



Comparing Google Consumer Surveys to Existing Probability and Non-Probability Based Internet Surveys

Paul McDonald, Matt Mohebbi, Brett Slatkin
Google Inc.

Abstract

This study compares the responses of a probability based Internet panel, a non-probability based Internet panel and Google Consumer Surveys against several media consumption and health benchmarks. The Consumer Surveys results were found to be more accurate than both the probability and non-probability based Internet panels in three separate measures: average absolute error (distance from the benchmark), largest absolute error, and percent of responses within 3.5 percentage points of the benchmark. These results suggest that despite differences in survey methodology, Consumer Surveys can be used in place of more traditional Internet based panels without sacrificing accuracy.

Paul McDonald is a Product Manager at Google.
Matt Mohebbi and Brett Slatkin are Software Engineers at Google.

Please address correspondence to:
Paul McDonald (consumer-surveys-support@google.com)
Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043

Introduction

Data collection for survey research has evolved several times over the history of the field, from face-to-face interviews and paper based surveying initially, to telephone based surveying starting in the 1970s to Internet-based surveying in the last 10 years. Three factors appear to have played a role in these transitions: data quality, data collection cost and data timeliness.

The Internet has the potential to collect data faster, with similar data quality and less cost. Chang and Krosnick (2009) compared random digit dialing (RDD), a probability based Internet survey (where respondents are chosen to be representative of the population) and a non-probability based Internet survey (where no effort is made to ensure the sample is representative) over the course of the 2000 presidential election. They found the probability based Internet survey produced results that were more closely aligned with actual voting behavior than RDD and non-probability based surveying.

However, there continue to be challenges with Internet surveys. Bias can be introduced into surveys when attempting to represent the entire U.S. population. While Internet use in the United States is approaching 78% of adults, these Internet users tend to be younger, more educated, and have higher incomes (Pew, 2011). Despite these challenges, for many types of surveys, the trade-off between acquiring the most representative sample and acquiring a sample quickly and inexpensively has led many commercial and academic institutions to favor these Internet based surveys.

In this paper, we introduce Google Consumer Surveys, a new method for performing probability based Internet surveying which produces timely and cost-effective results while still maintaining much of the accuracy of pre-existing surveying techniques. Section one provides an overview of the product, including survey sampling, data collection and post-stratification weighting. Section two compares the results obtained from Consumer Surveys to well known baselines and results obtained from other commercial Internet based surveying techniques. Section three provides a summary of the system along with its limitations.

Section I: Google Consumer Surveys Overview

Product summary

The core of Consumer Surveys is a “surveywall.” A surveywall is similar to the paywalls used by publishers to gate access to premium content but rather than requiring payment or subscription, visitors can instead choose to answer one or two survey questions. By reducing the burden to just one or two clicks, we increase the response rate of the survey. In conducted trials, the average response rate¹ was 16.75% compared to the latest industry response rates of less than 1% for most Internet intercept surveys (Lavrakas, 2010), 7-14% for telephone surveys (Pew, 2011. Pew, 2012) and 15% for Internet panels (Gallup, 2012).

Consumer Surveys involves three different groups of users: researchers, publishers, and consumers. Researchers come to Consumer Surveys and pay to have consumers answer them. Consumers encounter these survey questions on publisher websites and answer questions in order to obtain access to the publisher content. Publishers sign up for Consumer Surveys and are paid by Google to have surveys delivered to their site. Thus, Consumer Surveys provides a new way for researchers to perform Internet surveys, for publishers to monetize their content and for consumers to support publishers.

Data collection

Many researchers perform multi-question surveys in which the same respondent is asked to fill out a several minute questionnaire. With Consumer Surveys, researchers create and run multi-question surveys but the system only asks users one or two questions per request. Two-question surveys are called screening surveys and require the first question to have a simple yes/no response (e.g. “Are you a dog owner?”). Both the sole question of a single question survey and the second question of a screening survey can be of the form multiple choice, image choice, or five point scale rating. The completion rate for screening surveys average 9.25% but trial participants have expressed that the value of the screening question outweighs the increased non-response.

Unlike traditional surveys which explicitly ask respondents for demographic and location information, Consumer Surveys infers approximate demographic and location information using the respondent’s IP address² and DoubleClick cookie³. The respondent’s nearest city can be determined from their IP address. Income and urban density can be computed by mapping the location to census tracts and using the census data to infer income and urban density. Gender and age group⁴ can be inferred from the types of pages the respondent has previously visited in the Google Display Network using the DoubleClick cookie.⁵ This information is used to ensure each survey receives a representative sample and to enable survey researchers to see how sub-populations answered questions. Inferring this demographic data enables Consumer Surveys researchers to ask fewer questions in a survey which in turn increases response rates.

