October 2009
Vol. 49 No. 10

MRA SM
CMOR imro
A Family of Services Now Under One Name

Alert!

Magazine

Data Quality

Consistency: the New Quality Concern

By Steven H. Gittelman, Ph.D. and Elaine Trimarchi, PRC

Consistency of online samples is a core issue for market researchers. After all, much of the value we provide is in the tracking studies we perform – even one-time studies – should relate to some reference and not float in a sea of variability between panels. If your data changes, it is essential to know if the changes are real or the inadvertent product of sample inconsistency.

In the past, we had no reason to fret over sample consistency. At the core of every research career, there is a fundamental reliance on probability. Toss a coin, any coin, and it will reliably come up half heads and half tails. There is no magic in it, but small samples are predictive of the results we find in larger samples. No one would expect to toss a coin a million times in order to prove its "fairness." A sample of some smaller number will do.

Market researchers have drawn samples from known commodities for decades, always relying on the fairness of a coin toss. Households could reliably be reached by telephone almost 99% of the time, and the small fraction of non-phone homes mattered little. Yes, we had to adhere to strict calling regimens, callbacks, refusal conversions and most of all recovering a large percentage of the sample. The key here is the telephone sample replicated the Census because it reached most segments of the population with equal penetration, it earned its name "probability sample." It was reliable, predictable and repeatable, and thus consistent.

As refusal rates began to climb, and "do not call lists" became good politics, the ability to reach some segments of the population dwindled. The all-sacred concept of a high recovery rate began to crumble. Phone, with an increasing percentage of line cutters and cell phone users, has begun to dwindle, as its ability to replicate the Census was impaired. And why does the Census matter? We anchored ourselves to the Census. If demographically we matched the count that the founding fathers demanded by law there was a built in reality check. Where once research was well grounded in a probabilistic framework supported by an underlying Census of the population, online market research has moved into a new era, from a probabilistic framework to "working without a net." In the absence of a probabilistic net to anchor samples, they can drift without our knowing.

"Market researchers have drawn samples from known commodities for decades, always relying on the fairness of a coin toss."

One, now historic, example of this happening was presented by Ron Gailey, now of Coca Cola, but previously of Washington Mutual, who disclosed how 29 studies representing 40,000 online interviews had gone astray due to panel inconsistency. In the WaMu research, the change was due to shifts in respondent tenure that resulted from changes in the panel's constituents over the two-year span of the base research. Gailey's research showed a 30% drop in buyer demand for WaMu's financial products; a result (2006-7) unsupported by sales. His conclusion, after much study, was long-term panel members were less optimistic about their purchases than new panel members. Others have since corroborated these finding. The lingering question, now that WaMu is gone, is how the tainted research impacted on critical business decisions.

The affects of hyperactive respondents and other online respondent ills were brushed under the rug. Gailey unknowingly had to use a sample that showed aging affects that took time to evolve. Gailey had to do a lot of digging to find the root problem within his data. If online samples changed as they aged, then they could not be counted on to provide reliable data through time. And there are a host of reasons that could change them. For example, mergers bring together samples of different sourcing and aging profiles. Management makes decisions influencing the frequency of hyperactive respondents by increasing the number of surveys that they are invited to and allowed to complete. It is evident that a panel that is used many times a month is different from one that has new respondents all the time. Panels differ for a wide variety of reasons, many of which are not disclosed to clients since

(Continued on page 20)

(Continued from page 19)

currently there are no standards.

Luckily, there is a world out there of science that has long ago learned to collect data and make decisions based upon sampling frames that are non-probabilistic. When Charles Darwin hauled himself onto the volcanic shores of the Galapagos Islands, he took samples of as many islands as he could reach. For the most part, these isolated little islands were different from one another. Even birds that could theoretically fly from one to the next differed. He didn't have a Census to draw his conclusions: He was the Census!

Charles took samples of a few islands and wrote a pretty good book. The samples were not grounded in probability theory, and he could not generalize from island to island. It was the differences that gave him clues. Each island was an ecosystem unto itself and the differences that species on the islands had to endure shaped them into the specialists that they became.

The hard sciences cannot afford the luxury of probability theory because usually there is no Census to hang a hat on. It is the body of knowledge that is the Census. A bit tougher, but it works.

