

By Elaine Trimarchi and Steven Gittelman

Growing pains

Speed, price and convenience are just a few of the selling points of online research. These assets have carried the methodology to a level of dominance. Online research, a mere teenager, has reached maturity before its time and now must defend its weaknesses just as other methodologies have before it. The representativeness of its sampling frame is in question, the respondent pool is reaching saturation and overuse syndromes are rearing their heads. End users are increasingly uncomfortable as disaster stories become legend.

Online market research operates under different rules from the telephone interviewing that it replaced. It has moved into a new era, from a probabilistic framework to “work-

Online research comes of age—and hits some snags along the way.

ing without a net,” in a non-probabilistic sampling frame. The explosion of online research is not the root cause of the erosion of our probabilistic framework; its origins go far deeper. Random-digit-dialed telephone samples once provided the industry with the ability to relate back to a sampling frame rooted in the census. If we made enough call backs, converted refusals and adhered to strict calling patterns, it was possible to approach a random selection of households that seemingly gave us a reasonable facsimile of a probabilistic world. Alas, the telephone too has run into trouble as

technology, increasing refusal rates and do not call lists confound its very way of being.

Internet-based research has been troubled from its advent. Our fondest wishes cannot grant it a true probabilistic frame: The offline population is different from those online. Early on, we turned to phone-online comparisons to provide credibility to our new format. One has to wonder if our efforts were not misguided in the first place. As the telephone standard drifted from gold to brass, we were seeking a safety net in the wrong place.

Sample frames can change for a variety of

reasons, such as change in the respondent aging profile, the merger of two or more samples, recruiting respondents from new and/or different sources etc.

Market researchers collect data to answer questions. By use of questionnaires, we ask respondents to provide us with information so that we can answer questions that in turn offer our clients direction on their business decisions. If we ask the wrong questions, then we get the wrong answers. If we ask the right questions of the wrong people (an unstable sampling frame), then we still get the wrong answers. End users rely on us to deliver answers but have traditionally trusted us to ask the right questions of the right people.



Executive Summary

Online research has struggled to come to grips with challenges to the sampling frame. In the past three years, spotty progress has been made, but the credibility of market research remains at risk. New metrics for stabilizing online access panels are needed, but progress has been slowed by a lack of transparency and the need to understand the issues. This article proposes multivariate metrics to combat this global problem.

There was a day when our samples were rooted in comparisons to the census. We could always compare the subsample we were using to assure ourselves that the sampling frame represented the census and thus the population at large. Such samples were considered to be probabilistic in nature. We were all trained this way, right from Statistics 101 in college when the professor said, "...toss a coin, any coin. How big of a sample do you need before you know that the coin toss is fair?" Online samples struggle to meet this threshold; they are non-probabilistic and are difficult to define.

Variability

Fundamental to all research is knowing the differences between shifts in a sample and real changes in the behavior we are measuring. Variability is a key worry of all of us who practice in the research profession. As we try to interpret our data, there is an inner voice we hear that tells us to "watch out for the background noise." If respondents indicate that they have a decreasing desire to purchase a particular product, we must know if the shift is caused by a change in purchasing intent or a side effect caused by a change in the sampling frame.

The chorus of background noise gets louder and louder as respondents seem to do more and more online surveys for an ever increasing online community of sample providers. If all samples were alike and if they didn't change through time, life would be easier.

It seems as if anyone who has any kind of list is in the business of selling samples. The industry has yet to weed out those samples where the sources are unknown and are managed poorly. The problem goes even deeper as pricing pressures limit the willingness of the quality players to pay for a cure.

The panel providers bemoan low prices while the end users are just awakening to low quality. The researchers bid with purchasing agents intent on reducing cost without a meaning-

ful understanding of the quality they are buying. Academics are having a field day tearing into the commercial side as they understand what would be best in an ideal world but are not forced to apply it in the commercial world.

