On the road to clarity: Differences between sample sources

Authors: Steve Gittelman and Elaine Trimarchi

"It's not the panel quality that matters per se, it is the quality of the end data that the panel produces that matters." (Briggs and Walton 2007)

In the market research universe that boldly completes some $4 billion dollars of online research we still cling to a residual guilt that our sample has gone from a rigorous reliable foundation to the "wild west." We cling to the concept that online sampling must replicate the offline universe to be of value; that panels must provide decent recovery and that the crisis in quality is a myth. None of this is true. Online research is here to stay; it is cost effective and fast. Client companies will either buy it here or go across the street to someone else willing to fill the demand. It is time for the industry to obtain a clear perspective on how to use this valuable tool. This paper focuses on achieving clarity: online sampling frameworks are not interchangeable.

We begin with an essential basic: ***Our data is used to advise clients on how to best make their business decisions***. Critical to those decisions is not the demography of certain categories of "problem" respondents but rather how those respondents influence conclusions drawn about buying and purchasing behaviors.

It benefits the research industry to have a better understanding of its sampling resources. Online access panels are a critical element of that industry resource. This paper is the outcome of a need for transparency in understanding the potential errors in panel research. As a company that frequently uses online panels, we need to know more in order to advise our clients.

One category of respondents receiving intense attention are the frequent survey takers, the professional respondents. Our objective here is to understand, not to judge them. In fact we view them as a valued part of our sampling frame; after all they are the respondents who seem most anxious to participate in our most important measurement tool: the questionnaire. We do not intend to discuss the impact that professional respondents have on data quality. The word "quality" still requires definition. We are interested in determining how they influence our assessment of buying and purchasing behaviors.

INTRODUCTION

This study relies on a relatively large number of data sources to make its point. We could have used a large sample from one source to make parallel points, but we have a different objective. While we do not seek to pass value judgments (in fact we believe professional respondents should be re-named "acculturated" respondents) their potential to alter sampling frames and thus cloud our conclusions is a concern to us. If the professional respondents (as well as their sister behaviors, speeders, satisficers etc.) increase the variability of our data then we had better know. Our purpose is to describe how these different types of respondents affect the variability of data between online panel sources.

Thus we shall focus here not on describing different types of respondents but on how the variability they introduce impacts on the utility of our sampling frames. Simply put, we shall test the following hypotheses that appear to cascade from one to the other:

1. Respondents who choose to complete many surveys and/or belong to many panels are different demographically from the rest of the respondents in those sample sources.
2. These demographically different respondents also exhibit different buying behaviors.
3. These groups of respondents are found in different intensities from one sample source to the next.
4. The differences in intensity may be related to the source populations from which the sample sources are drawn.
5. As the sample sources contain differing levels of various respondent groups, they also differ demographically and behaviorally.
6. Thus the sample sources differ in their buying/purchasing behavior profiles.

The Professional Respondent

The concept of a professional respondent focuses on three characteristics: they readily remain on panels for an extended period of time, complete surveys often, and belong to multiple panels.

There is evidence that respondents who remain in panels over time actually change during that period (Coen, Lorch, Piekarski, 2005). Inexperienced respondents were more favorable on brand purchasing intent questions then experienced respondents. Completion of multiple surveys was also found to be a more "sensitive" measure than tenure. Panelists who participated in a greater number of questionnaires were more negative to a concept while those who had been long term but less active members of a panel did not share the same shift of opinion. The authors concluded: "This research underscores the importance of maintaining a stable panel and of using a panel which can provide a consistent mix of respondent experience over time." In fact Coen et. al. could not "separate the wheat from the chaff" and concluded that intent to buy responses given by frequent survey takers were more in line with reality.

Some have been inclined to accept professional respondents on a literal basis. Sparrow (2007) suggested that respondents as a whole be treated as employees, much as the name "professional" implies, and if they "don't play fair" certain rules should apply. His conclusion was that weighting was ineffective, financial motivations of respondents caused mode effects that could not be mitigated by weighting demographics, attitudinal variables and newspaper readership.

Belonging to Multiple Panels

Comscore Networks reported in October of 2006 at CASRO that less than 1% of respondents in the ten largest online survey panels in the United States accounted for 34% of the completed questionnaires. Comley (2005) appears to have been early to signal the trend toward a growing population of professional respondents who engage in surveys for a number of panel companies. He found that three-quarters of panel members belonged to three or more panels. These multi-panel participants hungrily sought to complete a few interviews a week.

Major attitudinal differences were found in respondents who belonged to multiple panels (Casdas, Fine and Menictas, 2006) although they were demographically similar to those in only one panel. The differences went far beyond demographics and required weighting by covariates. The authors used propensity score adjustment wherein parallel data collected from one method (telephone) is used to bolster that of another (online). Statistical models estimate the probability (the propensity score) of falling into the telephone data and then the online study is in turn weighted to insure that the propensity scores of the two modes match. Casdas et. al. found that the increase in multiple panel membership is accompanied by a correlated increase in dissimilarity from the telephone data set.

De Wulf (2007) found a silver lining: those who belonged to multiple panels were better respondents in some respects. They were more positive towards the process, willing to complete additional surveys, and viewed others in the panel process in a positive light. Although it was thought that this should generate a higher overall response rate among multiple panel members, it was not clearly the case. The most striking finding was that multiple panel participants had "slightly superior" response quality. De Wulf found that money was not the lead motive driving multiple panel participants. They seemed more moved by intrinsic motivations such as the need to learn new things (31%) and the desire to help others (25%) with "interested in receiving a financial reward" trailing at (16%).

The reported characteristics of multiple panel respondents was refuted by De Wulf (2007) who concluded that "The often used image of lower social class, poorly educated housewife filling out surveys as a complementary source of funding does not seem to be true." Among many variables tested (age, gender, education, income and professional status) only one appeared to be significant: they had a higher chance of being unemployed or looking for a job.

De Wulf (2007) tested a number of sources and concluded "…we should not be too concerned about the fact that people join multiple panels as their quality is not inferior to people who have only joined one panel." Further, "…we should abandon stereotyping people subscribed to multiple panels." On the issue of sourcing: "…multi-method recruitment for building online panels is not a necessary condition for building a good quality panel. Most recruitment methods have shown to recruit quite similar profiles."

Panel Sources

There is quite a bit of argument about the validity of response rate in evaluating research quality using online research panels. Given the propensity of certain respondent groups to leap to the task of doing online surveys it appears that response rate can easily be managed by choosing among frequent participants. Thus questions arose about various sourcing methods. As previously mentioned De Wulf (2007) tested five sourcing methods: online ads, e-mail, Intercept, phone, and pop-up surveys. Only the personal intercept method yielded different results, attributed "to the nature of the events." In terms of quality, all methods except e-mail were approximately equal, with online methods yielding higher click and participation rates.

Taylor (2007) called for additional study on the differences between panel methods. De Wulf et. al. (2008) compared access and custom panels. Online access panels are described as a pool of adult subjects who have agreed to be surveyed online. Custom panels focus on interest groups, perhaps customers or others that should have greater subject interest and a sense of community toward the research. De Wulf et. al. concluded that there were slight differences in responses in concept tests between access and custom panels, but the differences faded over time; there were no large conditioning effects and quality, measured by response rate overwhelmingly benefited custom/dedicated panels. There are many that believe that the recruitment of panel members is the driver of differences between the panels and a statement of their inherent quality (Pineau and Slotwiner, 2003).

Balancing Samples

Research data has migrated to a non-probabilistic frame from a probabilistic frame that we can only long for as a fond memory of the old days. A year ago we witnessed a practitioner publicly confess that the days of probability sampling were, well, "forget it." (IIR Chicago, April 2008) Others have joined the pragmatic chorus (Huizing, Ossenbruggen, Muller, van der Wal, Lensvelt-Mulders, and Hubregste, 2007; Kellner, 2007) refusing to abandon a valued resource under fire, while clinging to a concept that may have dwindled in practicality, if it remains achievable to any real economically viable sense at all.

Harlow (2008) compared online panels that were created through probabilistic and non-probabilistic methods as employed in business to business (IT) research. Speeders were found to be far more prevalent in non-probabilistic (19%) than in probabilistic (1%) samples. Straight lining, suggestive of inattentive respondents, (Baker and Downes-Leguin, 2007) was far more present in non-probabilistic panel data. More confounding was that substantial "noise" was found in the non-probabilistic data that clouded the interpretation of research results. Bivariate analyses were "masked" making it difficult to discern important relationships. Only the presence of probabilistic sampling frames helped to identify hidden patterns that were clouded in the non-probabilistic data. Given that bivariate relationships were hard to interpret, it was no surprise that multivariate relationships were further illusive. The probabilistic sampling frame yielded a robust factor analysis that discriminated four factors that drove purchase decisions. A parallel analysis in the non-probabilistic data was different and did not tie into an accompanying telephone data set. In essence, Harlow was unable to achieve his business goals on behalf of his client using the non-probabilistic data as it was misleading and led to clouded interpretations.

Huizing et. al. (2006), proposed the use of what they termed *response inclination*, "the probability of an element of the designed sample to respond to a survey invitation" as a means of correcting sampling bias. Their view was that neither quota sampling nor weighting schemes takes into account differing individual response inclinations. To calculate the respondent's response inclination, the response history must be tracked and added as a second step to the propensity scoring. By taking response inclination into consideration low responders could be attracted in greater frequency. As good as it all sounds Huizing et. al. admits that there is no way for their method to adjust for missing elements in the source population from which panels are drawn.

Attempts to purge the data (Harlow, 2008) prove to no avail. By removing suspect respondents from non-probabilistic data too much had to be extracted and the value of the data set was put into question. It did not appear to be enough to remove speeders and other forms of inattentiveness; the remaining data still lacked precision.

In market research we must begin to think of sample resources as specialists. Though they might aspire to be our wistfully remembered probability samples, they are sourced in an eclectic fashion. It is best to think of them as specialized resources and use a combination of them to maximize their strengths. The differences between sourcing categories is a magnet for discussion, but the homogeneity of data obtained within a category has received little attention.

METHODS

As our intent was to understand the differences in purchasing behavior between data sources we were privileged to obtain cooperation from sixteen different data sources ranging from river to panel to phone in the United States, and to liven up the comparisons, one panel in the United Kingdom. Two companies participated twice and a third five times. Four panel companies participated on a blind basis and were paid for their participation. Two companies participated on a paid basis but were not blind and the balance provided panel members at no charge, but was aware of their participation.