Our use of online data has much to learn from island biogeography. Think of each online panel as an island. They have similarities, but are drawn from different sources. We should not expect them to be identical, we should expect them to be different. Our research has shown them to be quite inconsistent. The panels are not interchangeable. The online panels are drawn from different sources, are subject to differing management practices, and for a host of reasons yield different results.

> "Consistency is a complex concept. We need to know the differences between panels at any given moment so that we can use multiple panels in our research."

Hidden in all of this is the concept of consistency. After all, if we measure bias and can't anticipate its shifts over time, then we will not understand which changes are coming from our data or from background noise in the sample. The issue of consistency is the most important area of concern. We must learn to measure not only what the constituent elements of our data sources are, but also how they change in time. In other words, we have to enter a new world of consistency analysis.

As we do our research, we must know what the changes in our data mean. Are they the product of shifts in opinions or changes in the sampling frame? To get a grip on this we need parallel studies that document the consistency of our samples.

Consistency is a complex concept. We need to know the differences between panels at any given moment so that we can use multiple panels in our research. And if we must change panels, we must know as much as possible about the old and the new. Moreover, even if we stick with the same panel we must know how it changes through time and events. And, lest we forget, blending samples is a good way of spreading risk among many to avoid the potential ills of just one. Be prepared for the use of sample blending techniques to become an industry standard for achieving consistency; in the United States. We have developed almost a thousand blends with about a fourth of them within 1% RMS error from the Grand Mean.

We have moved onward from our initial study of the American markets and have expanded its research to include 140 panels in 35 nations. In each, a standard instrument is used in a tracking study that includes a diversity of measures but mostly focuses on buying behavior segmentations. By conducting repeat waves of this consistency study, a local Grand Mean is calculated for each market. In addition, using standard quality control techniques an analysis of the consistency of each panel is conducted.

The Grand Mean is an aggregate statistic. It is a measurement of consistency that should be reliable and yield a sense of predictable change. No panel represents the universe as well as the Census, but the sum of many panels represents it better than any single panel alone. Think of the Grand Mean as a group of indices that are measured from the sample of each panel over and over again: tracking panel quality through time.

When we relied on the Census we

assumed that if we matched its numbers through quotas we would be within the realm of reality. We knew that the Census was a once in a decade count and had to weight for some changes. The further we drifted from the Census, the weaker our reliability became. We could not anticipate seasonal changes from the Census, we had to calibrate for that. It is quite predictable that buying behavior will shift on a seasonal basis. It is also predictable that the ice will melt in spring and that the rains will come. We bet on it all the time. They say that the only thing that you can count on is death and taxes – wrong. Predictability is as much a part of consistency as is reliability. Consistency does not mean staying the same, but rather having predictable patterns of change. The Census did not provide us with that.

If all panels were required by their users to show that they were consistent, we would have both a measure of quality in their samples, as well as a new set of indices to replace the Census. Certainly, if all panels provided data on how their members responded to a battery of purchasing questions and the segmentations were tracked, Gailey would have had a reference to consider when his panel's composition began to change. Right now there is no such reference. The Census has very little relationship to the online community, and even less to the online community that participate in panels. The offline population is different from those online, and the panels themselves are a disparate group with almost no guiding standards. The whole process of belonging to a panel filters out an unknown sector of the population and no one knows how to weight them or what problematic and unweightable variables might be hidden in the data.

We propose that all panels test regularly for consistency: each would participate in at least four waves of audit per year. This would provide End Users with assurances regarding the stability of panel output. Their combined consistency data could be compiled to generate a Grand Mean within a market as a new metric to anchor panels and tracking studies alike. The composite data would provide rich insights into shifting in the sample universe and inconsistencies within individual panels. If needed, source blending using optimization modeling could serve to correct drift. The Grand Mean metric would itself be anchored to a battery of outside benchmarks. Consistent bias could be monitored against the metric.

Each panel company would participate in one multi-wave consistency audit. The company could distribute the auditor's reports to all of its clients. We are suggesting that End User companies require a standardized form of consistency analysis from all of their sample source providers.

Disclaimer: This article does not reflect the views or opinions of the Marketing Research Association.

Steven H. Gittelman, Ph.D. is the president of Mktg, Inc. and **Elaine Trimarchi, PRC** is the executive vice president of Mktg, Inc.