Figure 8 shows the distribution of media usage segments between the two sources. As it would be expected, this indicates major differences.

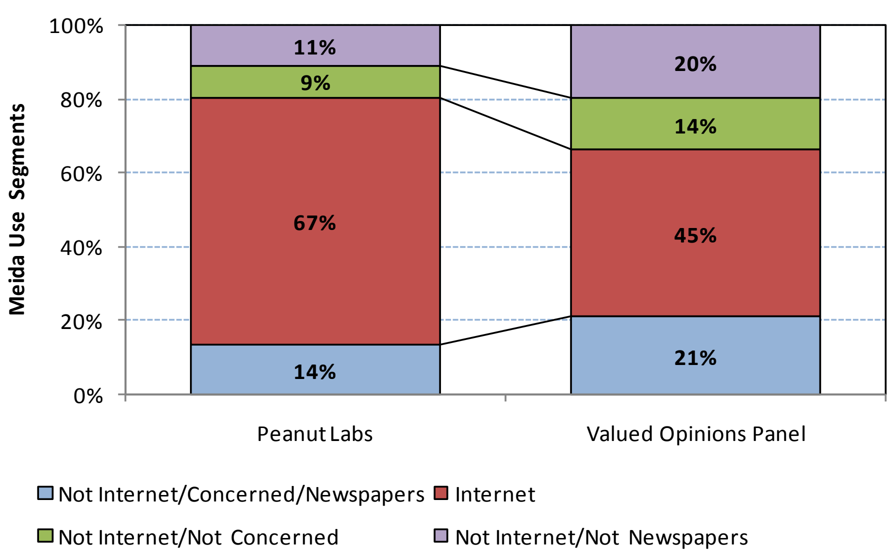

Figure 8, Media Use Segments Distributions

2 Maximum Blend Ratio

The behavioral differences between the VOP and Peanut Labs samples are significant. As a result, we suggest these sources are not directly substitutable for one another. When consistency of data is critical (wave studies, pre/post, tracking studies), uncontrolled introduction of Peanut Labs respondents into a Valued Opinions Panel sample could be problematic. Such a mixture may create considerable change in the characteristics of the original source. The practical question of blending, therefore, becomes one not of finding those source respondents who will exactly replicate the panel respondents. Instead, it becomes one of finding the correct amount of respondents who can be added without bringing about significantly different survey results.

While such a blending model could be developed for the sample as a whole, deviations between sources would likely exist within demographic cells. As such, a demographics based blending is called for. A demographic matrix (by age and gender) was used. The question was, for each cell in the matrix, what fraction of the source could be added to the host sample without materially altering the resulting characteristics.

2.1 Measures

There are two measurement issues; first is how to measure differences between the two panels and second what is the largest acceptable difference. Since this is a simple (linear) mixture, the acceptable maximum ratio would be equal to the largest acceptable difference divided by measured difference between the data sources to be blended. The measured difference is taken as the "Root Mean Squared Difference". That is, the square root of the average of the squared differences of the segments. For the Buyer Behavior, which has three segments, this becomes:

$$Distance = \sqrt{\sum_{1}^{3} \frac{[Segment(Host)_i - Segment(Source)_i]^2}{3}}$$

The Media Usage scheme has four segments and that increases the number of items in the average. Note that these measures are computed for each of the segmentation schemes.

The acceptable distance is related to the expected error around the distribution of segments. This is taken as a Root Mean Squared Standard Error. The standard error around each segment is given by the binomial formula:

$$Standard\ Error_i = \sqrt{P_i \times \frac{1 - P_i}{N}}$$

P_i is the fraction of the sample in $Segment_i$ of the Host and *N* is the number of respondents in the targeted sample. Note that the number of respondents in the targeted sample is not necessarily the size of the sample used in the measurement. It represents the size of studies for which the test is being run. The total measure of error is the Root Mean Square of these standard errors:

$$Total\ Standard\ Error = \sum_{1}^{3} \left(\frac{[Standard\ Error]_i^2}{3}\right)$$

