

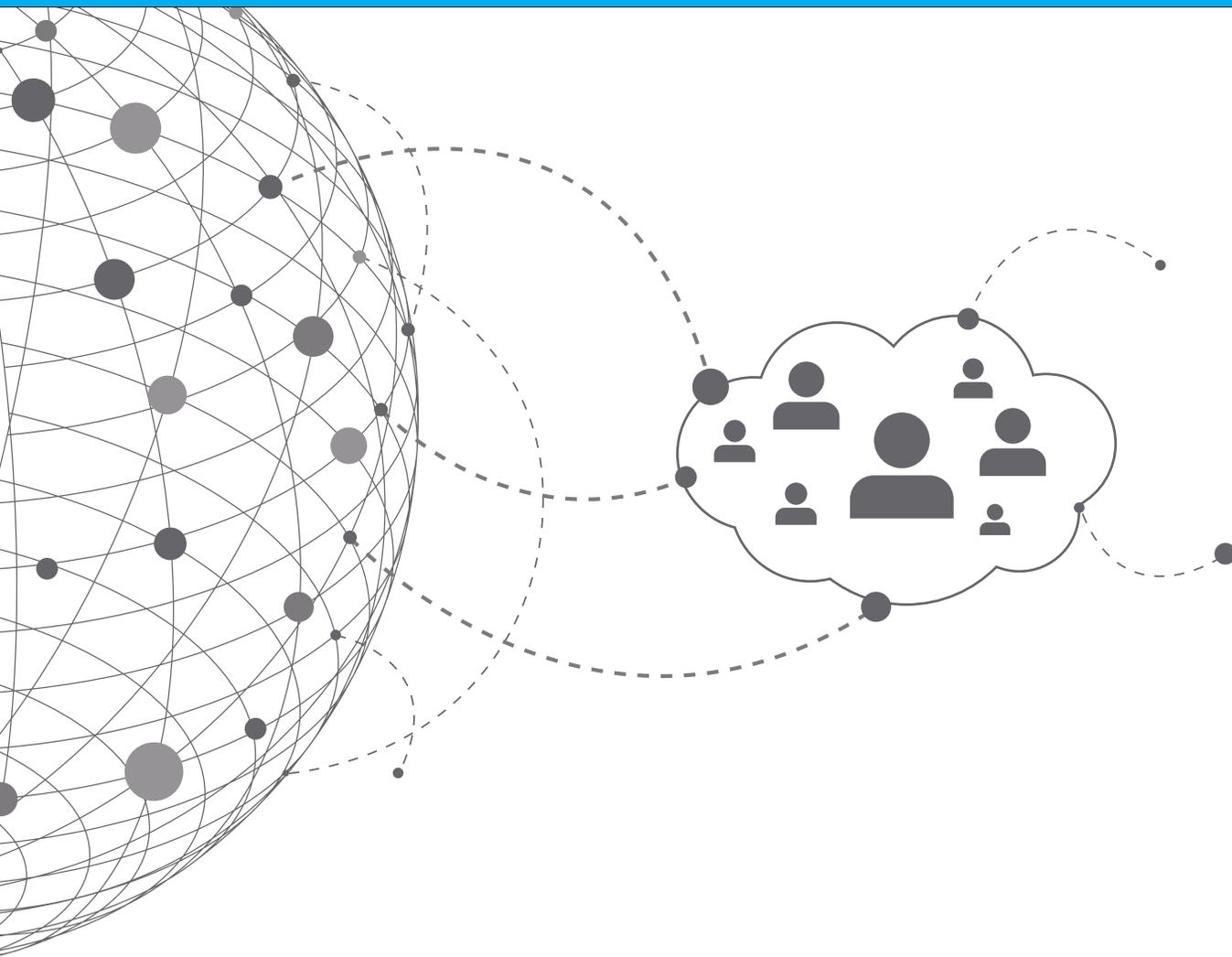


NIELSEN JOURNAL
of MEASUREMENT

JULY 2016-VOL 1 ISSUE 1

SURVEYS IN A BIG DATA WORLD

*By Michael Link, President & CEO, Abt SRBI
Jeff Scagnelli, Manager, Data Science, Buy Behavioral Methods, Nielsen*



EDITOR-IN-CHIEF
SAUL ROSENBERG

MANAGING EDITOR
JEROME SAMSON

REVIEW BOARD

PAUL DONATO
EVP, Chief Research Officer
Watch R&D

MAINAK MAZUMDAR
EVP, Chief Research Officer
Watch Data Science

FRANK PIOTROWSKI
EVP, Chief Research Officer
Buy Data Science

ARUN RAMASWAMY
Chief Engineer

ERIC SOLOMON
SVP, Product Leadership

The world of measurement is changing.