Sampling

Probability based Internet survey platforms typically recruit respondents via telephone using RDD telephone sampling techniques, but then require that panel members answer surveys online. By contrast, Consumer Surveys makes use of the inferred demographic and location information to employ stratified

1 Using Princeton Survey Research Associates International (PSRAI) methodology.

2 A number assigned to a device connected to the internet.

3 An advertising cookie used on AdSense partner sites and certain Google services to help advertisers and publishers serve and manage ads across the web.

4 Age groups supported: 18-24, 25-34, 35-44, 45-54, 55-64, and 65+.

5 See the privacy section for details.

sampling. The target population for Internet access among the U.S. population of adults is obtained from the most recent Current Population Survey (CPS) Internet use supplement (October 2010) and is formed from the joint distribution of age group, gender and location. Since this inferred demographic and location information can be determined in real time, allocation of a respondent to a survey is also done in real time, enabling a more optimal allocation of respondents across survey questions. This reduces the size of the weights used in post-stratification weighting which in turn reduces the variance introduced by weighting. Consumer Surveys gives researchers the ability to target specific sub-populations through two methods: demographic targeting and screening (described in Data Collection). Demographic targeting enables researchers to target a survey to a specific demographic sub-population (age, gender or location).

Post-stratification weighting

Consumer Surveys uses post-stratification weighting to compensate for sample deficiencies. Although Consumer Surveys attempts to build an optimal allocation of respondents to each question over the life of the survey, this is not always possible in practice due to additional constraints such as completing the survey in a timely manner, publisher inventory at the time, and competition with other surveys. Thus, post-stratification weighting is used to reduce this sample bias. The same CPS Internet target population used in sampling is also used in weighting. To reweight to this population, a simple cell weighting method is used, where cells are formed from a combination of age group, gender and location. The particular combination used for reweighting is a function of the number of respondents for the question and if the population is targeted (e.g. it is not useful to reweight by gender for a survey targeting males). The possible weighting options, ordered by priority, include: (age, gender, state), (age, gender, region), (age, state), (age, region), (age, gender), (age), (region), (gender).⁶ The weighting option is automatically chosen for a particular view and is shown as part of the report in Consumer Surveys.

Preview of post-surveywall content.

Refinery reports 135 percent earnings increase
 CALGARY, Alberta — Husky Energy Inc. is reporting a 135 percent increase in net earnings in 2011 from production growth, higher crude oil prices and improved refining margins... .

Answer a question to continue reading this page

What kind of packaging do you prefer to have for cereal?

Paper box

Plastic wrapping

[I don't know, show me another question](#)

OR

Existing User?

User Name:

Password:

[Sign In](#)

powered by Google™ [Learn more - Privacy](#)

Google Consumer Surveys surveywall.

Redacted post-surveywall content.

Figure 1. An example Consumer Surveys surveywall on www.limaohio.com. Once the user has answered this question, they can continue reading the article.

⁶ Google Consumer Surveys divides the United States into four regions, West (CA, AK, WA, OR, HI, ID, MT, WY, NV, UT, CO, AZ, NM), Midwest (ND, SD, NE, KS, MO, IA, MN, WI, MI, IL, IN, OH), South (TX, OK, AR, LA, KY, TN, MS, AL, FL, GA, SC, NC, VA, WV, MD, DC, DE) and Northeast (PA, NY, NJ, ME, VT, NH, MA, RI, CT).

Section II: Comparing Google Consumer Surveys to Existing Internet and Telephone Surveys

Method

Respondents

Two survey research firms were hired to administer an identical questionnaire, with one performing a probability based Internet survey and the other a non-probability based Internet survey. Both are well known for administering such surveys and each was asked to provide “1,000-2,000 responses for each question from a representative sample of American adults 18 years and older, residing in the United States.”

Consumer Surveys respondents were administered the same questionnaire. This process was repeated five times for Consumer Surveys to determine the variation of the responses. Each of the Consumer Surveys attempted to represent the U.S. Internet population as determined by the CPS Internet Supplement. The probability sample Internet survey was conducted with members of a panel recruited via RDD methods. Respondents that wanted to participate but did not have a computer or Internet connection were provided with one at no cost. The respondents were a subset of the individuals on the panel and they completed the survey via the Internet. The non-probability sample Internet survey recruited a stratified random sample of respondents via email. Consumer Surveys employed a surveywall on partner publisher sites and blocked access to the content of the site until the user either paid for the content or answered a one or two question survey.