Finally the acceptable level is taken as some proportion, "β", of the total standard error. We can look at this as a "Type I error", that is we seek the minimum acceptable likelihood that the two samples are the same. This is referred to as the "α" term. In typical statistical comparisons, an α term of 5% is generally implied, meaning the chances are less than 5% that the two samples are the same. This is a conservative threshold, chosen by scientists to minimize the chances that a given treatment is falsely said to have an effect. However, our intention is the opposite. We wish to establish at what levels our host and blended sample are *not* statistically different, and thus, a higher α is more conservative and appropriate. As such, we set our threshold at one standard error as the acceptable range which is equal to approximately α = 32%, rather than the usual two standard errors. This gives us two adjustable parameters in selecting a policy, the targeted sample size and the minimum acceptable likelihood.

Therefore the acceptable level is:

$$Acceptable\ Level = \beta \times\ Total\ Standard\ Deviation$$

And the Maximum Blend Ratio:

$$Minimum\ Blend\ Ratio = Acceptable\ Level/Distance$$

As mentioned earlier, this is done for each of the three segmentation schemes. The overall Maximum Blend Ratio is taken as the lowest of these. This is done for each of the demographic groups.

2.2 The Effect of Target Sample Size and Acceptable Likelihood

A Total Maximum Blend Ratio is computed based on the weighted sum of the individual demographic cells. Figure 9 shows the distribution of these total Maximum Blend Ratios as a function of α and N.

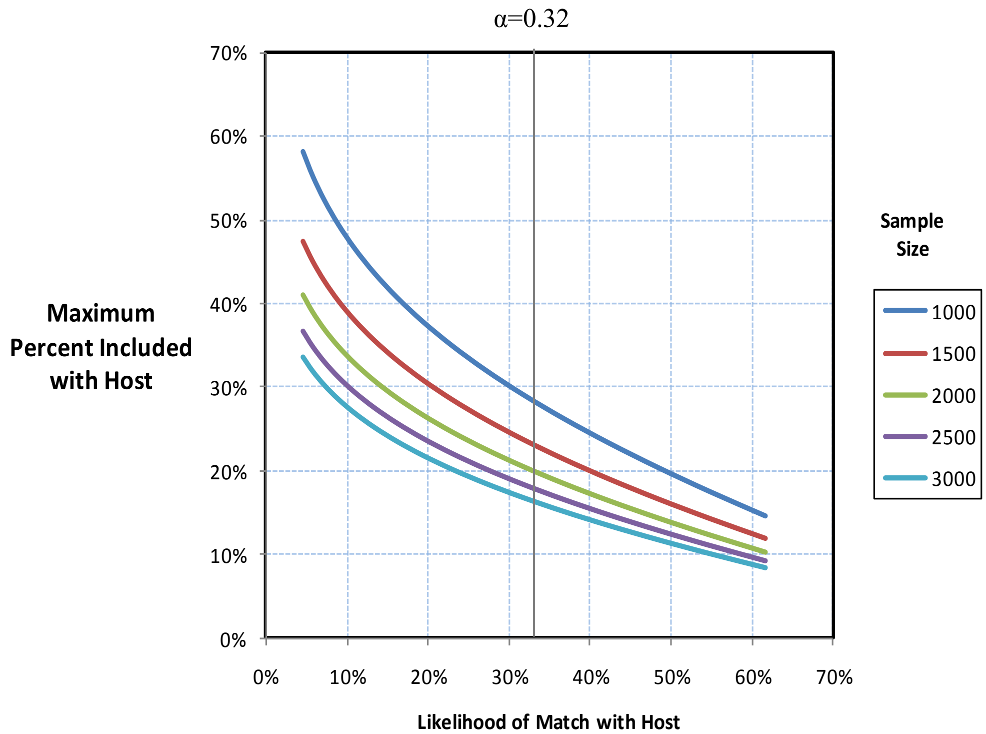

Figure 9, Overall Maximum Blend Ratio as a Function of Acceptable Likelihood & Size based on Average Values

Notice that this ratio decreases as α and N increase. As the tolerance, indicated by these factors, decreases, with increasing values of these parameters, the quantity of respondents that can be blended decreases.