Thanks to recent advances in data collection, transfer, storage and analysis, there's never been more data available to research organizations. But 'Big Data' does not guarantee good data, and robust research methodologies are more important than ever.

Measurement Science is at the heart of what we do. Behind every piece of data at Nielsen, behind every insight, there's a world of scientific methods and techniques in constant development. And we're constantly cooperating on ground-breaking initiatives with other scientists and thought-leaders in the industry. All of this work happens under the hood, but it's not any less important. In fact, it's absolutely fundamental in ensuring that the data our clients receive from us is of the utmost quality.

These developments are very exciting to us, and we created the Nielsen Journal of Measurement to share them with you. This paper is part of VOL1 ISSUE 1 of the Journal.

WELCOME TO THE NIELSEN JOURNAL OF MEASUREMENT

SAUL ROSENBERG

The Nielsen Journal of Measurement will explore the following topic areas in 2016:



BIG DATA - Articles in this topic area will explore ways in which Big Data may be used to improve research methods and further our understanding of consumer behavior.



SURVEYS - Surveys are everywhere these days, but unfortunately science is often an afterthought. Articles in this area highlight how survey research continues to evolve to answer today's demands.



NEUROSCIENCE - We now have reliable tools to monitor a consumer's neurological and emotional response to a marketing stimulus. Articles in this area keep you abreast of new developments in this rapidly evolving field.



ANALYTICS - Analytics are part of every business decision today, and data science is a rich field of exploration and development. Articles in this area showcase new data analysis techniques for measurement.



PANELS - Panels are the backbone of syndicated measurement solutions around the world today. Articles in this area pertain to all aspects of panel design, management and performance monitoring.



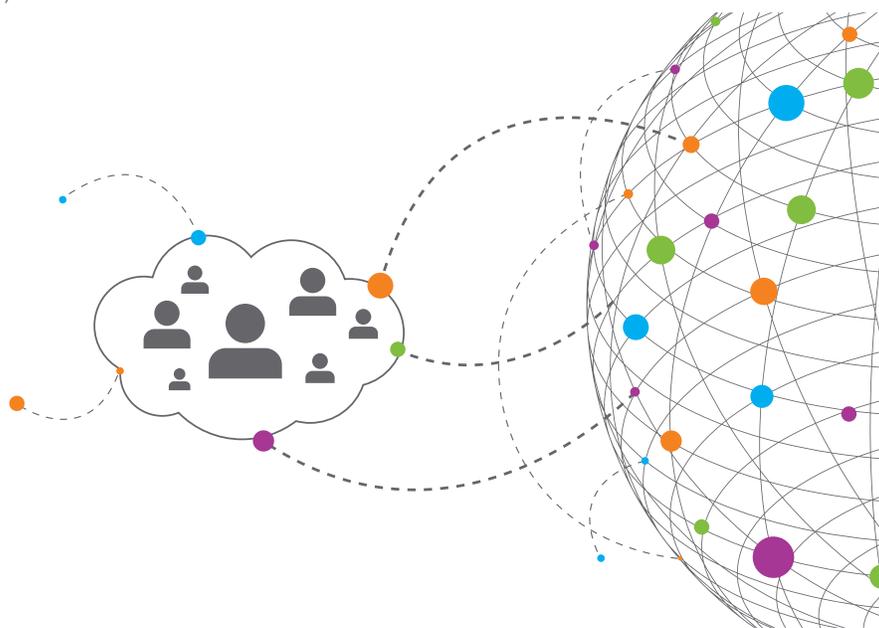
TECHNOLOGY - New technology is created every day, and some of it is so groundbreaking that it can fundamentally transform our behavior. Articles in this area explore the measurement implications of those new technologies.

SURVEYS IN A BIG DATA WORLD

By **MICHAEL LINK** *President & CEO, Abt SRBI*

JEFF SCAGNELLI *Manager, Data Science, Buy Behavioral Methods, Nielsen*

S



OVERVIEW

“Surveys are dead! Long live surveys!” We’re now living in the “Big Data” era – a world of voluminous, high velocity, and increasingly varied data sources. Surveys have been the workhorse of market research for nearly a century, but long lead times, small sample sizes, declining participation, and rising costs are making it far more difficult to conduct good surveys today than in the past.