Measures

Three questions measuring media consumption were asked in the same questionnaire administered to the probability based panel and the non-probability based panel. The same questions were also asked to the Consumer Surveys respondents; however, questions were asked one at a time and each respondent did not necessarily answer more than one question. Primary demographics (age, gender, income, race, education and location) were captured for each of the respondents in the panel surveys either via the questionnaire or at the time respondents signed up for the panel. Consumer Surveys utilized the users’ DoubleClick cookies to infer age and gender. Approximate location was determined using the IP address of the respondent. Income and urban density were computed by mapping the location to census tracts and using the census data to infer income and urban density (for details, see Data Collection in Section I).

Benchmarks

Three media benchmarks were measured in a large (200,000 respondent) semi-annual RDD telephone survey conducted by a respected research provider. The benchmarks measured Video on Demand (VoD), Digital Video Recorder (DVR) and satellite dish usage in American households. Four health benchmarks were also measured against responses drawn from the Consumer Surveys respondents. Large government surveys from the CDC with response rates of over 80% were used to obtain health benchmarks for smoking, asthma, disability and car accident incidence rates.

Weighting

To reduce the effects of non-response and non-coverage bias or under- and over-sampling resulting from the study-specific sample, post-stratification weighting was done on the responses from each of the methodologies. The non-probability panel data was weighted to reflect the latest Census including age, gender, household income, and region. The probability based panel and Consumer Surveys respondents

were weighted by the most recent data from the Current Population Survey (CPS) Internet Supplement, including age, gender and geography.

Analysis

Accuracy measure

We considered the media and health benchmarks to be the gold standard and used those to determine the error of the various survey platforms by computing the deviations between the benchmarks and the other survey methodologies. For each survey, the average absolute error was computed for each question and the survey as a whole as well as the maximum of the absolute error and the percent of responses within 3.5 percentage points of the benchmarks (Yeager, Krosnick, Chang et al., 2009). The above analyses were first conducted with post-stratification weights on the panel data and the Consumer Surveys data and then repeated without post-stratification weights on the Consumer Surveys data.

Variation between survey platforms

Comparisons of variation between samples on each platform were limited due to the number of surveys run using the probability and non-probability based panels. However, the goal of this study was to compare the differences between platforms so most of the analysis focuses on the differences between the average Consumer Surveys sample and the samples provided by the other platforms. The benchmark and media consumption based surveys were run five times in a six month period to compare the variation of responses across surveys and to measure how the publisher website mix affected the validation metrics. Consumer Surveys were targeted so that some attempted to get a representative sample of the U.S. Internet population while others only asked respondents of a given publisher site. Using this methodology we could test the bias introduced by each publisher independently. Subsequent studies will focus on reducing the bias introduced by individual content providers.

Results

Accuracy across benchmarks

The average absolute error (the absolute difference between the percentage of the population that choose the answer and the benchmark) was measured for each survey platform (including Consumer Surveys) using a seven question validation survey (see Table 2).

Raw results

We were only able to capture raw results and weighted results from the Google Consumer Surveys platform. Both panel results came back weighted, without the raw counts. In both the media consumption based comparisons and the health benchmark comparisons, the raw results were less accurate than the weighted results (average absolute errors of 4.66% and 4.52%, respectively).

Weighted results

After post-stratification of the Consumer Surveys results, accuracy for the media benchmark was best (average absolute error 3.25%), and slightly less accurate (average absolute error 4.20%) for the government health survey. The probability sample Internet survey and the non-probability Internet survey were both marginally less accurate than the Consumer Surveys (4.70% and 5.87% respectively). As expected, post-stratification significantly increased the average accuracy of the Consumer Surveys. The

average absolute error for the non-Google samples was 5.29% across all benchmarks, while the Google samples averaged 3.76%.