We have chosen the targeted sample size to be **1500** and used an α = **32%** or one standard error. This corresponds to what we believe to be reasonable conditions for a typical mixed source application. In the case of US VOP and Peanut Labs, this allows an average maximum blending ratio of **23%** covering all demographic cells, though in reality the specific percentage will differ between the cells. Increasing the tolerance would result in a larger maximum blending ratio, as well as the reverse, should more conservative estimates be desired.

2.3 Variation within Demographic Cell

Figure 10 shows the distribution of Maximum Blend Ratios across the demographic cells using averaged values. It ranges from 14% for the Female 55+ to a high of 43% for the Male 18-24.

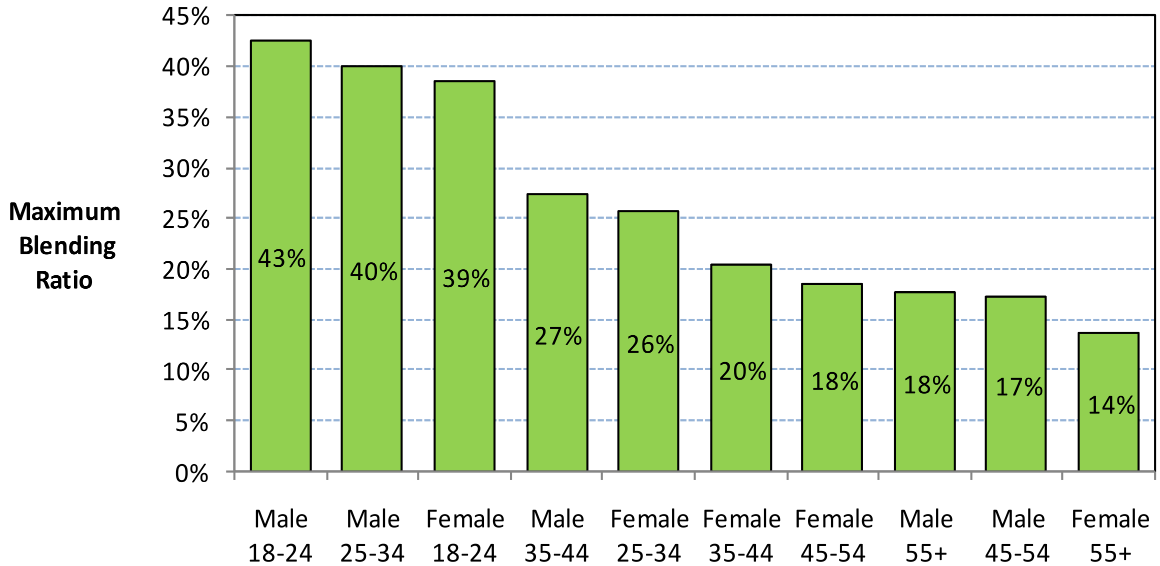

Figure 10, Distribution of Maximum Blend Ratio across Demographic cells for Average Values

3 Final Blending Model and Maximum Effect

Figures 14-16 show the effect of the blending process based on the three main segmentions. It is expected that there should not be any major differences between Valued Opinions Panel, and the blend, even though 18.8% of the blend is respondents from Peanut Labs. Figure 14 shows the results for the Buyer Behavior Segments. While there are differences between the host and the blend, they are relatively minor.