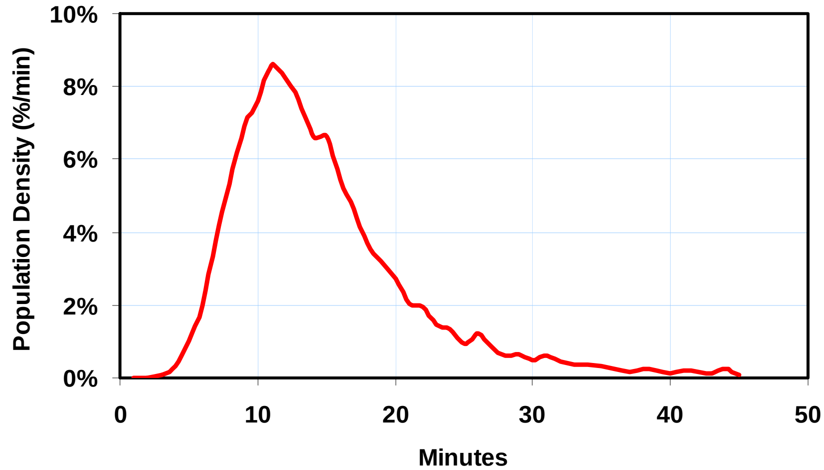
Data was collected from December 2007 to December 2008 at our offices in Long Island, New York. The survey instrument was approximately thirteen minutes in length. A total of 7,600 interviews were collected, about 400 completes per cell. Quotas by ethnicity, income, gender and age were set by completed interviews to replicate the US census. In the United Kingdom income was replaced by social grade. Telephone data collection was performed at our facility on Long Island, New York.

Respondent Challenges

The literature on various respondent groups is growing rapidly and is exhaustive. Here we investigate four different respondent challenges:

1. ***Failure to follow instructions***: Respondents failed to answer particular questions with designated answers.
2. ***Inconsistencies:*** We asked a few questions that provided statements that appeared mutually exclusive such as: “Brand over Price vs. Price over Brand”.
3. ***Speeders:*** The average length of the survey was thirteen minutes. We chose to designate the quickest ten percent as speeders. Examination of the questionnaire duration curve (Figure 1) gave us no rational point to discriminate speeders so we used this rather arbitrary number. Other practitioners have used half of the average as the division point, a no less arbitrary determination.

Figure 1. Survey length in minutes



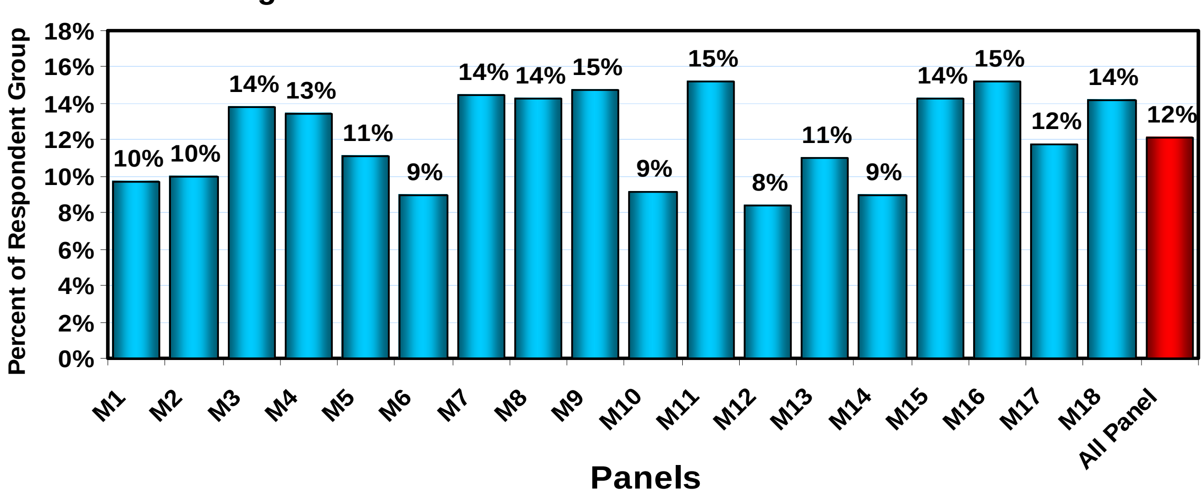
4. ***Professional respondents***:

- Respondents who reported that they take online surveys "practically every day" representing 25% of the total respondent pool.
- Respondents who reported in an open ended format that they took over thirty online surveys in the past month, representing 15% of the total.
- Respondents who identified on an aided basis $\geq$5 panels that they currently serve as members representing 36% of the total respondent pool.

RESULTS

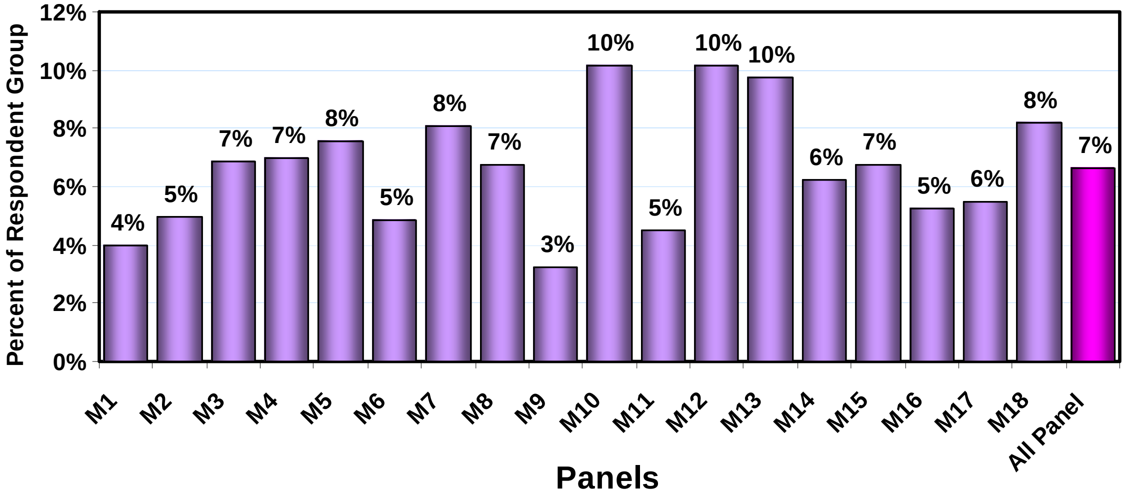
The failure to follow instructions, varied from a high of 15% in panel M16 and a low of 8% in M12, with a median across panels of 12% (Figure 2, panels presented in the order the data was collected). The differences between panels are significant in many comparisons but we were unable to attribute the variation in panels to a failure to follow instructions.

Figure 2. The failure to follow instructions.



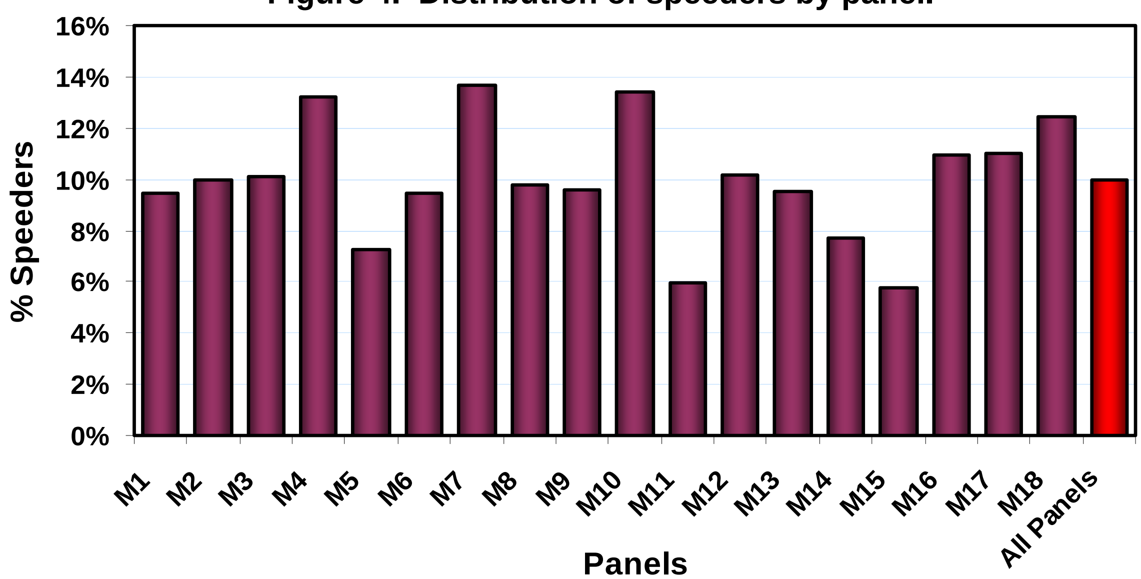
The brand over price consistency error varied dramatically between panels (Figure 3, panels presented in the order data was collected). The three-fold spread between low (3%) and high (10%) is clearly significant and meaningful.

Figure 3. Inconsistency, "Brand over price vs. price over brand."



Speeders, although somewhat arbitrarily designated at the quickest 10%, proved to be another measure that varied substantively between panels (Figure 4).

Figure 4. Distribution of speeders by panel.

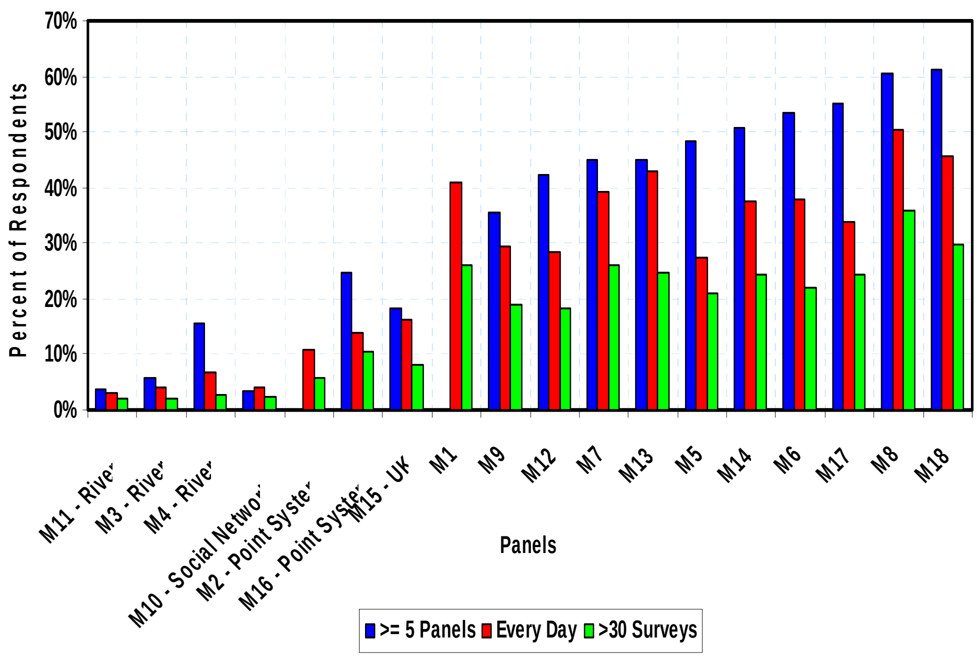


Panel Sourcing and Respondent Types

Perhaps the most widely discussed measure of panel quality has been the presence of professional respondents (Figure 5, panels presented by sourcing subcategory). All three measures were highly correlated (average ρ = 0.85) and are thus somewhat redundant. It is here that the variation between panels is overwhelming and not just statistically significant but meaningful. Such vast shifts in the frequency of professional respondents imply immense differences between panels. If professional respondents represent different purchasing behaviors it is likely that the panels will also be different by that measure. The root of these differences appears to be related to the sourcing models used by the sample providers.