End users have their own internal people to answer to. As research stumbles over sampling frame issues, those that purchase our wares are challenging the value of research. Pity the poor in-house research manager who tries to sell an engineering VP on the need to change his ways, revamp his department, reallocate resources and risk his retirement over some new data. That engineer is likely to approach a VP up the ladder pointing to a sampling frame conundrum. The internal researcher finds it harder to sell our wares. Why? Research has lost its credibility.

The samples used for online research mostly come from online access panels. The access panels use a double opt-in process where respondents essentially agree twice to participate in research and then are offered a sequence of studies that they can choose to take. The panels grew quickly and became a big business practically overnight.

When the recession hit, prices were driven down. Some said that the market research industry was being "commoditized." It was only a few years ago when one could attend a market research conference and hear a positive message on the sampling frame issue and then attend an academic conference and listen as the sampling frame got hammered. Academics railed away at sampling frame error; market researchers, who made their living by it, were forced to defend it.

Most market research practitioners are not in a position to research the science behind the sampling frames. When RDD telephone sampling was prevalent, there was little to defend. As the recession pressured for lower prices, telephone studies

If market research is to retain its seat at the corporate decision-making table, it has to bring more than information.

were switched to online at an alarming rate. Researchers were walled in by client demands and competitive forces.

If market research is to retain its seat at the corporate decision-making table, it has to bring more than information. The results of our work have to provide direction that is actionable. We cannot, and must not, be on the defensive, embroiled in an argument about the quality of our research. The times are difficult. The perceived value of research is the casualty of our own reluctance to deal with quality issues. We must be the keeper of the sampling frame, or at least hold the intellectual gate to its use, if we are to avoid commoditization.

The Quality Crisis

While our credibility suffered, we offered sales talk in our defense. End buyers of research, such as Kim Dedeker of Procter and Gamble, protested openly. The industry resisted.

Sparks of leadership came from nontraditional corners like Bob Lederer, a journalist with a feisty newsletter. Mainstream researchers still held to the old model: Buy a sample, execute questionnaires, analyze data and try to deal with sampling frame issues.

Lederer came up with a novel idea: He would bring panels and research buyers into the same crowded arena and convene a dialogue. The clients grumbled, and the sample providers stonewalled. No dialogue was to be had. It was called the “road to the client congress,” but it was not going to be paved with progress.

One end user, Ron Gailey, the head of research for Washington Mutual, brought a valuable lesson. His was the tale of some 40,000 interviews, covering 29 research studies, conducted for the financial giant in 2006-2007. According to Gailey’s research, demand for WaMu’s financial products was on the wane; purchasing intent had dropped some 30 percent over the term yet the market place was not experiencing the same results. (See Gailey, Ron (2008), “The Quest for Sample Integrity: Exploring Panel Sample Characteristics That Impact Survey Results,” Presentation at The Research Industry Summit: Solutions that Deliver Quality, IIR, Chicago (November).)

For a researcher, bringing bad news to management is not good news. What could be worse? The answer: bringing the wrong news.

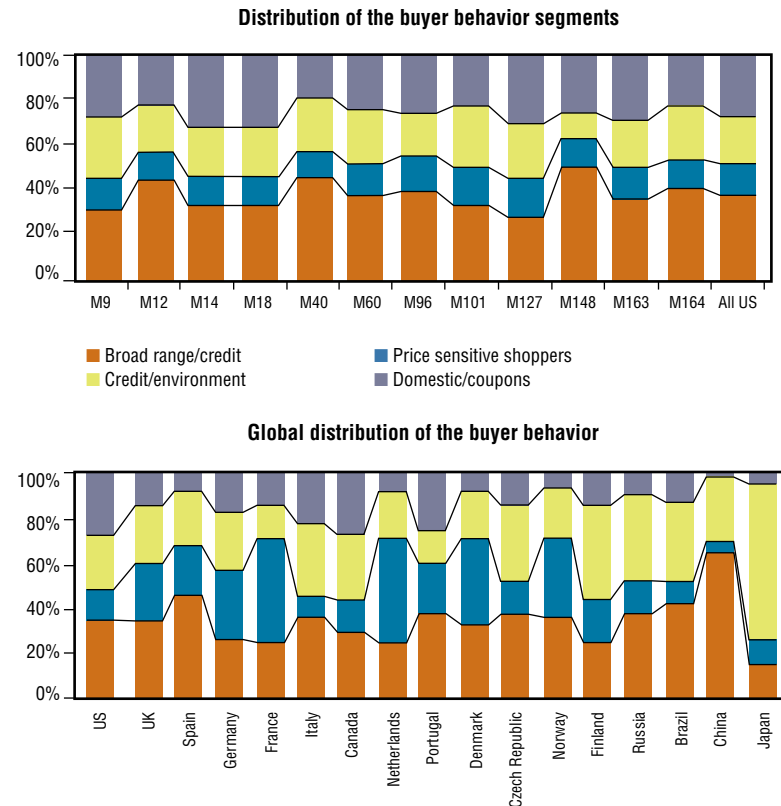
Apparently, Gailey was the victim of a change in his sampling frame. A variable that had not been previously identified as an important culprit—panel tenure—had crept into his data. As respondents served longer and longer terms on the panels that fed Gailey’s sample, they became more conservative. As Gailey pointed out in 2008, “In every study examined, people with more panel tenure gave lower demand.”