Recent headlines from political polling fiascos in the U.S. and abroad have added fuel to the fire, raising serious doubts about whether accurate information can—and should—still be collected through interviews with a sample of the population. The reality, however, is far more complex. While

alternative information-collection techniques (based on data from social media, global-positioning systems, stock-keeping units, sensors and the like) are being used today to supplant surveys in many cases, they cannot tell us “why” the same way as simply asking the question of a respondent.

For most (if not all) clients, “why” still matters. Fortunately, there are a number of solutions to the current challenges.

SURVEYS IN FLUX

The ways in which public opinions, attitudes and behaviors are formed, expressed, conceptualized and measured are now more diverse than ever. At the same time, the need to monitor and understand consumer behaviors and preferences has risen dramatically in recent years.

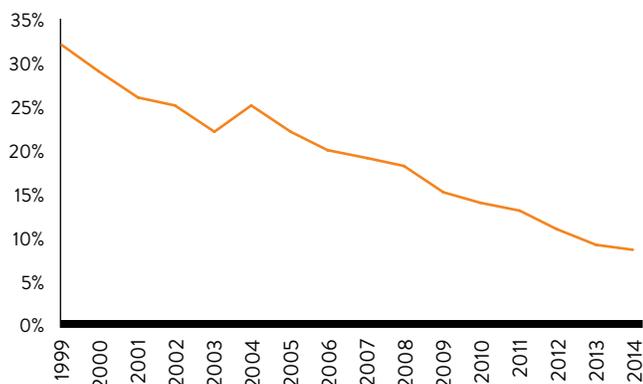
Businesses, government agencies, academics and others make decisions and create policies based on knowledge of populations and markets that has traditionally come from surveys. For example, surveys are still relied on for measurement of the unemployment rate, the inflation rate, and many other national indicators used widely in America today.

While the need for high quality data on consumer attitudes and behaviors is great and rising, the challenges of continuing to collect this information via traditional surveys are substantial. In the face of these challenges, many users of survey data have grown skeptical of contemporary survey methods, leaving survey research at a crossroads.

The problem is driven by societal changes and technological advancements. Telephone surveys, the decades-long mainstay of market research, now face a simple problem: people refuse to answer their phones. Telephone researchers also now face greater suspicion of solicitations from strangers. The decline of landlines, prohibitions on autodialing cellphones and the proliferation of caller identification systems all contribute to weaker response (see figure below).

DECLINE IN SURVEY RESPONSE RATE (1999 – 2014)

(Average of CBS/NYT/ABC/WP/Pew Polls)



Source: David Dutwin (Jan 28, 2016). "Political Polling Isn't Dead Just Yet." *The Washington Post*, *The Fix Blog*

Moreover, society has moved from a more altruistic view of information exchange ("I'll provide you, the researcher, a response on how I feel about X, so that my voice can be heard") to one that is decidedly more transactional ("I'll provide my data for a price"). That transaction may be monetary (a paid incentive for taking a survey) or an exchange for "like goods" as in the case of social media or mobile applications, where participants agree to give companies access to their personal information when they use the platform to research products, communicate or share their thoughts and pictures with others.

Online surveys face their own set of challenges with regard to response; only a small fraction of those invited to participate in surveys (often via banner ads or other similar mechanisms) do so. This shift in public attitudes and changes in communication habits have produced an environment where good surveys are difficult to conduct and results are increasingly called into question.

Political polls are perhaps the most visible arena where survey effectiveness has been called into question. The failings of Mitt Romney's presidential campaign in 2012 demonstrated just how easily political "pollsters" can miss an election's result, as both the candidate and his team believed—based on their polling numbers—that they would win the election, right up until election eve. There have been numerous other instances in recent years where polls got it wrong, sometimes determining the correct winner in a race but by substantially different margins than the actual election outcome. In other instances, surveys have gotten the winner of the race wrong altogether. And this issue is not confined to the U.S., but is, rather, a global one. In 2015 alone, political pollsters failed to make the correct election calls in the U.K., Argentina, Poland and Israel.

Granted, calling election outcomes from polls does have some unique properties not often shared with other forms of surveying or market research—such as the need to accurately model the election turnout when making final predictions. Yet political polls have often served as a bellwether for the survey industry, because they are one of the few forms of polling where reality or true public preference is actually known – the vote outcome. So while political polls are unique in some respects, their failings, while occasional, are very visible and tend to call into question other forms of survey research, such as market research and the determination of consumer preferences.