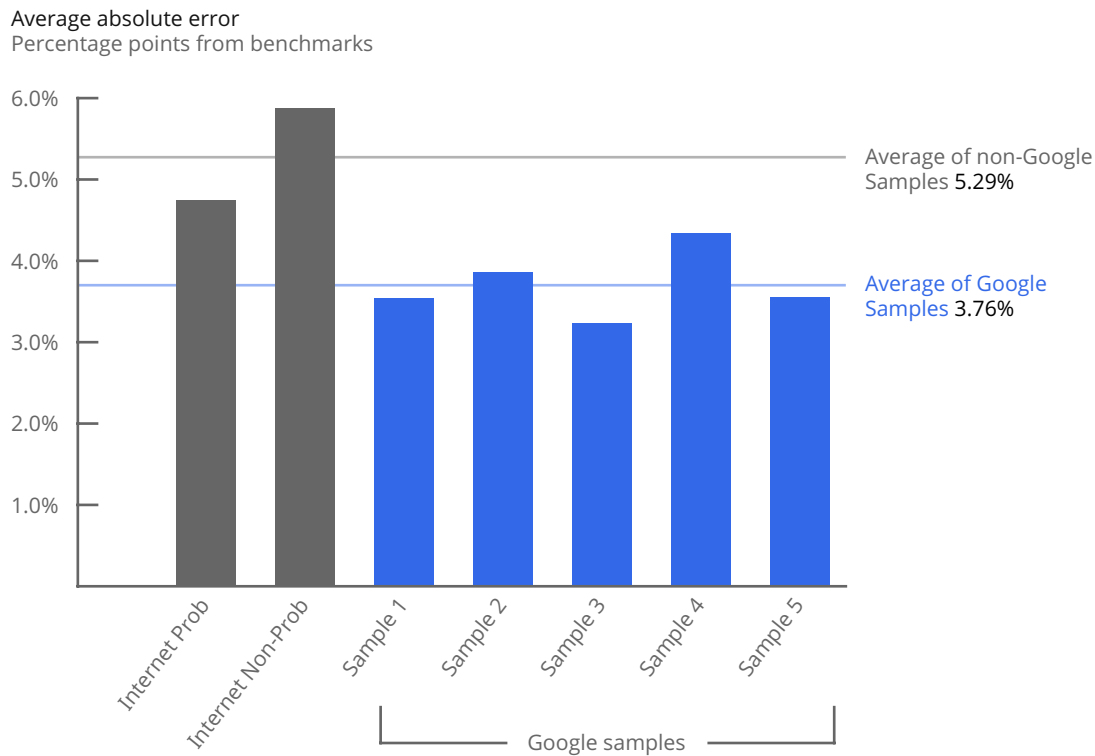


Figure 2. Average absolute error over the five trials conducted on Consumer Surveys compared with single trials from the probability based and non-probability based Internet panels.

Other accuracy metrics

Largest absolute error

The largest absolute error is another measure of accuracy, identifying the largest spread from the health and media benchmarks to each of the survey platforms. Measures without post-stratification weighting were less accurate based on this metric (11.70% for unweighted Consumer Surveys vs 8.40% for weighted). After post-stratification both the probability based sample and the non-probability based survey had larger absolute error when compared to Consumer Surveys' 8.40% (vs 9.20% and 11.40%

respectively). The average of the largest absolute error measures across all benchmarks was 10.3% for the non-Google samples while the Google samples average 7.20%.