It was the train wreck that helped force the dialogue. Panel providers were slow to turn around, but a few tenacious buyers kept up the pressure. Lederer prepared to bring a list of suggestions formulated from his joint think tank to a gathering of end users. His intent was to create guidelines for conducting online research.

As researchers will do, the decision was to research the problem. A flurry of white papers collected on Web sites across the land. It was a time for clarity and transparency, but what we received were “proprietary solutions.” There was little transparency.

The years brought a numbing flood of data that resisted admission that there was a problem. The various associations wore themselves out, and eventually guidelines tumbled out of the Client Congress in Chicago. ESOMAR generated 26 questions that we all should be asking the sample providers, and the Advertising Research Foundation (ARF) launched a research study, purportedly with a one-million dollar market value, to bring clarity if not transparency.

The brand value equation



Research Findings

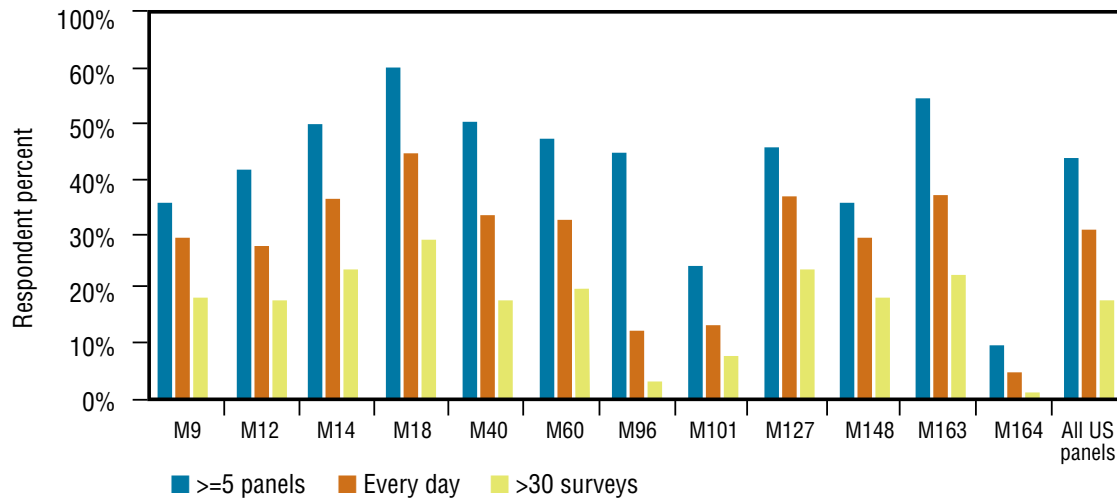
There were quite a few conferences in the past three years. The take-home message remains inconclusive. As always, the commercial side defended online research as a cost-effective means of obtaining needed data in a cost-conscious time. The research conducted by various associations became bogged down in panel demands for anonymity. The overriding conclusion that the panels were different and therefore not interchangeable was finally accepted.