“River” sampling represents our first three sample sources. These represent three waves of the same online sample product. We propose here an ideal model for the River methodology which we designate “theoretical River (TR)”. In this ideal methodology, which we also believe to be economically unfeasible, portals are randomly opened in a great variety of websites representing a large percentage of the online universe. In our dream world these portals allow respondents in but are closed to re-entry for that respondent. Further the probability of finding the portal at any given time in any location is random.

Figure 5. Professional respondents, sample sources grouped by type[1].



Our mythical TR respondents are asked very few questions about their demography. Little or no information about them is collected and they are not acculturated to the market research experience before being exposed to the screener of the one and only questionnaire that will be offered to them. Upon completing a screener they are either terminated or complete the questionnaire and then returned to the river where they are not allowed re-entry to the process.

This mythical practice approaches ideal in many ways. Respondents having the predilection to do frequent surveys cannot be aggregated since they are denied re-entry. Further those going through the interviewing process cannot learn from the experience and carry it over into the next

[1] Panel membership was not asked in M1 and M2.

survey. Essentially they remain un-acculturated to the market research experience and cannot develop traits that are attributed to increasing tenure in some panels. Essentially exposure to the process has little impact because they cannot easily return for second surveys.

Unfortunately TR is impractical. The inability to collect data about the respondents eliminates the possibility to target them for lower incidence studies. The mere concept of releasing willing and able online respondents without using them again is a business model heresy. The costs would be prohibitive and no company would be able to provide the traffic required to meet needs.

Thus companies who have River do not have TR. They must modify it somewhat to achieve reasonable business objectives. Examples of such changes are the collection of some data on the respondents as they first enter the system so that they can be targeted to various screener buckets. A second modification is to expose respondents to a second screener bucket if they fail to qualify at first.

The further we get from TR the more we weaken the value of the research model and favor the panel business model. Some companies have utilized a large flow of respondents who have not been entered as a panel as a "River." To us a river has to approach TR to retain the designation. Those who expose respondents to an endless stream of questionnaire screeners until they qualify are mitigating the value of river. We would also react similarly to those who expose respondents to an endless stream of questionnaires. Obviously once a respondent is exposed to either a long string of questionnaires or their associated screeners we drift further from TR and closer to panel.

One of the artifacts of River methodology should be that repeat responders cannot in fact return with any frequency. Thus we would expect "Rivers" to have a low frequency of professional respondents. M11, M3, and M4 have low professional respondent rates and represent a river source that collects a small amount of data on incoming candidates and exposes them to a small number of (less than three) screeners and no more than two questionnaires. They also claim to block respondents from re-entering the system and do not use the respondents again as part of their River sourcing model.

Source M10 is a social networking site of the Web 2.0 genre. These sites are non-commercial in that its members do not complete financial transactions as a part of being at the site. Instead the motivation is social exchange. Demographically, the site is younger in age (the provider was unable to complete the older age quotas), better educated (Figure 6), have higher frequency of minors in their households (Figure 7), and more frequently employed or a student and far less often retired than found in all three professional respondent categories (Figure 8).

At this juncture professional respondents may be excluded from the social networking universe as they do not find it appealing. Conversely, those who frequent social networks may find doing an abundance of market research interviews unappealing. Sources M2 and M16 consist of members who were drawn from point system cultures where certain purchases provide them with a point reward interchangeable with cash or other gratuities; e.g. frequent fliers. We are not speaking here of the incentives that respondents who complete questionnaires receive as a reward.

M2 and M16 consist of the most highly educated panels (Figure 6). Their households have fewer minors than other panels (Figure 7). This group consists of the full time employed and the less frequently retired, quite the opposite of the professional respondents (Figure 8).

Inherent in these sources is the ability to make purchases but as professional respondents tend to be less often employed and make few online purchases they are unlikely to participate in such buying communities. We believe that the mix found in the sourcing models of these two panels exclude a great percentage of the professional respondents that might be found elsewhere. We are choosing to ignore the possibility that these panels have found a way to remove professional respondents through panel management. Of course this is possible but as we have no conclusive evidence as to how effective their methods are to either support or refute the contention; we will ignore the issue.

Source M15 is our red herring in the mix: The United Kingdom. We would expect that broad differences exist in the social structure of the UK and the US. In these data they are less often employed; less often receive college and post-graduate educations, but are most likely to receive trade or technical school educations (Figure 6). However, from discussions with the panel provider we have learned that a large proportion of the sourcing here is also due to a point system culture. As British and American markets are so different we will just leave it at that. The EU has mandated strict privacy regulations that have yet to reach American shores. European panel managers may be approaching the question of appropriate practice differently. This is an area which shall be the focus of a further analysis.

Figure 6. Educational distribution by panel. No quotas were set by education.

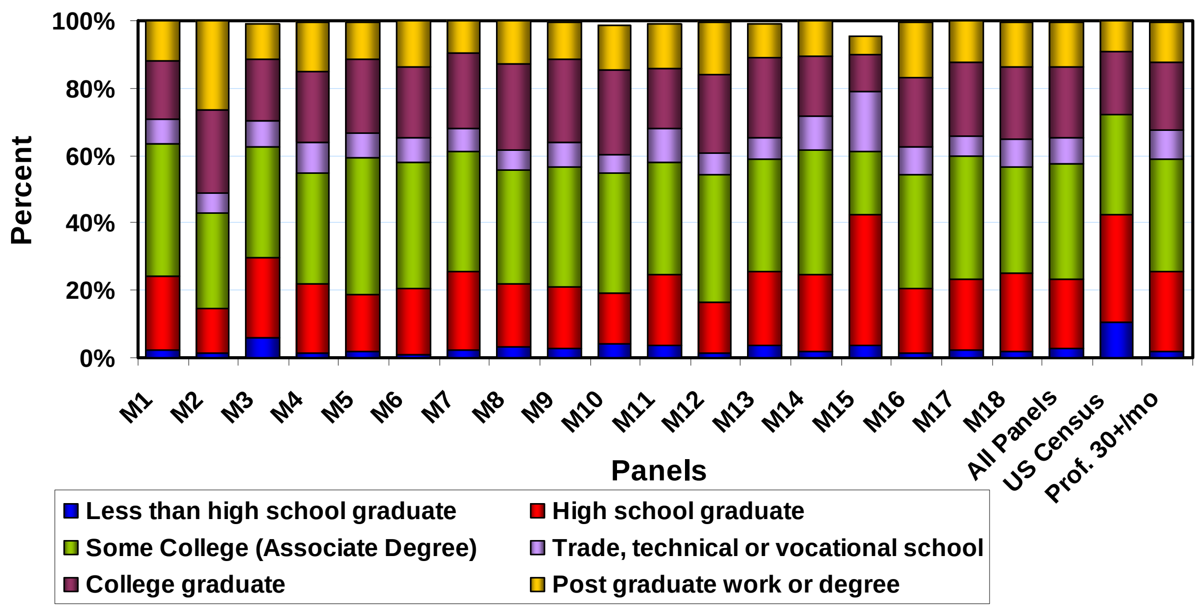


Figure 7. The presence of minor children under the age of 18.

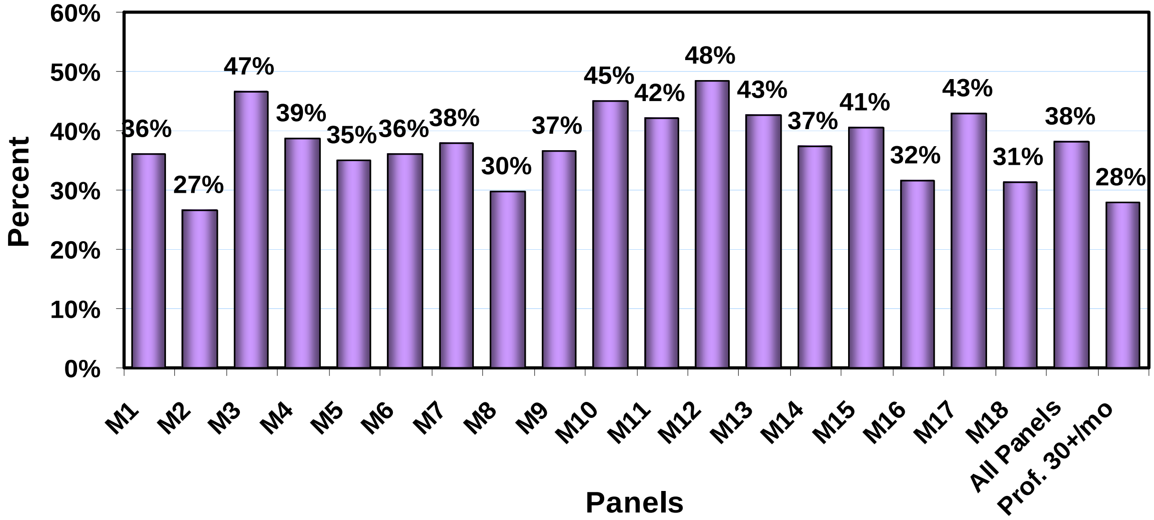
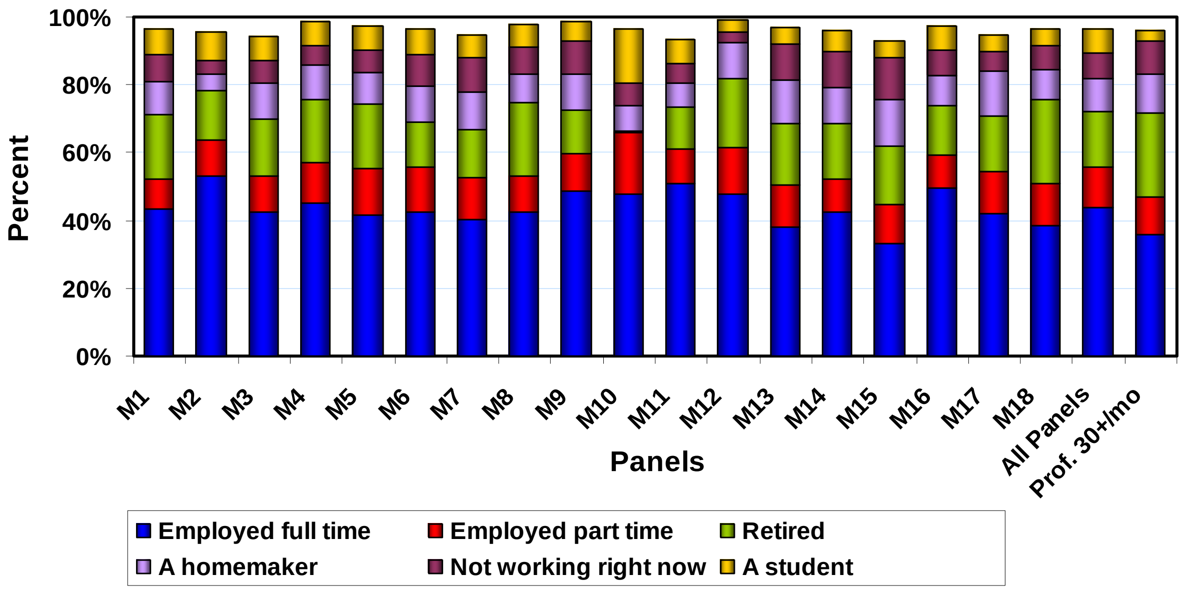