Largest absolute error

Percentage points from benchmarks

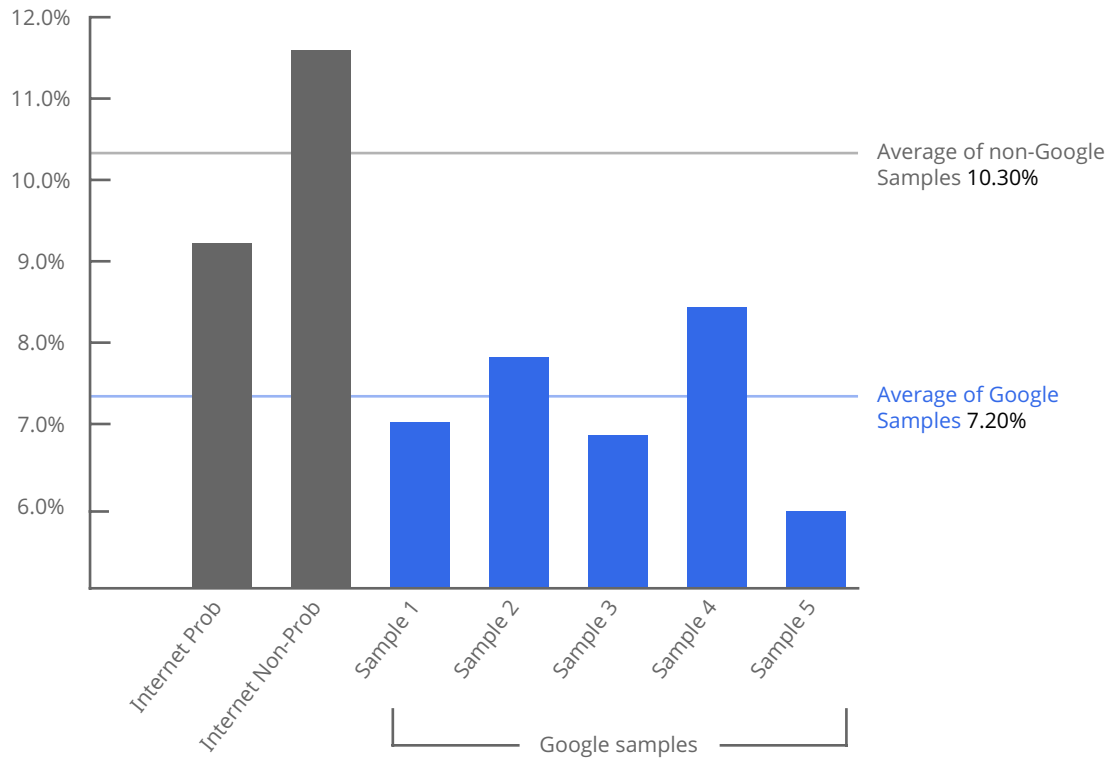


Figure 3. Largest absolute error over the five trials conducted on Consumer Surveys compared with single trials from the probability based and non-probability based Internet panels.

Percent of results within 3.5 percentage points

The final measure of accuracy used is percentage of measurements within 3.5 absolute percentage points of the benchmarks. Post-stratification weighted Consumer Surveys data averaged 48.54% of results within 3.5 percentage points while the unweighted Consumer Surveys data averaged 34.29% of results

within 3.5 percentage points. Both the probability based sample and the non-probability based Internet survey had 33.33% of results within 3.5 percentage points.

Percent within 3.5 absolute percentage points

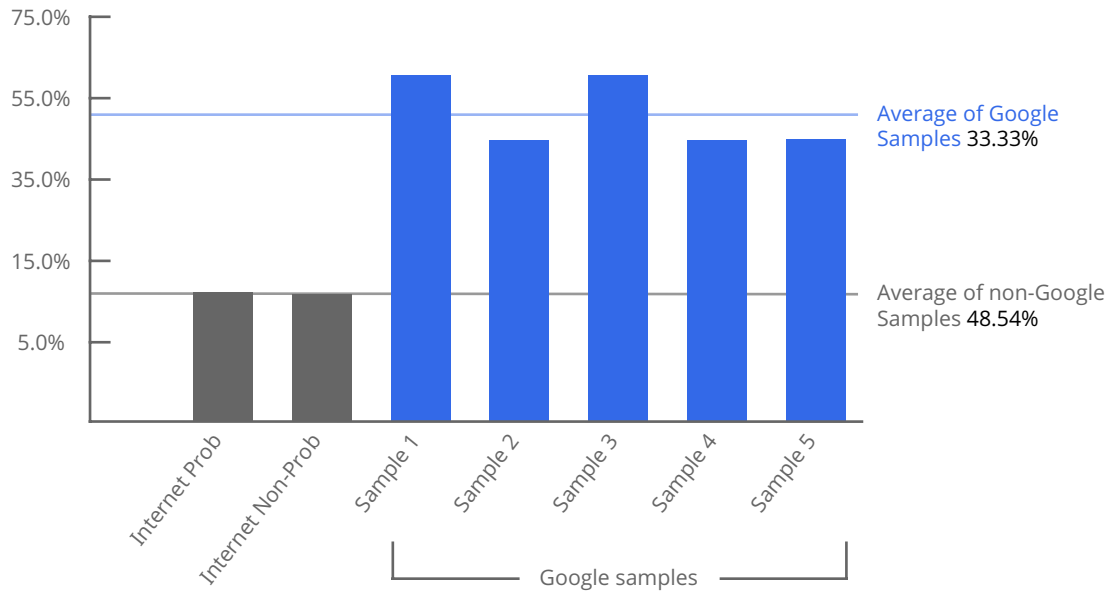


Figure 4. Percent of responses within 3.5 absolute percentage points over the five trials conducted on Consumer Surveys compared with single trials from the probability based and non-probability based Internet panels.

Section III: Conclusions and Limitations

Conclusions

Google Consumer Surveys provides both a new way to perform Internet surveys and a new method for publishers to monetize their content. Since Consumer Surveys run directly within publisher sites, the respondents may be more representative than respondents of more traditional internet surveys.

