

Assessing the Quality of Data from Online Panels: Moving Forward with Confidence

Written By:

Renee Smith, Ph.D.

Senior Research Scientist
Research and Methodology
Harris Interactive

Holland Hofma Brown

Vice President, Internet Panel Management
Harris Interactive

Introduction

Clients using online panels for Internet-based market research must have confidence that they are obtaining high quality, unbiased results from the data collected by their research partners. Attention has recently focused on whether “professional respondents” exist and threaten the validity and reliability of results from Internet-based market research.

For example, a widely known 2004 press release from comScore® Networks suggested that *“more than 30% of all online surveys are completed by less than 0.25% of the population”* due to panelists with memberships in multiple online panels. According to comScore, the market research industry faces a *“crisis of ‘over-fishing’ and poor respondent quality”* that demands solutions to *“improve the quality of online samples and associated responses.”*

Harris Interactive takes these concerns seriously, and we continue to ensure that data of the highest quality are delivered to our clients. While we may disagree with comScore’s estimates and with their statements about the existence of a “crisis,” we do agree that perceived or real threats to accurate and unbiased data must be addressed.

Respondent Types

Implicit in concerns about “professional respondents” are assumptions about the degree to which respondents who participate in numerous surveys provide inaccurate, fraudulent, or biased responses. The quality of responses from “professional respondents” is automatically assumed to be lower than the quality of responses from those with lower levels of survey participation.

“Professional respondents” are assumed to be motivated by incentive maximization rather than intrinsic interest in survey participation, behavior which is assumed to be inconsistent with accurate and unbiased responses. When

incentives are absent, respondents who take numerous surveys are suspected of posing threats to data quality due to “conditioning” or “learning” that occurs in the process of taking numerous surveys.

Harris Interactive leads the industry in maintaining high panel standards and delivering accurate and unbiased data. As part of that leadership, we contend that the phrase “professional respondents” should be replaced by terms that allow us to address concerns about respondent quality in a more systematic manner.

For example, we can classify respondents as follows:

- **Fraudulent respondents:** Those who intentionally misrepresent themselves or provide inaccurate information—often because they desire to maximize the incentives they earn.
- **Inattentive respondents:** Those who, because of survey length, time constraints, or other reasons, do not appear to be providing thoughtful answers during the course of a specific survey.
- **Hyperactive respondents:** Those who participate in numerous surveys—especially online surveys—and who often belong to multiple online panels.
- **Conditioned respondents:** Those who, because of exposure to information in prior surveys, provide answers that are different from respondents who share their characteristics but who have not been exposed to the same type or degree of information.

Fraudulent and inattentive respondents may or may not be hyperactive respondents, but not all hyperactive respondents are fraudulent or inattentive. Conditioned respondents may or may not be hyperactive, but not all hyperactive respondents are conditioned.

Quality Maximizing Practices

Using these definitions, we now focus on Harris Interactive's research-on-online-research and solutions to concerns about panel and respondent quality to answer the following key questions:

- Is there evidence that these respondent types exist in the Harris Poll OnlineSM panel (HPOL)?
- If they exist, is there evidence that these respondent types threaten the quality of the data delivered to our clients?
- How does Harris Interactive ensure that potential threats to data quality are minimized?

Fraudulent Respondents

Our research on online research indicates that only a small proportion of our respondents behave in ways that raise suspicions as to whether they are providing fraudulent answers. Harris Interactive also sees little evidence that respondents are misrepresenting themselves as unique panelists by joining the Harris Poll Online panel using multiple email addresses.

- For example, across more than 20 large studies, approximately 3% of respondents who claim to be the sole user of their email accounts provided answers for age or sex that conflict with the demographic information stored for them from prior surveys.
- In 10 different controlled experiments that compared groups of respondents who were offered an incentive to those who were not, we found no evidence that those offered an incentive qualified for the study at higher rates than those not offered an incentive.
- In a large-scale study with many thousand respondents, Harris Interactive found that only 0.11% of respondents appeared to be registered in our panel under more than one email address given the pattern of responses provided.

Nevertheless, we continue to monitor standard responses to ensure data integrity. Age/sex consistency checks between survey data and stored panel member data can be done on virtually all of our online studies. Most clients choose to take advantage of this option. If a respondent provides information that does not match our records, the respondent is required to answer all demographic questions again. We do not rely on the stored data.

Given our strong panel and survey quality practices, the tiny amount of demographic instability found among respondents, and evidence from controlled experiments that qualification rates did not vary for those offered incentives, Harris Interactive clients who are willing to follow best practices for survey and invitation design can be confident that fraudulent respondents are having little or no effect on the quality of the data we deliver.

Inattentive Respondents

Research on research conducted by Harris Interactive and others indicates that in well designed surveys, only a small portion of respondents behave in ways that cast suspicion on whether they are paying full attention to the survey questions.

- Across a series of concept tests, Harris Interactive found that about 1% of respondents failed to differentiate their responses across a large set of attributes. While it is somewhat more difficult to identify random patterns of response, we found that another approximately 2.5% of respondents *might* have been responding randomly given the variation in their responses.
- Means and correlations were virtually identical regardless of whether respondents who appeared to be “clicking through” (also known as “straightlining”) or answering randomly were included or excluded from the analysis.