Figure 8. Employment distribution by panel. (No employment quotas were set).

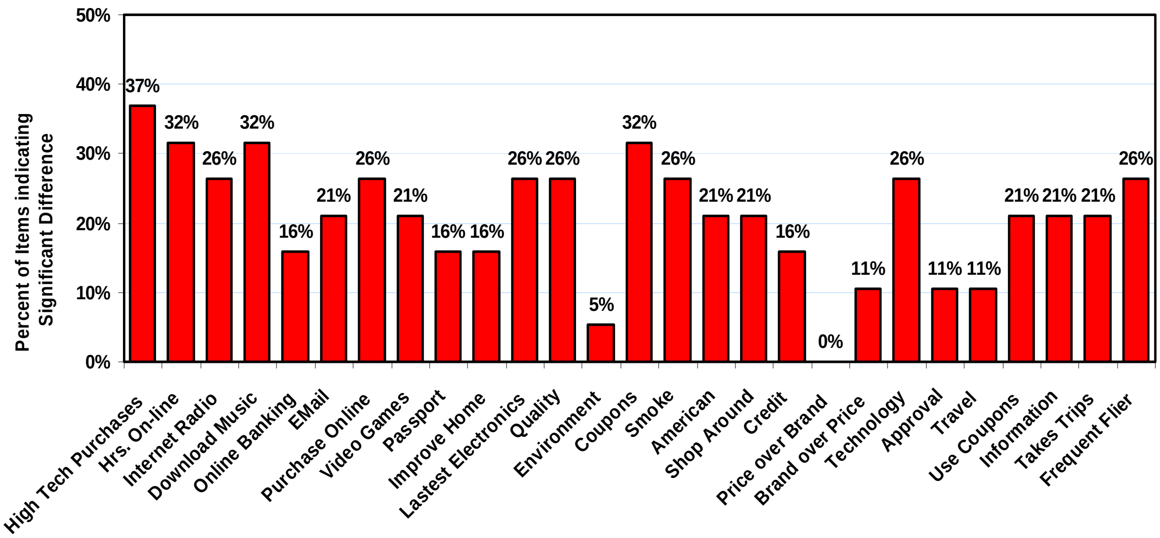


Where the rubber hits the road: Buyer behavior

Data was collected on a total of 27 purchasing related variables (figure 9). Some appear more germane than others, but in total they represent a broad spectrum of questions that give us a first look across panels on this measure. There is probably an endless array of possible ways to measure purchasing behavior and it is not our mission here to exhaust the possibilities.

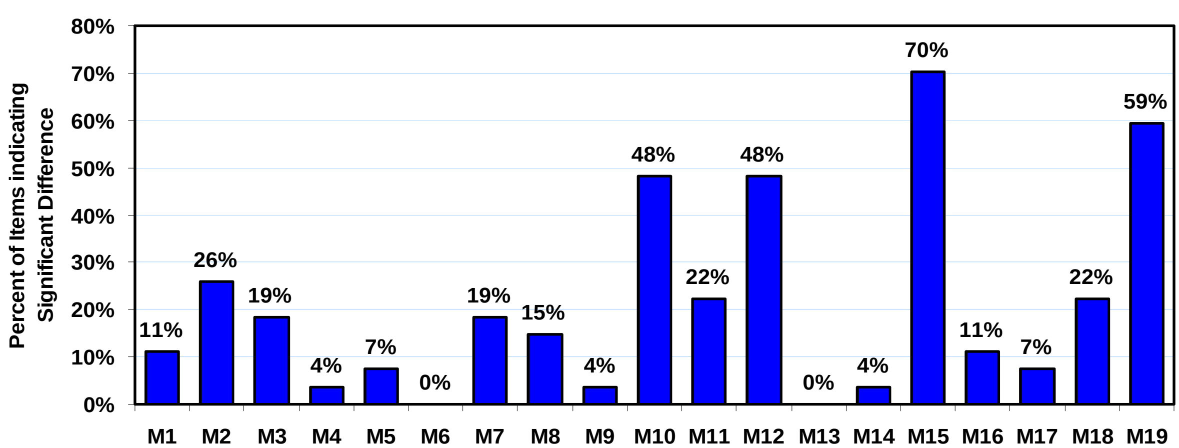
We have summarized the purchasing variables in Figure 9, where the percentage is the number of data sources that were significantly different from the mean of all sources, or the "Grand Mean." Included in that mean are the United Kingdom and one cell of CATI interviews. The panels range dramatically around the purchasing measures. The range in variability peaks at high tech purchases where 37% of the panels significantly differ from the grand mean while there are none who do so around "price over brand".

Figure 9. Buyer Behavior Variability across panels by behavioral measures.



Perhaps more important is the variability between panels, Figure 10. It comes as no surprise that buyer behavior of a panel in the United Kingdom (M15) would differ from American sources. And given our previous analysis of M10, the social network site, we are comfortable with its distinguishing level of difference (48%). It should also come as no surprise that M19, RDD telephone, represents a different state of mind when it is compared to the online sources. There is an abundance of data that shows telephone and online as providing comparable information but as sourcing methodologies they are quite different and there is a commensurate difference in the purchasing patterns evidenced here. Our two point system models M2 and M16 are not extreme outliers. And the three rivers, M3, M4 and M11 seem reasonably clustered.

Figure 10. Frequency of significant differences from the grand mean of purchasing data[2].



[2] Significance at three standard errors, where the total n=27 possible buyer behavior measures.

We clustered the data using 27 purchasing measures into five purchasing segments. Table 1 shows the components of the segments in terms of major differences.

Table 1, Buyer Behavior Segment Description

	Description	Internet	Credit	Brand over Price	On-line Banking	Shopper	Scale	
Segment 1	*Purchasers/Credit Cards/ Not On-Line*	blue	red	orange	blue	red	Very High	red
Segment 2	*Shoppers/No Credit Cards*	white	blue	light blue	white	orange	High	orange
Segment 3	*Non-Purchasers/On-line/ No Credit Cards/Price*	orange	light blue	white	orange	blue	Neutral	white
Segment 4	*On-Line/Credit Cards*	white	red	white	light blue	white	Low	light blue
Segment 5	*On-line/Not Price/OLBanking*	red	white	red	red	light blue	Very Low	blue

The solution provided us with five exclusive segments that reflect combination of the underlying variables. The five segments appear to vary broadly between panels. Once again our sourcing models seem to impact the behaviors we witness (Figure 12). M10, the social network, shows the highest frequency of traditional purchasers, who use credit cards, off line. We have learned to expect the reverse of these internet savvy respondents. They will shop until they drop but are on the web for social exchange not purchasing. It would appear that the social network sites employed by this vender are very different purchasers.

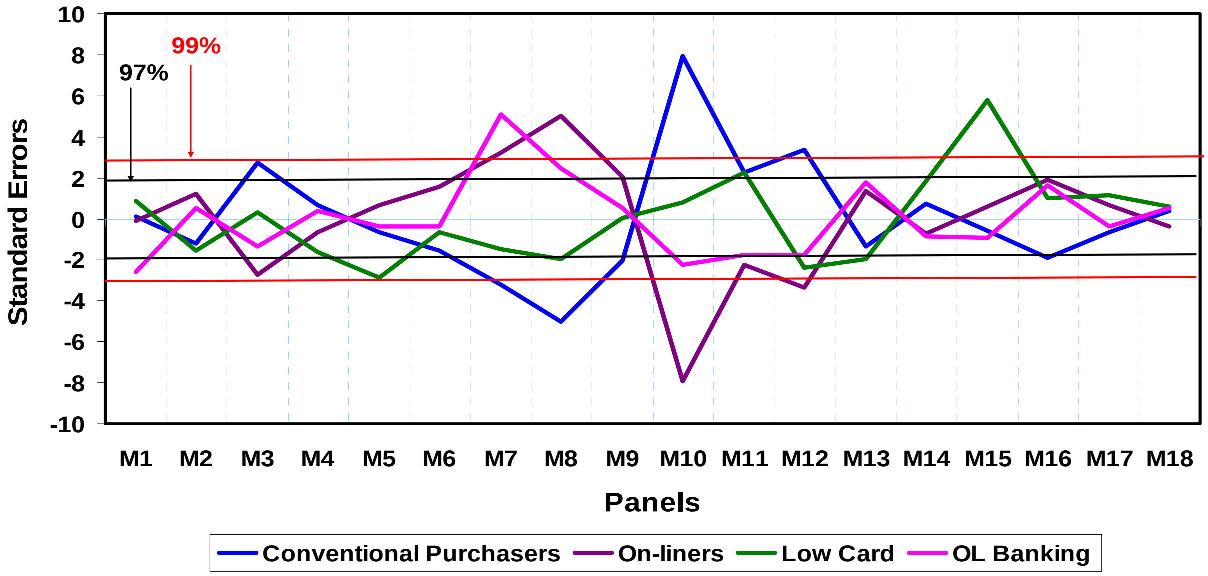
Our British cousins M15 celebrate their difference by being price conscious online purchasers who appear to be less attuned to brand than they are to price. M2 and M16 are apt to be found doing their banking and purchasing on line by credit card as one might expect of those who belong to an on-line purchasing community driven by a point reward system.