SURVEYS ARE CRITICAL TO DELIVER INSIGHTS

Nevertheless, data from surveys is still necessary for truly understanding consumer attitudes and behaviors. Many inroads have been made in recent years in the use of big data sources, such as identifying patterns in credit card transactions, loyalty card data, Twitter posts, cell phone GPS coordinates, and the like. Yet while these approaches can help us understand what consumers are doing and to some degree who is engaged in these behaviors, they often come up short in determining the true reasons behind consumers' actions. And understanding "why" is still very relevant for companies making critical decisions regarding what new goods and services to offer, to whom they should market these services, how and where to advertise, and a whole host of other vital business decisions. Inferences can be used to develop a set of insights, but relying on indirect data can be risky when you are basing financial decisions on it.

For this purpose, Nielsen utilizes surveys on a number of fronts (see table below for a list of some of our main survey

series). Ongoing surveys are used, for instance, to better understand consumers' attitudes and behaviors with regards to advertising content, purchase decisions, technology use, financial and spending habits, and a range of other issues.

Granted, there are some circumstances where knowing "what" in a quick and granular manner may be enough for some subset of business decisions. For instance, to discover problem areas that need to be addressed to restore service, or to detect financial fraud, or to respond quickly to increased demand. However, there is often little substitute for understanding why consumers truly hold the attitudes and take the actions they do. Neuroscience and biometrics are new techniques that are becoming more widely used to help answer "why" questions, particularly for specialized marketing communications such as video ads, new package designs and multi-platform testing. However, such solutions are confined for the moment to a relatively narrow set of questions compared to the broad universe of applications for surveys. Surveys remain, therefore, the critical tool for gaining insights into "why."

But they face many challenges.

SURVEY SERIES	DESCRIPTION
NIelsen SCARBOROUGH	More than 200,000 surveys are conducted per year in over 120 local areas across the United States. Information is gathered through a combination of phone and mail surveys about radio listening, newspaper and magazine reading, television viewing, leisure activities and shopping habits
GLOBAL SURVEY OF CONSUMER CONFIDENCE AND SPENDING INTENTIONS	Online surveys are conducted with more than 30,000 consumers in 60+ countries throughout Asia-Pacific, Europe, Latin America, the Middle East/Africa and North America
NIelsen MOBILE INSIGHTS	A monthly survey of 30,000+ mobile subscribers aged 13+ in the U.S. Data is collected primarily online and also via a supplemental Spanish-language phone survey seeking to reach Hispanic respondents aged 18+
NIelsen INSURANCE TRACK	An online survey of approximately 35,000 adults is conducted biennially, collecting consumer-level data on behaviors related to auto, residential, life, and other insurance coverage types
NIelsen TECHNOLOGY BEHAVIOR TRACK	An online survey is conducted annually with approximately 32,000 adults, collecting consumer-level data on behaviors related to home technology use
NIelsen NATIONAL PEOPLE METER SERVICE	A combination of face-to-face, telephone, and online surveys is used to screen eligible homes for inclusion in these measurement services and to collect demographic and consumer and media behavior information throughout the tenure of participating panel homes

UNDERSTANDING THE CHALLENGES

Surveys face bias and accuracy issues from a number of sources, such as the manner in which a questionnaire is constructed (measurement error); the way in which it is administered: by telephone, in-person, online, mobile device, etc. (mode bias); and the amount and type of incentives used, if any. However, the key area of contention among survey practitioners today—and often the source of most error—can be traced to the manner in which people are identified to participate in a survey. That is, the way that respondents are sampled and the success of that approach in providing sufficient coverage or representation of the broader population of interest (say, e-commerce customers within a particular retail segment, television viewers within a particular local market, or individuals who visited a particular set of websites).