All surveys conducted by Harris Interactive must be approved by the Survey Methodology group in our Research and Methodology Department. They review each survey to ensure it meets our requirements for an optimal respondent experience and to minimize survey practices that might cause respondents to lose interest in the survey. Harris Interactive has also developed a “response equivalence” metric that allows us to assess the respondent burden posed by different types of survey questions. Our survey methodology standards cover a range of topics from right-to-left scrolling, to the number of questions to be banked on a single page, to the length of the interview and more—all geared toward ensuring that HPOL panelists do not grow frustrated with the survey experience or become inattentive.

Hyperactive Respondents

Harris Interactive's research on research into the characteristics of hyperactive respondents—panelists who belong to more than one online panel and who take numerous surveys—is ongoing. Our initial results show:

- Hyperactive respondents—identified by observed and self reported measures—are somewhat more likely to be female, age 50+, and to have lower income and education levels. Hyperactive respondents are less likely to be employed full time outside the home and are more likely to be answering the survey from home.
- Harris Interactive’s propensity score adjustment methods can be used to correct for potential attitudinal differences between hyperactive and non-hyperactive respondents. Propensity score adjustment, in conjunction with standard demographic weighting, can be used to bring demographic groups into proper alignment and to minimize potential attitudinal or behavioral differences.
- Ongoing research is being conducted by many organizations, but to date, there is no evidence indicating that existence of hyperactive respondents biases online survey data or causes clients to reach incorrect business decisions.
- If such evidence is found, the existence of hyperactive respondents will pose a threat to *all modes* of research because 19% of hyperactive respondents say that in a typical month they participate in one or more telephone surveys, 8% say they participate in one or more in-person at home surveys, and 47% say they participate in one or more mail surveys. These rates are significantly higher than those of their non-hyperactive counterparts.

Harris Interactive encourages telephone screening of focus group participants recruited from online panels as well as restricting access to such groups because the high cash incentives awarded for this type of research may be especially appealing to hyperactive respondents. Among hyperactive respondents, 29% indicated that they participate in one or more online focus groups in a typical month, which is significantly higher than the rate among their non-hyperactive counterparts.

Conditioned Respondents

Our research on research into conditioning, or learning effects, is ongoing. We have found few differences between those who have participated in many surveys over time versus those who have not. Our results show:

- To the extent that statistically significant behavioral or attitudinal differences occur between those who have recently taken more HPOL surveys and those who have

taken fewer, the differences arise on variables that measure the degree to which a person is engaged in society or the degree to which they are participatory. These differences are not necessarily the result of conditioning or learning, but simply reflect the fact that a tendency to participate in social activities is correlated with the tendency to complete surveys.

- Harris Interactive’s propensity score adjustment methods, in conjunction with standard weighting techniques, are specifically set up to remove potential biases due to differences in the participatory nature of different respondents.

Panel Management Practices

Harris Interactive makes use of a variety of panel management practices across all stages of research—from panel recruitment to sampling to survey invitations and incentives—to ensure the quality of our HPOL panel data.

- We recruit HPOL members using multiple methods, including websites, opting in client lists, telephone interviews, mass mailings and limited viral recruitment. This ensures a broad distribution of panelists across a variety of demographic categories and interest areas.
- Frequency of contact with HPOL panelists is actively managed. We apply limits on frequency of participation across all surveys and also apply exclusion rules across tracking studies.
 - As a general matter of policy, HPOL panelists receive 2-3 invitations to unique surveys on general topics each month. Some panelists with special interests (e.g., movies, product testing, and elections/political issues) or special characteristics (e.g., chronic illness sufferers) have agreed to take surveys specifically tailored to their interests or characteristics. These members may sometimes receive another 1-2 survey invitations per month.
- Survey invitations are written to minimize bias as appropriate, mentioning only general topics and purpose of the research and not revealing qualifying criteria.
- Panel membership is managed at the respondent, not household, level. Survey invitations are directed only to HPOL members. Only in rare instances, such as when

we contact parents to gain permission for their child to participate in a study, are panelists asked to forward the invitation to another member of the household.

- To ensure that our “novice” versus our “experienced” survey participants are distributed evenly across jobs, we pull samples based on HPOL members’ historical survey participation. Stratification by prior survey participation is superimposed behind the scenes on every standard sample pull.
- Our loyalty and retention program was developed to enhance the panel member’s experience, while avoiding the introduction of bias. This program relies on mix of incentive types including a sweepstakes called HIstakesSM, a points program called HIpointsSM, and the delivery of instant results at the end of each survey. Using survey results as an incentive along with token point levels ensures that intrinsic motivations for survey taking such as the desire to help others or voice one’s opinion are not overwhelmed by financial motivations.
- The token number of points awarded for survey participation is not substantial enough to be the sole influencer of panelist behavior. Additionally, our survey invitations do not typically inform HPOL members of the number of HIpoints they will earn for survey completion, decreasing the likelihood of respondents’ screening themselves into a survey.
- Harris Interactive HIpoints accounts are established on email address and cannot be transferred, removing motivation to register for HPOL under multiple email addresses. All prizes earned through this program are sent to a physical addresses provided by the respondent.