The variability we see in other sources remains unexplained. We have given the ESOMAR 25's and 26's from each source close scrutiny but have not been able to explain the variation further. In fact we found answers to the sourcing questions quite difficult to mine for information. One supplier replied that sourcing was "proprietary." We agree, but believe that information is essential to understanding how to use these online resources.

We recombined the underlying variables into four segments (Figure 11) that represent a non-mutually exclusive solution. The four segment solution in Figure 11 allows us to clarify the groups but requires a different graphing technique to provide an image. The line chart highlights differences between panels and at the same time gives insights into the degree of statistical significance and to how meaningful the impact.

Once again, the differences between panels seem somewhat illuminated by sourcing exceptions. M10, the Web 2.0 social networking site, represents the extreme on two accounts, highest on conventional off-line purchasing and lowest for those who purchase online. The United Kingdom, M15, is once again a haven for low credit card use. And the rivers, M3, M4 and M11, while not lying in the extreme seem relatively variable. One would be hard pressed to know in advance which panel to select.

Figure 11. Statistical panel profiles against buyer segments



When we cast the five segment buying behavior solution against various categories of respondent behavior; patterns of similarity and difference standout (Figure 12). Professional respondents, Figure 13, (in this case those who completed over 30 surveys in the last thirty days) might be best understood as price conscious, non purchasers of many of the items we listed, who are frequently on-line but prefer not to use credit cards. Instead they like to shop around looking for bargains.

Those who failed the validity test, that is, did not follow a simple instruction to enter a particular answer appear strikingly similar to the professional respondents, except they have a higher predilection to bank online and are not as price conscious. Those who expressed an inconsistency on brand vs. price and price vs. brand profess to be off-line traditional purchasers who use their credit cards. Given the choice they would choose brand over price, spend considerable time online, do their banking online and will use their credit cards.

Those who are caught up in an inconsistency over happiness are strikingly similar to the speeders. They are traditional on and off-line purchasers who like the convenience of their credit cards.

Figure 12. Buyer behavior segments by panel.

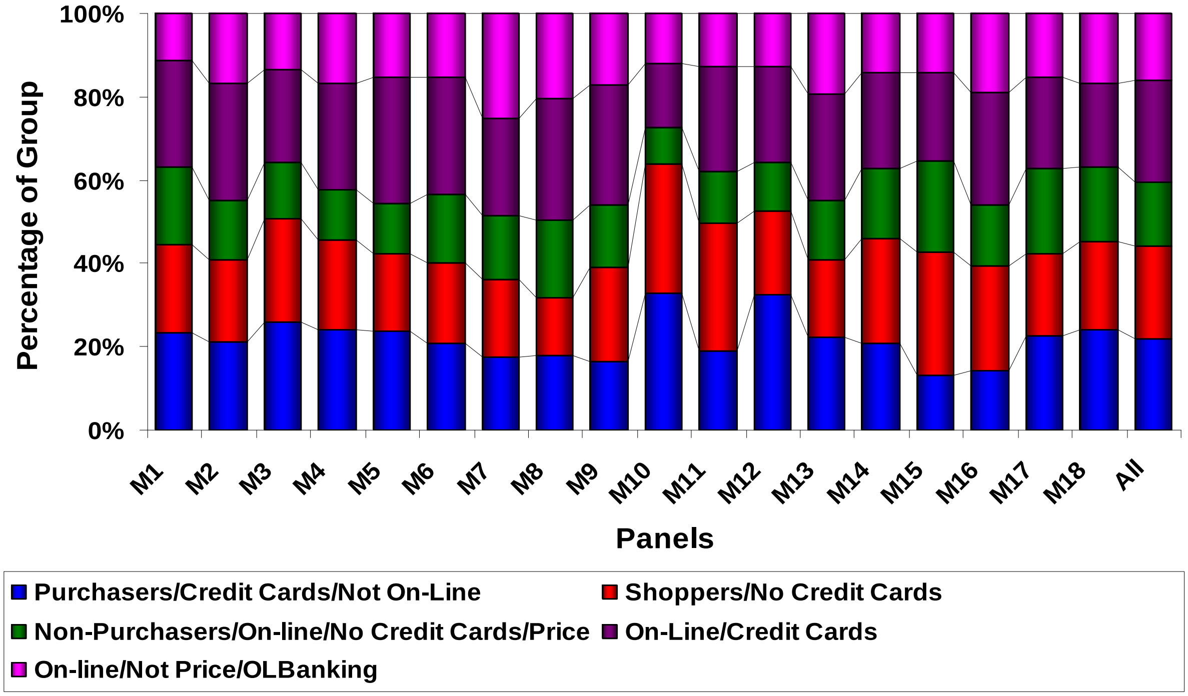
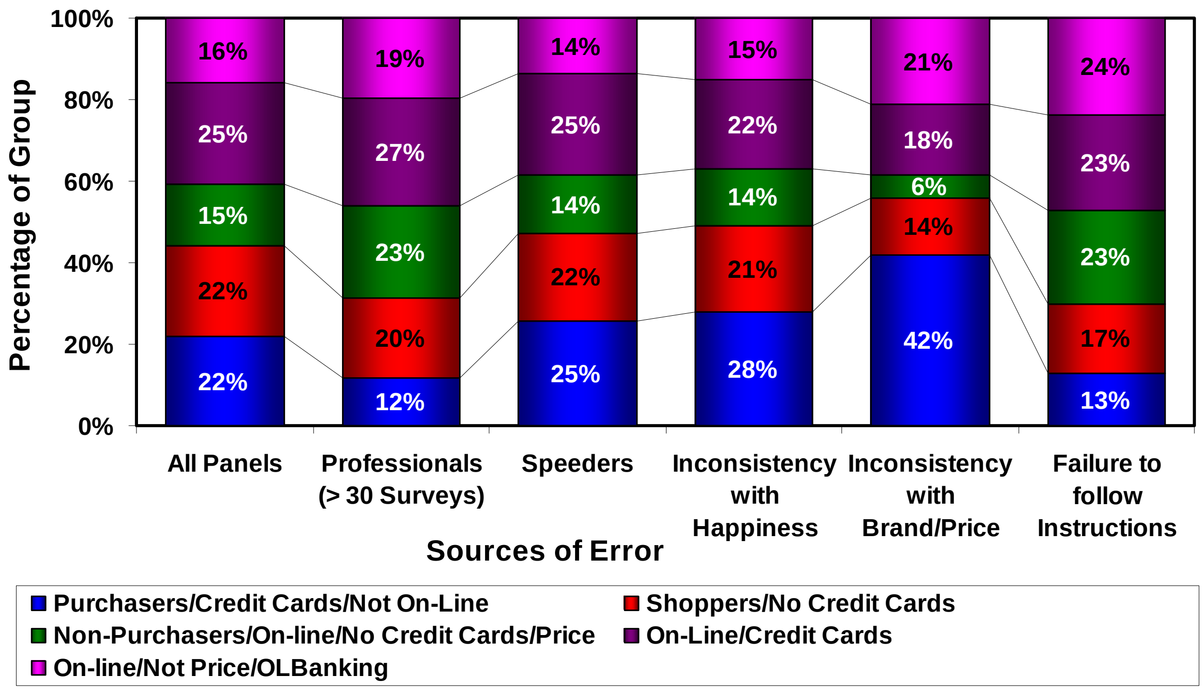


Figure 13. Buyer behavior segments by respondent type.



Discussion

Cross panel comparisons are rare. In what must be considered a landmark study Vonk, Ossenbruggen and Willems (2006) compared 19 of 30 panels then existing in the Netherlands.

Whereas their methodology was different, e.g. they did not quota control to replicate census, but instead relied on release of a 1000 respondents per participating company, many of their conclusions parallel this study.

The panel universe does not replicate the real world. Prudent business decisions are real world and force recruitment and panel management models to fit an optimization between profits and sample quality. So far the business model has worked; the online access panels represent some of the fastest growth centers in our business.

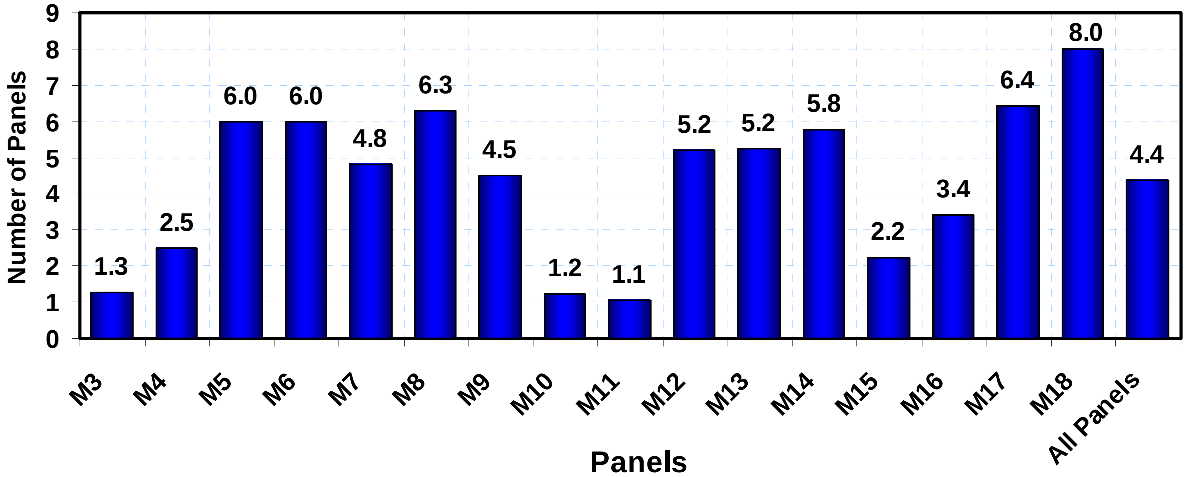
Given the absence of best practice standards, sourcing models are actually business decisions that are dictated by the access afforded a company and the reach of its resources. Panel management is also an evolving art where best practices are in the process of formulation. It is best to consider the combination of sourcing and panel management to be business decisions until a better course is on the horizon.