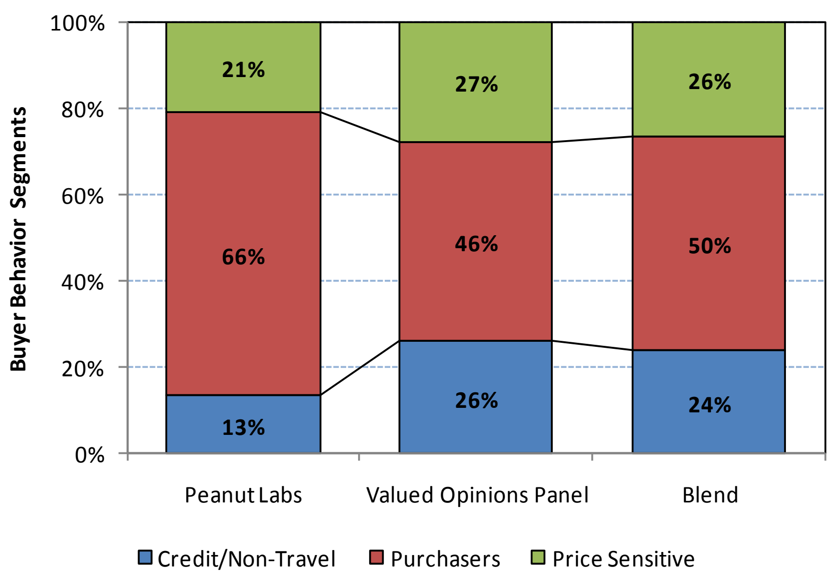

Figure 14, Blend Buyer Behavior Segment Distribution

Figure 15 shows similar results for the socio-graphic segments. Clearly, as previously noted, Valued Opinions Panel and Peanut Labs are very different. But the blend is very close to that of the original panel. The Blend and Peanut Labs sample sets were significantly different at p<.01.

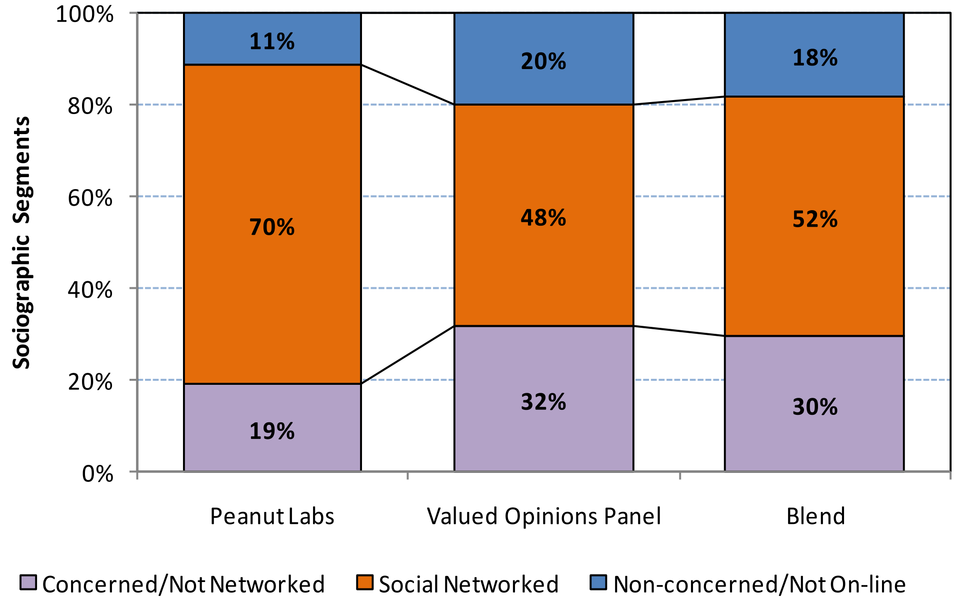

Figure 15, Blend Socio-graphic Segment Distribution

Figure 16 shows the results for the media usage segments with the same conclusions. The difference between the blend and the host are minor compared to that against the total for Peanut Labs. It is this similarity of characteristics that allows for the blend to be used as an extension of the original panel without major concern regarding consistency. The Blend and Peanut Labs samples continued to be significantly different.