Panelist Privacy and Validation

Recently, one or two clients have asked about third party validation of our panel. These clients have asked whether we can release email addresses or other personally identifiable information to another vendor who would use this information to determine whether an individual had taken other online surveys or participated in other online research. We are currently unable and unwilling to participate in such types of validation for the following reasons:

- First, the validation in question is not validation in the traditional market research sense. Traditionally, validation of market research has been undertaken to detect fraud—committed either by the interviewer or the

respondent—in data collection. The fact that some individuals belong to multiple panels does not necessarily mean that they are providing inaccurate information; that is, hyperactive panelists are not necessarily fraudulent.

- Second, as shown above, provided that best practices for panel management and research design are followed, the effects of fraudulent, hyperactive, inattentive, or conditioned respondents on the quality of the data obtained from the Harris Poll Online Panel are small and unlikely to bias results.
- Third, validation of email addresses across panels will not detect the most damaging type of fraudulent respondent—individuals who join a panel multiple times under various email addresses for the purpose of accumulating incentives. Detection of this type of fraud is best done at the panel level through database and data file monitoring and cleaning activities.
- Finally, as a company in the business of market research, an industry that relies on the good will of the survey-taking public, we must ensure the longevity of that goodwill. While privacy protection has recently entered the forefront of public concern, protecting survey participants’ privacy has been an important part of our business since long before the advent of the Internet. We participated in industry-wide efforts to protect consumer information long before we began conducting surveys online and will continue to do so as a vested member of the market research community. For these reasons, we strictly adhere to our privacy policy, which dictates how we manage our relationship with our panel and the data we capture in the course of recruiting panelists and conducting surveys. Purposes of email contact with the panel are narrowly defined, and the release of personally identifiable information is only done with panelist consent. Even when asked, a significant portion of HPOL panelists will not agree to share personally identifiable information. These are standard tenets of online privacy and are likely in effect with most, if not all, online research providers.

Therefore, we think it likely that attempts to conduct panel validation by detection of duplicate email addresses across panels will be ineffective, and we think such methods, as currently espoused, are inappropriate to the online environment.

Summary

Harris Interactive’s panel recruitment, panel management, survey design, and bias correction methods ensure that our clients receive accurate data and results of the highest quality. Harris Interactive continues to lead the industry in the amount of research on online research we conduct, and we will continue to investigate whether the existence of hyperactive respondents—who belong to multiple panels and take numerous surveys—poses a threat to the validity and reliability of the results we deliver to our clients.

Harris Interactive and its clients do not face a “crisis” of poor respondent quality due to hyperactive, inattentive, fraudulent, or conditioned respondents. When appropriately

designed surveys are used, in conjunction with sampling and weighting methods that minimize bias, to collect data from properly recruited and managed panels, these respondents have little or no effect on data quality.

Conclusion

When appropriately designed surveys are used, in conjunction with sampling and weighting methods that minimize bias, to collect data from properly recruited and managed panels, these respondents have little or no effect on data quality.

Moving Forward with Confidence: Key Aspects of Harris Interactive’s Approach to Respondent Quality

	Fraudulent Respondents	Inattentive Respondents	Hyperactive Respondents	Conditioned Respondents
Definition	Those who intentionally misrepresent themselves or provide inaccurate information – often to maximize incentives	Those who do not appear to provide thoughtful answers during a specific survey	Those who participate in numerous surveys – especially online surveys – and who belong to multiple online panels	Those who, due to exposure to information in prior surveys, provide different responses than those not exposed to the information
Harris Interactive Research on Research	Less than 3% of respondents provided inconsistent age/gender information. Qualification rates did not vary for those offered incentives	Only 1% of respondents appeared to be “straight lining.” Another 2.5% <i>might</i> have been answering randomly.	There is no evidence to date indicating that existence of hyperactive respondents biases online survey data or causes clients to reach incorrect business decisions	We have seen no evidence of conditioning or learning effects in appropriately designed surveys or tracking studies
Harris Interactive Panel Management Solutions	Use points rather than cash when possible. Send incentives only to physical addresses. Limit viral recruitment	Use appropriate incentives based on survey length. Allow respondents to finish the survey later	Follow policies that limit frequency of contact with panelists. Use diverse recruitment methods including exclusive recruitment partnerships. Continue ongoing research to understand hyperactive respondents	Exclude panelists from repeated participation in tracking studies. Invite panelists to surveys on a wide variety of topics. Follow policies that limit frequency of contact. Use ongoing recruitment to keep panels fresh
Harris Interactive Research Design Solutions	Use consistency checks of survey information. Use general survey invitations. Design screeners to make it hard to know how to qualify	Use invitations to set expectations of survey length. Use survey design practices that decrease respondent burden	Use propensity score adjustment to minimize potential biases due to differences in participation or non-participation rates	Use propensity score adjustment or other adjustment techniques to minimize potential biases from learning effects