This data shows considerable differences between sample providers. Vonk et. al. found substantial differences between panels in the above mentioned study with response rates ranging from 18 to 77%. Response rates seemed to be a function of the way in which panel members were recruited. Those obtained through traditional research methods such as questionnaires seemed far more responsive than those who are self selected. The authors concluded that the existence of a prior relationship, such as the completion of a market research interview, would drive an emotional reason to respond to an invitation, whereas those who approached the process on their own volition would have little loyalty generated by human connection.

Whereas response rate was drastically different between panels "there was no difference in outcomes between panels related to response rates." in the Vonk et.al study. They concluded, "In fact, the respondents in online panel surveys are a homogeneous group of people…" They further conclude that the "double opt-in seems to work as an extra filter to potential panelists." Given that the only way that response bias can influence research results is if the respondents who respond are different from those who don't, then the potential response error was determined to be a non-issue. Higher response rate could have no bearing on quality of a panel, it is "…a measure of the efficiency of panel management. It is a panel strategy component driven by rules of economics."

While some may conclude that online respondents are a homogeneous group there are others who would disagree. These data indicate that respondents differ dramatically and that sourcing mode is a driver. One of our measures of professional respondents is particularly telling, the average number of panel memberships per panel (Figure 14) ranges from 1.1 to 8.0. Once again, certain sourcing models seem to generate different results: River respondents, M3, M4 and M11 all taken from the same source were 1.3, 2.5 and 1.1 respectively. Our singular social networking site, M10, was once again an outlier at 1.2 panels per respondent. M16 sourced from a point system environment that appears to exclude the demography associated with professional respondents was low at 3.4 and the United Kingdom, a completely different sourcing environment was at 2.2.

Figure 14. Average number of panel memberships per panel[3].



With these somewhat unique sourcing models eliminated the remainder range from 4.5 to 8. Vonk et. al. found that panels recruited by telephone (2.0), snowballing (2.3) and traditional research (2.4) were meaningfully different from more abundant and typical processes of self registration (3.3), via website surveys (3.3),from existing panels (3.7), via banners (3.7) and by purchase of e-mail lists and subsequent offerings of participation (4.3). River sampling, social networks and point system cultures were not identified in their data.

To further exacerbate the situation those who respond seem to fall into the 80/20 rule where a majority of the questionnaires are completed by a minority of respondents. Professional respondents and the cloud that surrounds them seem to drive this situation as the respondent pool within the panels is different from the respondents who actually complete the questionnaires. Given the overbearing presence of a minority of panel members in the majority of interviews it is possible that the sourcing differences, though high in impact, may be smothered by avid survey takers. However, here the sourcing differences seem to impact on the very presence of the professional respondents. The essence of the outliers is that they have a low frequency of such respondents.

Casdas et.al. (2006), found that multiple panel members were younger, less educated; more often female; not working full-time; more often worked part-time and were more likely to rent as opposed to own their principal residence. As previously mentioned they deviated more from phone data. In addition they showed less reliability on attitudinal in contrast to behavioral measures. They also drank less wine, invested less, smoked more, read magazines more, more often owned pets and had more of them. In many cases our data here agreed.

The impact of these respondents on behavioral data resonated in Casdas et. al. (2006). The predilection of multiple panel respondents to purchase a motor vehicle in the next twelve months correlated with the number of panel memberships. As the number of smokers, home owners, and broadband users increased in their sample so did motor vehicle purchasing intent. Weighting did not reliably cure the situation. They observed, “Where we have obtained independent measures,

[3] M1, M2 and CATI were not asked.

we find that neither single nor multi-panel members accurately reflect the true penetration." Further they conclude, "Every panel will have its own unique profile of panel membership and will require its own weighting approach."

Vonk et. al concluded that professional respondents are, "…to be found on the internet more often, is educated below average, more often unfit for work, less satisfied with his or her income, less satisfied with how democracy is functioning and feels less healthy. In the Netherlands 8% of respondents were deemed to be professional respondents, while in our measure the number appears to be considerably more than that ranging from 37% who belong to greater than four panels; 26% who complete a survey just about every day to and 17% who report completing 30 or more surveys in the past month.

We find that professional respondents are statistically less well educated, but we question how meaningful the difference is. Their incomes are lower, they are less likely employed, their households have fewer young members and their outlook seems less confident. In essence, we agree with the prior authors.

One result reported by Vonk et. al. is that professional respondents were more aware of brands than the average respondent but less familiar with advertisement. The authors concluded that the typology they described might be interesting but probably "nothing more than the heterogeneity of the population. Different people think different things."

They also discovered that different respondent characteristics appeared in different frequencies among panels. They attributed the differences to panel recruitment models. "Both professional and inattentive respondents are both much more likely to be approached from purchased databases or by self-selection…" While inattentiveness and professionalism was highly correlated in their results (ρ =0.76, n=19) it does not appear to be in ours (ρ=0.06, n=18).

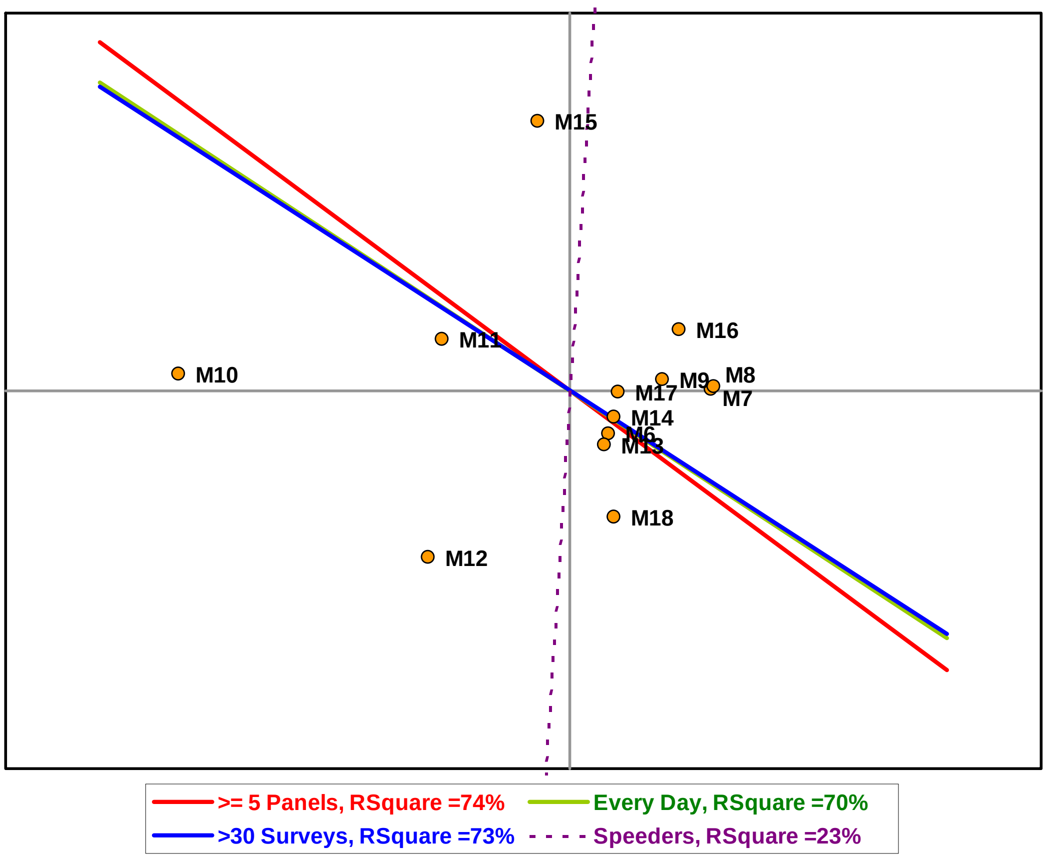
Vonk et.al. carries their analysis to the next logical question: Does the differential distribution of respondent groups who think and behave differently explain the differences between panels? While we are on the road to analyzing purchasing behaviors and buying predilections it seems worthwhile to subject these data to the same analysis that our colleagues in the Netherlands entertained.

As in our predecessors, we calculated the deviations in a matrix, in this case between 13 panels and 27 purchasing questions. This was used as an input for an MDS analysis (Multiple Dimensional Scaling). The resulting coordinates are plotted in Figure 15. Each point on the MDS plot represents a panel. The positioning of the panels to each other translates to a degree of similarity. Thus those panels that are found clustered together would be considered to be most similar. Conversely, those farther apart would be considered to be most different. Once again we find that the sourcing of panels is somehow connected to the difference. M11, river; M10, social network and M15 are clear outliers. M12, previously un-described here is also the product of the same company that offers the river data and is derived from a similar source, albeit using a different methodology. Here, unlike Vonk et. al. our matrix consists only of purchasing/buyer behavior measures. The three professional respondent vectors are highly correlated (R^2 of 0.70 to 0.74, n=13) with panel distribution and provide a connection between purchasing/buying behavior in panels and the presence or absence of these respondents.

In these data the speeders, were significantly correlated (R^2 = 0.23) but not to the degree of the professional respondents. Unfortunately, we did not ask all of the purchasing questions of panels

M1 through M5. Thus the analysis here is restricted to M6 through M18. CATI data is excluded as this is intended to explain the variation between panels.

Figure 15. MDS Plot. Variation between panels appears attributable in large part to the presence of professional respondents[4].



We subjected the data to a principal component analysis to obtain a different analytical perspective, particularly on the speeders. Two examples of the four analyses are depicted in Figures 16 and 17. One of three professional respondent measures, multiple panel membership

[4] Data in this case is based upon 27 purchasing attributes.

and the speeders are shown as vectors in their own analysis in formulation with principal buyer behavior in panels M6 through M18 with the United Kingdom deleted for clarity.

In each measure professional respondents explained the majority of variability in buyer behavior represented by the panels (R^2 > 36%). In contrast, speeders appeared to explain little of the distribution with R^2 approaching 0.4%. There is a considerable body of research that links speeders with professional respondents as a quality issue. However, here they are de-linked. Simply put, while some report that professional respondents speed, in this data it is the behavior of the professional respondents that drives changes in purchasing patterns and not that of speeders.

These data connect professional respondents with differential buying/purchasing behaviors and demonstrates that panel sourcing concentrates these respondents and intensifies their effect. To the extent that panel sources can minimize the frequency of professional respondents the data is actionable, but not practical.

Panels can identify returning respondents by use of digital fingerprinting and other technological advances available to the Market Research industry in recent years, sadly subsequent to 2006 when Vonk et. al. wrote their paper. In the Netherlands, the overlap between panels and the overall online panel population is quite high ranging from a low of 29% to a high of 88%. The highest reported overlap between two single panels was 42%.