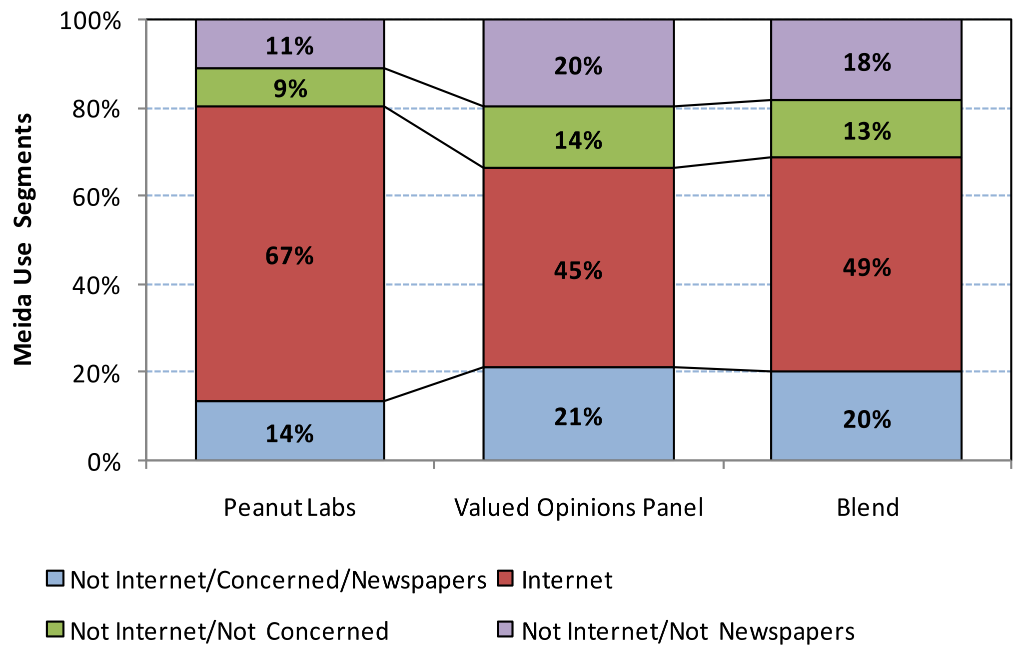

Figure 16, Blend Media Use Segment Distributions

2.4 The Effect of Blending on Survey-Taking Behavior

The blending procedure was designed to ensure that the structural segments of the blended sample remain statistically similar to the original panel when controlling for demography. However, the introduction of blended sample may result in differences with regard to survey-taking characteristics of the sample such as panel tenure, survey-taking hyperactivity, and quality metrics. The changes that one can expect are detailed in Figures 17-19.

Figure 17 shows the results for "performance", a measure of respondents' susceptibility to 'trap' questions, through which respondents' engagement in the survey is tested. Three such questions were used; first, an instructional question where respondents were asked to enter a certain value. Those who entered an incorrect value received a mark for 'failure to follow instructions'. Two other questions asked logically identical but oppositely worded questions regarding their quality of life and their preference for brand over price. An attentive respondent should give opposite answers to these questions, and those who did not were coded as being 'inconsistent'. As shown by the RMSE (Root Mean Squared Error) statistic, the blended sample was not significantly different from VOP in any measure of performance, but was significantly different from the source sample set in all cases except "standard of living".