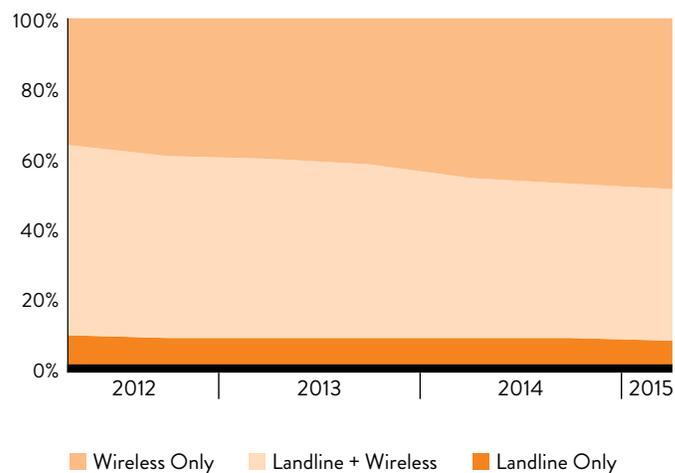
Representation

Surveys produce results that can be considered “representative” when they can assess and account for deviations in responses to a question (or set of questions) between the sample gathered and the population of interest. When survey participants differ systematically from the broader set of individuals they were selected to represent, “bias” is likely to occur in the survey estimates produced. This can result in misleading or fallacious conclusions about what consumers are actually thinking or how they are behaving. For example, a recent postmortem of the failure of pollsters to correctly call the 2015 general election in the UK traced the main problem to a failure to reach particular types of people – in this instance, the more conservative voters in the electorate (who tended to side with the Tories).

Coverage

Every survey has some target population or group of people (e.g., the general population, likely voters, members of an association, customers) that it is meant to represent. In an ideal world, all or nearly all of this “target population” would be identifiable and accessible. In other words, there would be a comprehensive list of all target population members available. But rarely is this the case, and in many instances some lists relied upon in the past are not what they once were. For example, as shown in the following figure, the landline telephone frame no longer contains all or nearly all households. Other theoretical frames, such as all or nearly all individuals with email addresses, do not exist in practice (at least not for general populations).

RECENT CHANGES IN HOUSEHOLD TELEPHONE STATUS (USA, 2012 – 2015)



Moreover, non-response is often not random. For example, the U.S. Census Bureau finds that single-person households have a much higher “not at home” rate—and therefore a lower response rate—than multi-person households. This type of systematic non-response could skew sample data and lead to under-representation of certain groups unless special efforts are made to include these respondents.

To reduce the chance of bias in these instances, researchers have often developed new methodologies. For example, maximizing population coverage when doing telephone research now requires a dual frame approach that includes both a landline and a cell phone frame. Nearly 10 years ago, Nielsen transitioned one of its services from an approach that selected households using a random selection of household landlines to one based on a random sampling of residential addresses. Given the restrictions on using automatic dialers to call known cell phone numbers, a dual sample (landline + cell phone) approach was simply cost prohibitive. The move provided a solution to the coverage issue facing the diary at the time, and ensured the inclusion of all—or nearly all—households in each Designated Market Area.

Source: Blumberg SJ, Luke JV. Wireless substitution, National Center for Health Statistics.

Challenges specific to online surveys

Online research has developed as a means of countering the cost of telephone and in-person approaches, but it faces its own challenges. An obvious issue is that people who are not online have no chance of being included in the sample. Beyond this, online survey and panel members typically do not come from a known list or sampling frame, but rather opt in to the survey effort by clicking on banner ads or by being part of large ongoing online survey panels.

Unfortunately, this often makes it difficult to evaluate the quality of the survey process and its outcomes. For example, understanding how online survey participants differ from target populations of interest, such as the general population or all online consumers, and whether they can represent those populations accurately, requires empirical evidence, deep analysis, and good judgment. Other lists, such as customers who bought a specific product or registered for a loyalty program, may be less than complete or otherwise imperfect as well. There may be gaps, duplicates and even errors. For its online panel surveys, Nielsen often utilizes multiple vendors that use different strategies to recruit potential panelists, in an attempt to have as wide and representative a pool of online survey respondents as possible.

Low participation

In addition to issues of coverage, surveys often suffer from lack of participation among those who are asked to participate. To the degree that those individuals who are contacted and agree to participate in a survey differ in their attitudes and behaviors from those who do not, the resulting survey estimates may be misleading or incorrect (i.e., “biased”). While declining survey participation has occurred generally across people of all walks of life, it is particularly notable among younger and racially diverse populations. They are among the hardest to reach and most difficult to convince to complete a survey.

Researchers often go to extraordinary lengths these days in an attempt to improve participation in surveys. Over the past decade, Nielsen has conducted considerable research to identify and refine methods for improving participation in survey samples: using multiple contacting attempts, offering different ways or modes of completing the survey (i.e., telephone, online, mail, interactive voice response, etc.), and offering various forms of incentives, both monetary and non-monetary. While these efforts have largely slowed the decline in participation, they have not halted the trend – a trend seen across the market research industry.