While Ron Gailey’s work is considered seminal, the potential drivers that make respondents differ over time have been left in a cloud. However variable the panels appeared, no one seemed ready to admit that they do change. Much of the analysis was pegged on singular test variables, despite the fact that some measures are just hyper-stable. Other analysis focused only on the shifts that occur over short time periods.

To fix the problems of panel differences, panel blending has begun a slow rise as a means of mitigating potential bias. In the face of alternatives to the standard business model for research, the industry remains resistant to transparency although slow progress is being made.

In 2007, Mktg Inc. began a massive research-on-research endeavor called the Grand Mean Project. Resistance was strong throughout 2008 when data was being collected in the U.S. market. An analysis of 17 American panel companies

Hyperactive respondents by U.S. sample



was delivered at CASRO in January 2009. (See Gittelman, Steven and Elaine Trimarchi (2009), “On the Road to Clarity: Differences in Sample Sources,” CASRO Panel Conference, New Orleans.)

The panels were clearly different. Hyperactive respondents, tenure and sourcing models appeared to drive the differences. The panels were not only different demographically, but also behaviorally. There was a strong implication that the differences found between panels would drive differences within panels over time.

Data coming from the ARF contradicted Mktg. It concluded that, although the panels were different, there were no changes through time (waves were separated by only a few weeks). Further, while panel tenure was important, the elements that define hyperactivity were benign and in some cases actually provided better respondents. In essence, belonging to multiple panels and doing a multitude of surveys had a beneficial side. (See 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.)

Lederer and the panel of end users struggled to move forward. One, Steve Schwartz from Microsoft, was at his limit. For Schwartz and colleagues, the ability to sell research within the software giant was a difficult climb. He had his own internal “people” to deal with, and the research industry seemed to have no inclination to provide the information he was seeking. The research companies were simply saying, “Trust us.” But Schwartz and other end users were not willing to go along. To them, the sampling frame issue was rotting credibility, making the concept of “garbage in and garbage out” all the more real. Microsoft began to look for alternatives.

Panels Are Different

Between-panel differences in the United States are driven by differential sourcing, attrition and conditioning. Gailey’s findings are the key: A change over time can influence data and the decisions we make from it. If he is correct, then it should come as no surprise that there are respondent-related drivers that are the force for that change. There are large differences between buying behavior segmentations of American panels that stand as a cautionary tale for anyone planning to switch panels between waves of a tracking study or combine them without careful study. (See charts, page 23.)

The buying behavior differences are magnified when we look at the global scene. This is as one might expect because cultural differences appear to be an important driver. The variability is too great to allow commingling of sample without careful considerations for the blending approach being used.

We must understand the drivers of variability. Thus, in the past two years, attention was given to the impact of hyperactive respondents on three measures: respondent tenure, multipanel membership and frequency of survey participation.

When we measure the American panels, they differ on all three hyperactive measures. Global trends show how far afield the American data has become. The leading indicator of hyperactivity is multipanel membership. The impact of multipanel membership is reduced when it is not coupled with hyperactive survey taking. In the U.S., all measures are heightened well above the trends seen globally.

Perhaps the most critical difference is that panels change through time. (See chart, page 25.) In a new form of quality control measure, Mktg Inc. has been studying panels for consistency through time. The analysis includes more than 100 measures, most combined into various segmentations. Not all panels change through time, but on the battery of

measures studied all seem to change to some degree. In the media segmentation consisting of 31 input variables, we see a shift in the online components: Respondents rely increasingly on online sources to receive information. It is as one would expect.

The historical solution: Weighting has provided cover, but complex schemes heighten statistical error. Weighting against demography does little to protect against behavioral shifts resulting from conditioning effects and attrition. Yet it is used with abandon.

The use of nondemographic reference points and weighting against them provides additional false security. Again, the weighting process can be overdone, and the reference points do not protect against nonrepresentative behavioral shifts. The overuse of such techniques heightens the likelihood that we will fail to master the sampling frame conundrum, which will in turn further erode confidence in online research.