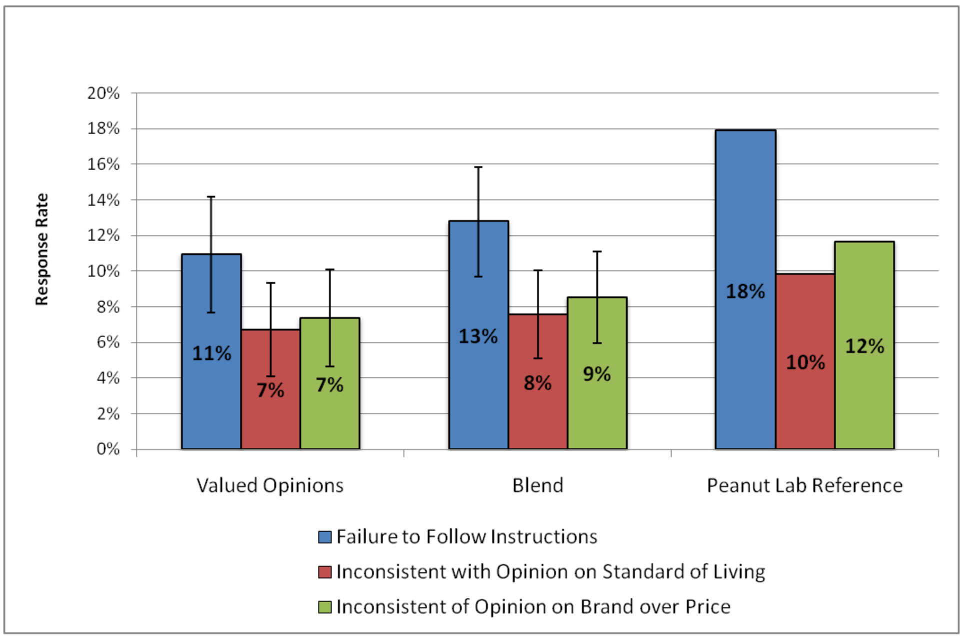

Figure 17 Measures of Performance

There is evidence that changes in panel members' tenure can cause shifts in data. In Figure 18, the comparison between the aging of panel participation distributions for Blend, VOP and the Peanut Labs Reference is shown.

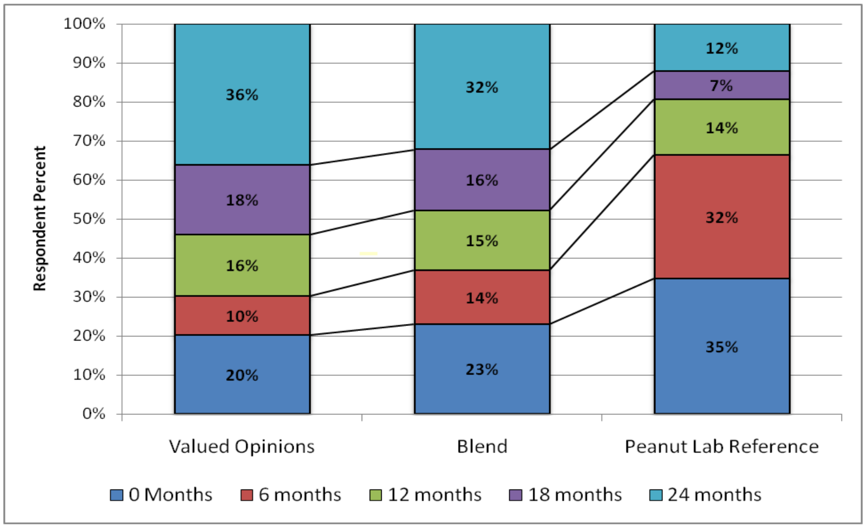

Figure 18 Distribution of Panel Tenure

The performance characteristics that were covered previously focused on the errors made by respondents and their tenure on panels. There is a third category of activities that are thought to possibly affect the quality of results. These are the participants who either speed through the survey (speeders) and those who give similar or identical values to blocks of questions in the surveys (straight-liners). These respondents can be viewed as potential satisficers. In Figure 19, the distribution of satisficing behavior is shown. Based on the RMSE, the number of straight-liners and speeders in the blended sample were not significantly different from VOP or Peanut Labs.