ADDRESSING TODAY’S CHALLENGES

To address these challenges, researchers can adopt three main strategies:

1. **Continue to invest additional resources in the current survey effort to ensure wider population coverage and improve response.**
2. **Explore the impact of switching to new data collection techniques, such as the use of social media data, to augment or perhaps replace current surveys.**
3. **Develop new analytic and modeling techniques, which could enable stable inferences from samples with sizable nonresponse and coverage errors.**

Most market research firms are still relying on the first strategy for a portion of their business (Nielsen included). Some of their surveys might have a long track-record—possibly going back ten, twenty, maybe 50 years—and they don’t want to jeopardize the reliability of the data by switching to a new data collection technique. The costs alone, which rise annually, make such a strategy difficult to sustain, but if the demand still exists and their clients are prepared to pay a premium in exchange for a stable data collection effort over time, it definitely remains a viable option.

For new surveys, or for surveys that are less sensitive to a switch to a different data collection method, there are new methodologies that show promise and are far less costly than traditional surveys. For instance, online search results, social media data and other modern consumer-based big-data datasets are relatively accessible and easy to mine for

insights, and a number of studies have found correspondence between survey data and data scraped from the web. However, predictions using these datasets have not always been successful.

It's important to recognize that big data has weaknesses on a number of fronts, including coverage issues and lack of respondent attributes. Social media and other forms of big data often do not cover well-defined populations. In fact, it is often very difficult to determine what types of people and/or geographies are covered by particular data sources. In addition, big data is often rich on a very small set of variables or areas of interest, but lacking in any demographic correlates or other variables of interest. This makes linking big data results to specific populations or tying trends across various areas of interest difficult, if not impossible.

The best thing to do with these types of datasets is to approach them with scientific rigor, and not hesitate to replicate their findings with more traditional methods, so that the two may be compared and their pitfalls properly identified. It may be possible to fully transition a survey from one collection technique to the other—for instance, from the telephone to Facebook—but such a transition needs to be handled slowly and methodically. Depending on the project, the solution may be to rely on big data but maintain a skeleton traditional survey to help calibrate the data.

It is interesting to note that some of the largest online and social media platforms today, such as Facebook and Google, haven't entirely turned their backs on traditional surveys: They rely on surveys of their users (along with a myriad of other internal and external measures) to better understand attitudes towards new features, user experience, advertising content and the like. These surveys tend to be of a different character than more traditional approaches (often featuring a very small set of questions across large numbers of different people rather than lengthier, more in-depth surveys of a smaller subset), but they do highlight the continued importance of surveys even in the big data arena.

To ensure accuracy of survey data within cost constraints, market researchers are increasingly relying on sophisticated statistical modeling techniques to produce reliable survey estimates. Sample matching is a good example: It is often more cost-effective to combine the information found in separate sample surveys than to recruit a new sample and field an entirely new survey. With careful consideration given

to the compatibility of sample designs across those distinct surveys, to the selection of proper matching variables, and to the weighting schemes used to harmonize the datasets and facilitate the matching process, it is often possible to impute data from one sample into the other. The combined—or matched—sample can then be used to infer new data relationships. With advanced statistical modeling techniques like sample matching, the initial quality of the starting sample is still a consideration, but far less so. Instead, researchers rely on statistical techniques to adjust the data after it is collected, to account for the sources of error in the initial sampling and data collection processes.

Granted, such adjustments have been made since scientific surveys began more than 80 years ago. The difference today is that these techniques are often more sophisticated, using additional data from outside the interview data itself, and are relied upon to correct for a much larger proportion of potential error than ever before. Such approaches also require new or at least more rigorous ways of validating the model assumptions and population inferences that are produced.

All serious survey research companies are now spending a great deal of time and resources to develop such models and make sure that they're tested scientifically and properly cross-validated – and that they can be adapted over time to account for changes in market conditions. This is an evolving area of statistical science. It requires a different way of viewing quality control in survey research, but it will quickly become essential to extend the life of surveys into the foreseeable future.

CONCLUSION

In today's big data world we can capture consumers' impressions, clicks, biometrics, GPS coordinates, demographics, cookies, transactions and even refrigerator temperature, all to identify relevant patterns of behavior upon which to drive insights and strategies. Knowing what consumers are doing is critical; however, in the end, nothing beats knowing why consumers make the choices they do. The most reliable route to understand their behavior is still to ask them directly. Conducting surveys in today's world is more challenging than ever before, but thankfully the industry is hard at work to develop the tools necessary to confront those challenges. Surveys still have an important role to play in consumer research. **n**

nielsen
.....