Unless panels pool their respondents, even technological advances will limit de-duplication efforts to within panels and between strategic partners. As frequent responders increase without bound, multiple panel participation should remain a growing concern. The critical question remains; What is a sample provider supposed to do?

Best practices would indicate that internal controls should limit the number of surveys performed by panel members. But in an environment where panels have to make business decisions that keep them profitable while providing the best in data quality it remains unclear as to how far an individual panel company can go to eliminate frequent responders.

There are considerable differences between the sourcing models we identified and the panels that employ them. Unfortunately, their differences are difficult to measure and most user companies, especially small and medium sized practitioners, cannot go to the pains that we have here. It would be unreasonable to expect all panels to provide identical participant pools, it is economically unrealistic. We have no choice but to celebrate the differences between sourcing models and capitalize on them as best we can.

Figure 16. Principal components analysis: Buyer behavior against professional respondents who belong to 5 or more panels.

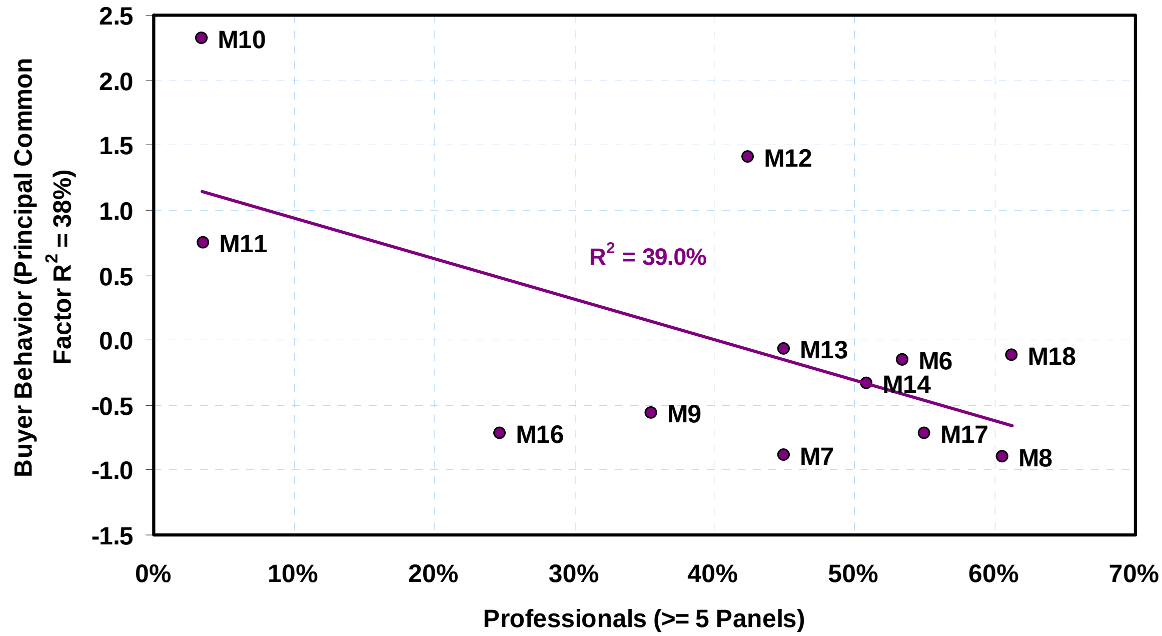
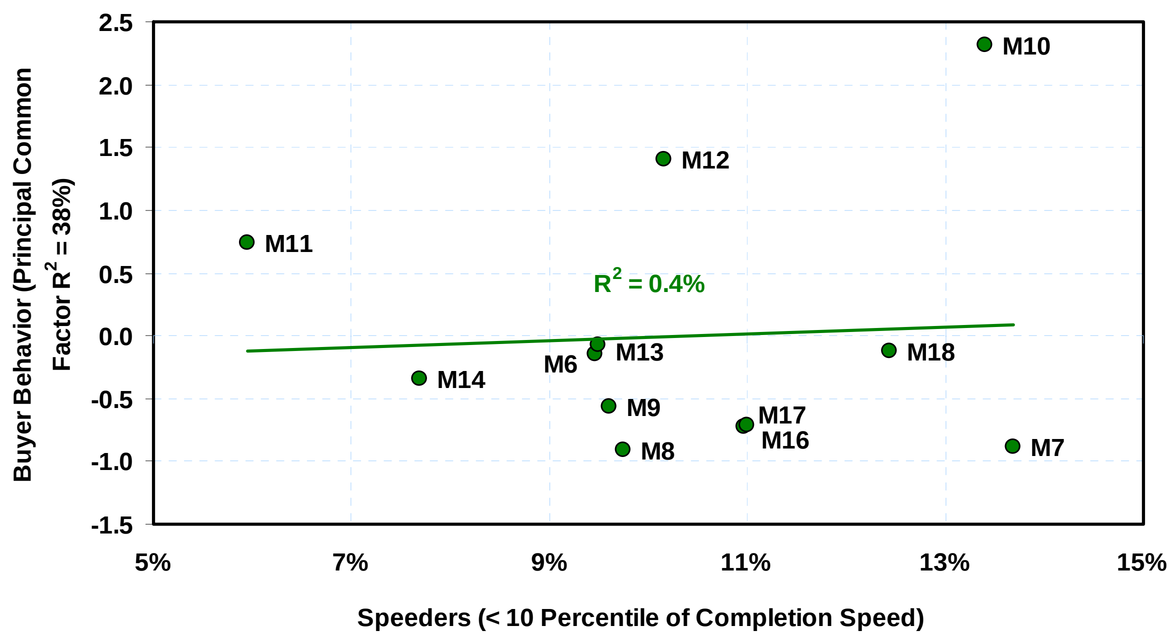
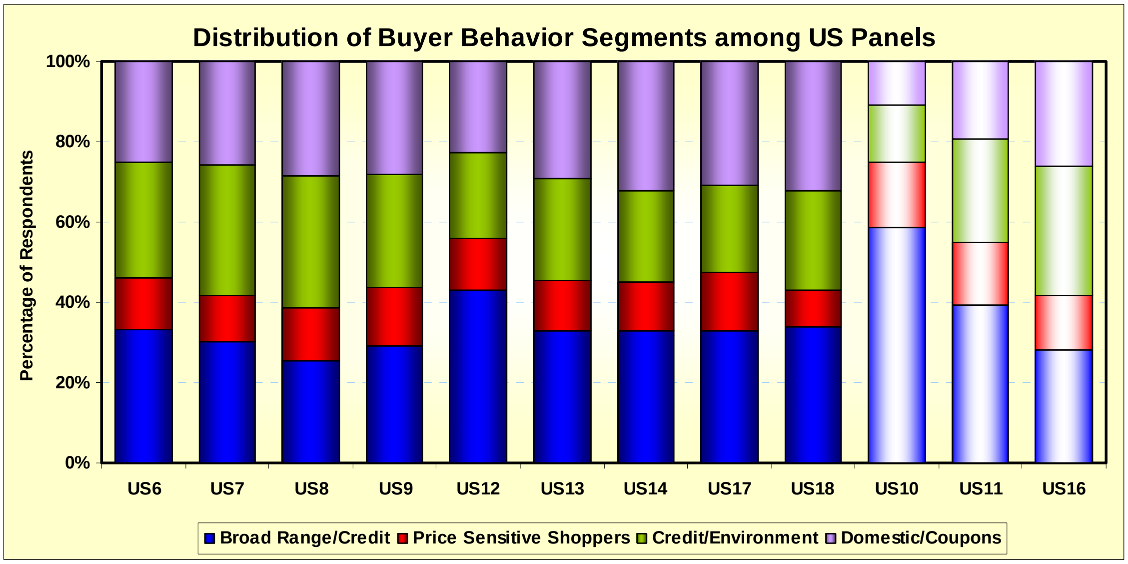


Figure 17. Principal Component analysis: Principal Buyer Behavior vs. Speeding (defined as the shortest 10% of the respondent population.)



However, the differences between sourcing models is not reflected within them. We do not have enough examples of river, point system, and social network sites to draw strong conclusions. A cursory examination of the American access panels (figure 18) seems to prove a point: the purchasing behavior within a sourcing model is quite consistent. We have collected considerable data on global panels and find, as one might expect, extraordinary differences between purchasing profiles across international barriers. One would expect the cultural differences to be resounding and they are. However within national boundaries we find considerable homogeneity. In figure 18, with the exception of panel 12 which we believe is sourced differently than other access panels the purchasing behaviors are more similar than they are different.

Figure 18. Similarity in buyer behavior segments among access panels in the US. Those in bold appear to have similar sourcing models with the exception of US 12 which is known to come from a unique source. US 10 is a social network, US 11 is river and US is point system.



Vive la difference! The argument for diversity.

It is clear from these data that panels coming from reward programs, social networks and river sampling have lower frequencies of professional respondents than do other online access panels. Due to these differences in sourcing models, buying/purchasing behavior is in turn variable over the panels as a whole. There is no wrong or right here, just different. Thus the selection of online research sample is difficult.

Practitioners need stable sample sources that are both predictable and reliable for the future. Given that it is difficult to select a sample source that is representative now, it would seem even more troublesome to find an enduring source that will protect the future. In a typical economy, the sage advice is that diversity provides predictability and stability. The old adage, “do not put all your eggs in one basket” holds true for market research sampling today. Researchers must learn to use multiple sources just as they diversify an investment portfolio.