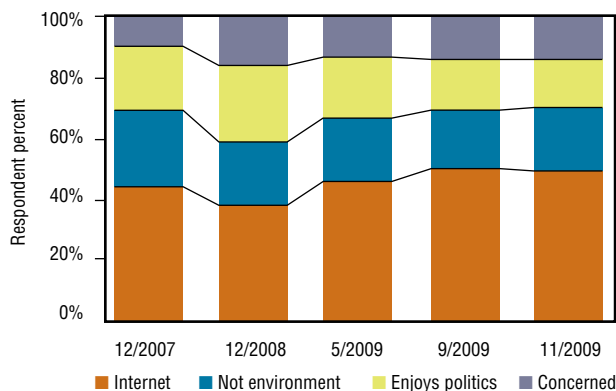
Measure Twice—Cut Once

A carpenter’s apprentice is taught this old axiom time and again. We as a profession measure behavioral trends for a business. It is our calling to interpret data on behalf of our clients. So why have we been slow to develop appropriate metrics to measure online quality?

The question of metrics is a hard one. After years of dedicated research-on-research into the drivers of online variability, an uneasy cloud hangs over the industry. Studies that say hyperactive respondents don’t matter seem self-serving and are a hard sell to anyone but the willing believer. Likewise, claims that hyperactive respondents behave better in their survey execution and provide more accurate results would seem to tell only part of the story.

Sadly, professing that the input variables have no impact appears conveniently self-serving and leaves us with no explanation for the variability that curses us. End clients use the output data to make decisions. If, as Gailey demonstrated, the data generated drives poor business decisions, then we had best surrender our chair at the corporate boardroom table now.

Variation over time for one panel on media



We need metrics that work to stabilize output. One possibility is a multivariate examination of buying behavior—social and media segmentations generated by structured samples of panel respondents. Call it a behavioral fingerprint of the panels. By looking at the average of such outputs, we can create a grand mean of the segmentations to guide us in comparisons. If panels regularly submitted samples for evaluation, they could be compared behaviorally; we would have the beginning of a new metric family.

Consistency is the driver of good samples. If a sampling frame meanders as a result of changes in sourcing and/or conditioning effects of panelists, the result is the same: We will be unable to separate changes in the underlying sample frame from real changes in our data. In essence, we will live with the Gailey syndrome. If the grand mean we spoke of earlier was rigorously maintained through time by repeatedly sampling the online sources in a country or region, we could create a time-sensitive metric that would be useful for detecting sample frame variability and defining normative data.

Pressure from Outside

If research practitioners remain incapable of grappling with sampling frame problems, then the users of research are justified in commoditizing the purchase of their services. Clearly, the collection of online data from one poorly designed sample is no better conceived than the collection from any other. If non-probabilistic sampling frames continue to grow unfettered by metrics, then the end users of research should not be criticized for their efforts to substitute less expensive alternatives.

For one, questionnaire design has been blamed as the culprit. Unable to fix the sampling problems, many researchers have blamed the obvious ills of poorly designed questionnaires for respondent fatigue. The only persons who can endure the grid-laden monsters that we have been offering are the truly tolerant. Those who are capable of sticking with a 40-minute questionnaire are clearly different from the rest of us who have no such patience. In fact, why bother with a questionnaire when alternative means of data collection seem equally productive? Online data mining makes the assumption that sampling frame is of no consequence because behaviors are measured and quantified. In fact, it has been postulated that a single question gets you most of the information gleaned from more complex designs.

Fortunately for the research industry, we are all creatures of habit. Like the respondents that we cajole into doing our questionnaires, we are wed to familiar methods. We are conditioned. Strangely, we reject the concept that our targets are equally conditioned. We can hope that our clients suffer similarly. The industry will survive but its professional respect is at stake. ●

Elaine Trimarchi is executive vice president of Mktg Inc. She may be reached at Elaine@mktginc.com. Steven Gittelman is president of Mktg Inc. He may be reached at steve@mktginc.com.