

*Mitch Sanders, Ph.D. is research director, analytics at Gongos Research*

*These days, online data for market research is widely available and easily accessible. But scrutiny is catching up with ubiquity, and some in the industry have expressed concern that problems are lurking in our datasets. This focus on data quality is well worthwhile – caveat emptor is as relevant as ever in the Internet age, especially when important business decisions are riding on the results. Fortunately, some good news is emerging from these inquiries. While possible threats to the quality of online data must be taken seriously, many conditions are actually benign, while others can be mitigated with modest effort. For researchers seeking guidance and reassurance, here are the answers to five important questions about data quality.*



#### **What is data quality, and why is it important?**

As with any product, online data “quality” has many facets, and different meaning for different users. But most people involved with quantitative market research likely agree on a core definition: high-quality data allows researchers to get the right answers from the right people. It’s important to keep in mind that there’s no value judgment attached to this definition. Results should be accurate reflections of respondents’ attitudes, preferences and behaviors – regardless of whether the study’s findings are encouraging or alarming. And those who respond to a particular survey should be suitable representatives of the larger population of interest, whatever that may be. If both of these conditions are met, analysts and decision-makers can be confident in the validity of a study’s results.

#### **What does it take to get the right people?**

Online access panels, which establish enduring relationships with willing participants and invite them to participate in online surveys as needed, have become the field’s standard mode of data collection. The fact that these relationships aren’t exclusive (i.e., a single person can join multiple panels) conjures visions of a small corps of professional respondents who dominate the universe of panels and repeatedly infiltrate the same surveys. However, recent research shows that these fears are unfounded. A study by the Advertising Research Foundation (ARF) found that only 16% of online panelists belong to multiple panels. More importantly, multiple panel membership did not have any noticeable impact on responses to survey questions. And in studies involving combined samples from more than one panel, ARF research revealed that less than 5% of responses were duplicates – too few to have a meaningful effect. If greater reassurance is needed, unique identifiers like postal addresses or digital “fingerprinting” can be employed to further reduce the likelihood of duplication. Arguably a larger question, and one that is relevant for any mode of data collection, is whether a given sample contains a suitable mix of people. The Internet has achieved widespread penetration into nearly all sectors of society, but we may not find perfect microcosms among those who are online, join online panels, or respond to surveys. These factors may or may not be relevant for any particular study – their relevance for results, if any, will depend on the relationship between respondents’ traits and their survey responses. Still, considerations of representativeness should encourage the adoption of meaningful standards, and, when necessary, trigger the use of tools like sample stratification and post-field weighting.

#### **What can be done to mitigate the impact of inattentive responses?**

Inattentiveness is an issue because online surveys typically offer incentives for participation, in the form of points (or a synonymous equivalent) that can be accumulated and redeemed for items of value. The amounts involved are small – even if one could take surveys for a living, the going rate would barely exceed minimum wage. The idea is that surveys should reward people for participation, without overwhelming their intrinsic motivation to share their opinions. Still, some rogue respondents endeavor to “earn” incentives as quickly as possible. By clicking without really thinking, they introduce random noise, or possibly bias, into results. Fortunately, they usually leave a trail, and can be spotted based on unreasonably short completion times, patterned responses, implausible open-ended answers, susceptibility to “red herring” questions or incorrect responses to data quality “traps.” Offenders can be dropped from a dataset, and repeat offenders purged from a panel.

#### **How can surveys elicit thoughtful responses?**

A few bad apples aside, online panelists are a well-meaning bunch, fully intending to provide honest and accurate responses to our queries. To make the most of this highly useful resource, surveys should be designed with the goal of helping respondents to help us. Empathy is essential, because for consumers (unlike researchers) any one product, or any one subject, is just a small part of their daily lives. The most important design issue is survey length – recent research by the ARF found that longer surveys dramatically increase inattentiveness, thereby weakening what we can learn from the results. But it takes more than brevity to make a good survey. Unclear, uninteresting or repetitive tasks will cause respondents’ attention to wane – grid after grid won’t do. The best way to engage respondents’ attention is to create engaging surveys, because even in market research, time flies when you’re having fun. Enjoyable as the experience may be, survey researchers should respect the value of the respondent’s time. Even the most dedicated panelists will be justifiably deterred by low incentives. Similarly, it’s important to avoid bombarding the same respondents with surveys, and potentially compromising data quality by dividing their attention.

#### **What new developments or new technologies may impact data quality?**

There is one constant in the field of survey research: change. And even though the future won’t truly be known until it happens, there are a few developments and technologies on the horizon that will surely have implications for data quality. Online communities offer respondents interactive and self-guided experiences, while serving as a useful platform for survey research. Greater engagement promotes data quality, by attracting and encouraging higher participation and attentiveness. These benefits do require a significant level of investment by community creators, in the form of careful cultivation and active management, which may deter widespread adoption. River Sampling takes an opposite approach, soliciting participation in surveys from the flow of visitors to numerous third-party Web sites, without requiring or even necessarily requesting that the respondent join a panel. Broader participation is potentially a major benefit; the question becomes how to use casual sampling to control the flow of respondents into surveys, especially for specifically-defined or low-incidence populations. Social networking offers a new paradigm for online behavior. The old-fashioned “e” Internet of e-mail and e-commerce is quickly being replaced by a new “i” Internet devoted to individualized content in the form of blogs, and social experiences such as Facebook and Twitter. These new online venues also offer potentially new ways to conduct market research. The challenge will be to find the harmonious patterns amid the cacophony. However the field of survey research develops, issues concerning data quality will, and should, keep the industry on its collective toes. With thoughtful investigation and wise investment, online research will continue to do what good research has always done – get the right answers from the right people. 