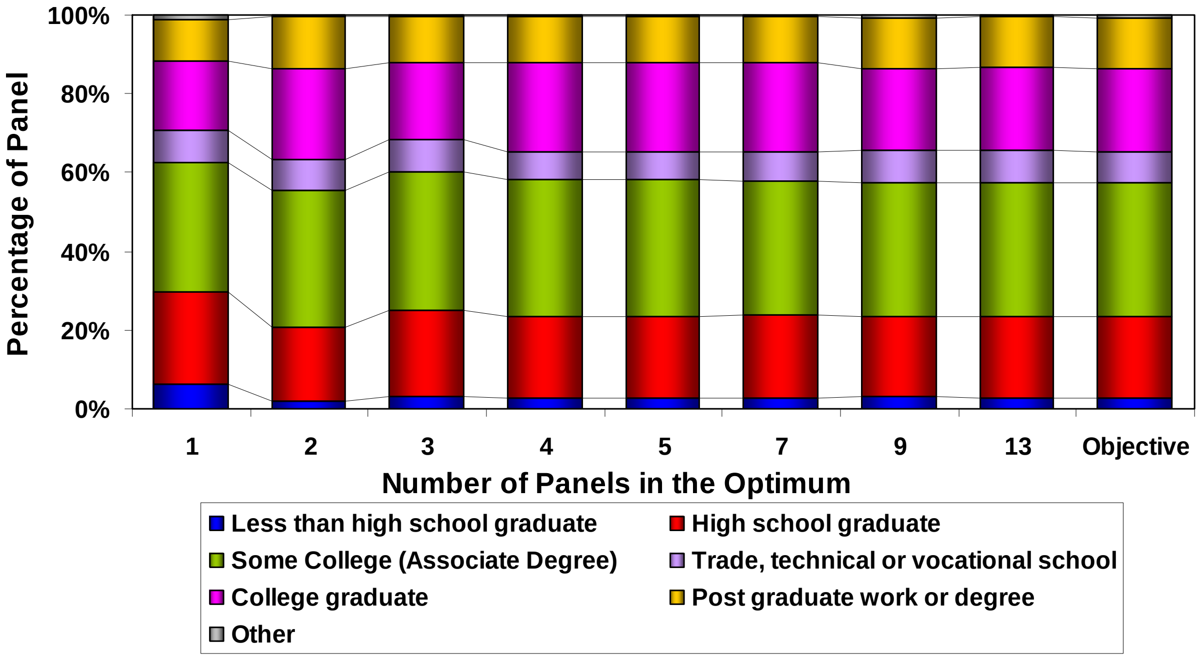
We can aspire to such a model. We have to learn how to subdivide the panel universe. We must capture information about our panel partners and learn to capitalize on their inherent strengths. Gaudemar (2006) collected data on nine American panels and proposed an elegant means of multi-sourcing. Their motivation for proposing the use of multiple panels was the same as ours is here: “…despite a complex and demanding set of nested quotas, quota sampling alone does not totally neutralize the inherent biases of each sample source. The results of a study fielded with one sample source could have been different with another sample source.” To meet the needs of combining sample houses on what appear to be large users of sample a “dynamic routing” system was proposed. The intent was clear, “biases would cancel each other out.” Gaudemar put it succinctly, “It is the sample buyers’ responsibility to use different variables…to combine potential sample partners into an ideal sample mix.”

While we concur, we propose an additional method to achieve targeted sampling. We agree that to achieve a probability sample of old would be a tall order. Instead we suggest that researchers chose their battles. An efficient mix of sample houses could be determined by optimizing against those variables considered of greatest importance to the research at hand. For example, in these data we established quotas to control age, income, ethnicity and gender; we allowed other demographics such as education to float freely.

Perhaps a client is a large university system that seeks a stabilized panel mix that will cancel out purchasing behavior bias, but seeks a balance of educational achievements at least representative of the online community.

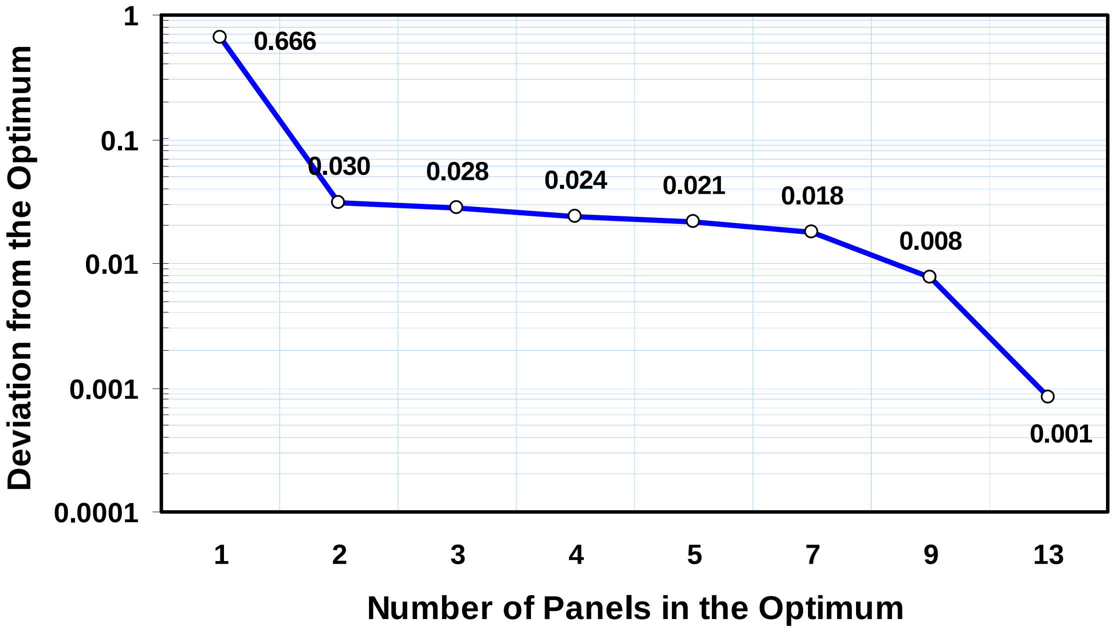
We used an optimization model to achieve at least part of this result. We sought to combine a minimum of panels, to keep administration practical, while maximizing the representativeness of education in our selection. To make the plan actionable we would hope to provide a small number of panels that are recommended for the mix as well as the percentages that they should represent in the final solution. In this case we minimized the panels to three M3 (40%), M6 (40%) and M14 (20%) while optimizing education distribution to simulate the panel universe. The solution is graphically represented in Figure 19.

Figure 19. Optimization model minimizing the number of panels and approximating the average online education attained.



In all optimizations the objective (here balanced education) is ultimately achieved when the entire universe is used. However, the concept is to obtain a reasonable return for effort expended. More simply put, no one is going to do this if it requires too many panels to achieve the desired result. As can be noted in Figure 19, the objective is achieved with trivial deviation at thirteen panels. However, the best cost/benefit return seems to be at between 2 and 5 panels, with two or three being a manageable number providing a balance between diversity and precision.

Figure 20. Return on optimization model. Approaches target as more panels are added into the selection. A three panel alternative was selected.



CONCLUSION

Panels are a diverse community. Sourcing models have the ability to drive the presence or absence of various respondent constituencies; particularly professional respondents. These avid survey takers can cause shifts in the purchasing behaviors that typify various sample sources. Sourcing models that aggregate demographic and psychographic concentrations of respondents likely to take frequent surveys will exhibit one set of buying/purchasing profiles, while those that lack broad representation may differ for other reasons.

Sample providers are likely to continue their evolution. As their business models shift so will the characteristics of their samples. The research industry has to welcome the incoming diversity as an opportunity to capitalize on specialists who are well adapted to their tasks. Web 2.0 and the social networking communities with their philosophical, typological as well as behavioral shifts will bring diversity to our sampling frame that we should not resist but instead embrace through multi-sample design.

We lightly play with the concept of using optimization models as a way to maximize stability and diversity while making sample source selections. We find that there are possibilities in the future to explore niches never before visible to us.

ACKNOWLEDGMENTS

This research was a big task for a small company. Many at Mktg, Inc. helped us nurse it along. We are particularly grateful to Gene Lieb at Custom Decision Support, Inc., for analytical advice

as well as the use of their fabulous Marketing Toolkit, an assemblage of statistical toys that helped us maximize our data.

REFERENCES

Bowman, Jo. (2007). Hands Up, Who's on an Access Panel? Should the market research industry worry about multiple panel membership and its impact on the quality of research results? Research World. ESOMAR Research World Article.

Baker, Reg. (2008). Without A Net: Making business decisions from online panel data. Panel Research 2008. ESOMAR World Research Conference.

Baker, Reg and Theo Downes-Le Guin. (2007), "Separating the Wheat from the chaff: ensuring data quality in internet samples, ASC International Conference on Survey and Statistical Computing

Briggs, Jon; and Mark Walton. 2007. Six Degrees of Online Quality. Panel Research 2007. ESOMAR World Research Conference.

Casdas, Dimitrio; Brian Fine, Con Menictas. 2006. Attitudinal Differences: Comparing people who belong to multiple versus single panels. Panel Research 2006. ESOMAR World Research Conference.

Coen, Terrence; Jacqueline Lorch, Linda Pierkarski. (2005) The effects of survey frequency on panelists' responses. In: ESOMAR World Research Conference, Developments and Progress, Budapest, 17-19 April 2005, p. 409-424.

Comley, Pete. (2005). Understanding the online panelist. Panel Research 2005. ESOMAR World Research Conference.

Comley, Pete. (2006). The Games We Play: A psychoanalysis of the relationship between panel owners and panel participants. Panel Research 2006. ESOMAR World Research Conference.

De Gaudemar, O. (2006). Benefits and challenges of multi-sourcing. Understanding differences between sample sources, Panel Research, 2006, ESOMAR World Research Conference, Barcelona.

De Wulf, Kristof; Sam Berteloot. (2007) Duplication and multi-source online panel recruitment: real quality differences or idle rumours? ESOMAR World Research Conference, Panel Research 2007: good quality, good business, Orlando, 28-30 October 2007, p. 49-62.

De Wulf, Kristof; Mike Friedman, Bert Borgreve. (2008) Pavlov Revisted: Comparing Panel conditioning and quality between panel methods. Panel Research 2008. ESOMAR World Research Paper.

Huizing, Lennart; Robert van Ossenbruggen, Martin Muller, Clasine van der Wal, Gerty J.L.M. Lensvelt-Mulders and Marianne Hubregtse. (2007). Improving Panel Sampling: embedding propensity scores & response behavior in sampling frames." ESOMAR World Research Conference, Panel Research 2007: good quality, good business. Orlando, 28-30 October 2007, p. 108-123.

Kellner, P. (2007). Down with random samples. Research World. May, 31.

Pineau, Vicki and Daniel Slotwiner. 2003. Probability Samples vs. Volunteer Respondents in Internet Research: Defining Potential Effects on Data and Decision-Making in Marketing Applications, Knowledge Networks, Inc.

Smith, Renee and Holland Hofma Brown, 2005. Assessing the quality of data from online panels: Moving forward with confidence," Harris Interactive.

Sparrow, Nick. 2007. Quality issues in online research. Journal of Advertising Research. June, 2007.

Taylor, Eric. (2007), Are panelists employees or respondents? Social versus economic exchange, in : ESOMAR World Research Conference, Panel Research 2007: good quality, good business. Orlando, 28-30 October 2007, p. 19-32.

Vonk, Ted; Robert van Ossenbruggen, Pieter Willems., (2006). The effects of panel recruitment and management on research results. A study across 19 online panels, ESOMAR Panel Research Conference, Barcelona.

About the Authors:

Steven Gittelman is President of Mktg, Inc.
Elaine Trimarchi is Executive Vice President of Mktg, Inc.