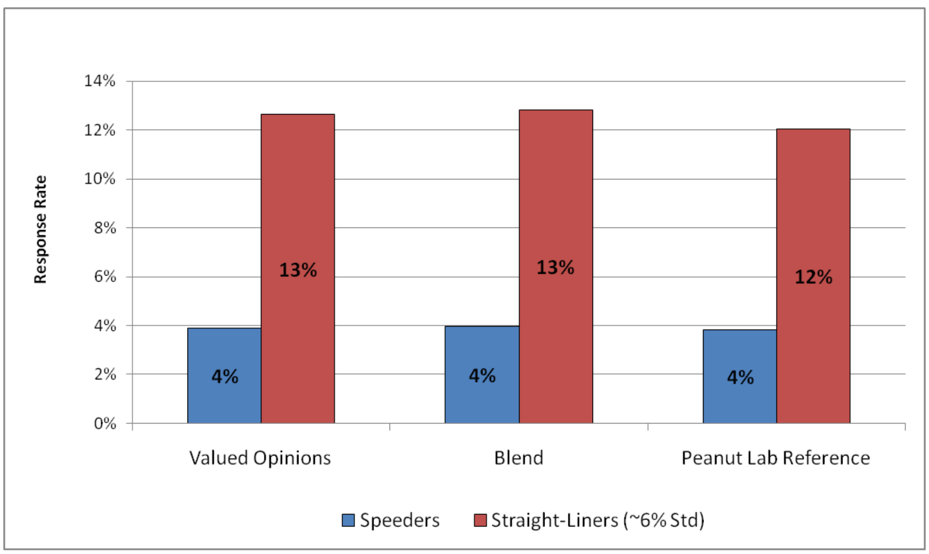

Figure 19 Distribution of Satisficing Behavior

Conclusion

Here we introduce the concept of a *minimum measurable difference*. It serves as the minimum change in our metrics where we conclude that samples are detectably different: At any lesser change the populations are considered the same. This contrasts with the standard statistical interpretation where we simply determine that two populations differ, without a measure at which point that difference was achieved.

Social media participants represent a large potential opportunity to source respondents for market research purposes. They represent a different population of respondents from those typically found in online panels. By virtue of their difference and abundance, we must find ways to include them in our online research.

However, their difference is both a resource and a potential problem. The existing panels have been providing valuable data for years and a sudden inclusion of new respondents has the potential to create data inconsistencies that should be cautiously avoided. We have proposed a conservative and measured way of including these new sources in a granular fashion. Their inherent difference within each demographic cell dictates the maximum blending percentage we feel can comfortably be added to a host population of online panel respondents.

At this time, it is better to err on the conservative side when merging these respondents into existing panels. Thus we have incorporated worst case scenarios involving sample size, income and the amount of statistically measured difference that we allow into our sampling populations.

The management of online samples is shifting from quota fulfillment to a concern for total sample frame. This type of approach is sensitive to the overriding philosophy that those who use these samples must be confident that the change that they see in their data is real and not an artifact generated by shifts in the constituent elements of the sample source being employed. Sample providers have a responsibility to be transparent about their sample frame. It is only through clarity that research practitioners can understand how to interpret their data and it is only through that clarity that end users will know what reliance to place upon it.

Once methods are employed to assure quality they cannot be “one time” credentials that pale with time: They are neither static nor do they transcend geographies. In the best of worlds they are sensitive to changing social, political, and economic conditions. As in all other quality metrics we do not consider the blending ratios to be static, therefore comparative analysis must be an ongoing endeavour.

References

Gittelman, Steven & Elaine Trimarchi 2010. , “ONLINE RESEARCH …AND ALL THAT JAZZ! THE PRACTICAL ADAPTATION OF OLD TUNES TO MAKE NEW MUSIC.” ESOMAR Online Research 2010.

Walker, Robert, Raymond Pettit, and Joel Rubinson. 2009. “A SPECIAL REPORT FROM THE ADVERTISING RESEARCH FOUNDATION: The Foundations of Quality Initiative, A Five-Part Immersion into the Quality of Online Research.” *Journal of advertising Research* 49: 464-485.

Steven Gittelman, Ph.D. | President

Phone (631) 277-7000 | Cell (631) 466-6604

Steve@mktginc.com | www.MktgInc.com

Adam Portner

Senior Vice President, Client Development

220 Montgomery Street, Suite 1058

San Francisco, CA 94104

Direct: (415) 948-2230