Response rates for Google Consumer Surveys are higher than telephone surveys and standard Internet panels, and are much higher compared to many Internet intercept surveys. This higher response rate is due, in part, to the short survey length of Consumer Surveys and the inferred demographic data.

Accuracy of Consumer Surveys is better than both the probability and non-probability based Internet panels on three separate measures: average absolute error (distance from the benchmark), largest absolute error, and percent of responses within 3.5 percentage points of the benchmarks. These results suggest that despite differences in survey methodology, Google Consumer Surveys can be used in place of more traditional Internet-based panels without sacrificing accuracy.

Limitations

Since Google Consumer Surveys only allows one-question or screening two-question surveys, analysis of the relationships between survey questions are difficult or sometimes not even possible. Bias can be introduced into surveys attempting to represent the U.S. population as Internet penetration in America is only 78% of adults. Internet users tend to be younger, more educated, and have higher incomes. Furthermore Google Consumer Surveys are served on our publisher network which, while large, does not fully encompass the breadth of Internet content available and therefore respondents can only be taken from a more limited sample.

Our initial study focused on two types of benchmarks (media usage and health) derived from large population surveys. We believe that these benchmarks represent the U.S. population but their reach is limited in terms of the diversity of the subject matter. It's possible that Google Consumer Surveys has inherent bias in other areas and the researcher should be aware of those potential limitations. Finally, because of the way Consumer Surveys are presented to potential respondents (a format that protects premium content and prevents the user from reading the content) some types of questions may be regarded as too sensitive or suspicious. For example, asking a user about their bank account or credit card usage may appear as an advance from an untrustworthy advertiser or website. Consumer Surveys attempts to mitigate these issues by branding the survey as "powered by Google" and providing quick access to more information about how and why the data is collected, but some bias may exist for these types of questions.

Acknowledgements

The authors would like to thank the entire Consumer Surveys team for their effort in making this product, Art Owen⁷ for his advice on statistical methods and for his feedback on this manuscript, Mike Murakami for his advice on survey methodology and for his feedback on this manuscript, and Adam Baker for his work on typesetting this manuscript.

Privacy

Privacy is very important to Google. Users always have the option to pass on a question and have the ability to opt-out of the DoubleClick cookie completely. IP location is approximate to the nearest city. The results of Google Consumer Surveys are only available to a survey researcher in aggregate form across many users and are never resold. For additional details, please see the Ads section of the Google Privacy Center (<http://www.google.com/privacy/ads>).

References

- Chang and Krosnick, 2009. National Surveys via RDD Telephone Interviewing Versus the Internet. Public Opinion Quarterly.
http://comm.stanford.edu/faculty/krosnick/docs/2009/2009_poq_chang_rdd.pdf
- Gallup, 2012. Gallup Research Consulting
<http://www.gallup.com/consulting/government/102403/research.aspx>
- Lavrakas, 2010. An Evaluation of Methods Used to Assess the Effectiveness of Advertising on the Internet.
http://www.iab.net/media/file/Evaluation_of_Internet_Ad_Effectiveness_Research_Methods.pdf
- Pew Research Center's Internet & American Life Project, 2011. Demographics of Internet Users.
<http://pewInternet.org/Static-Pages/Trend-Data/Whos-Online.aspx>
- Pew Research Center's Internet & American Life Project, 2011. Real time Charitable Giving (methodology).
<http://www.pewInternet.org/Reports/2012/MobileGiving/Methodology/Methodology.aspx>
- Pew Research Center's Internet & American Life Project, 2012. Tablet and E-book reader Ownership Nearly Double Over the Holiday Gift-Giving Period (methodology).
<http://www.pewInternet.org/Reports/2012/E-readers-and-tablets/Methodology/About-this-report.aspx>
- Pew Research Center's Internet & American Life Project, 2012. Teens, kindness and cruelty on social networks (methodology).
<http://www.pewInternet.org/Reports/2011/Teens-and-social-media/Methodology/Survey.aspx>
- Yeager, Krosnick, Chang et al., 2009. Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples.
<http://comm.stanford.edu/faculty/krosnick/Mode%2004.pdf>

⁷ Art Owen is a professor of Statistics at Stanford University. His contributions to this project were as a consultant for Google and not part of his Stanford responsibilities.

Appendix: Data

Table 1. *Survey sample descriptions.*

Survey	Invitations [†]	Responses	PSRAI Response Rate	Cost	Time to results	Average Absolute Error	Field Dates	Prob- ability Sample	Quota Used	Incentives Offered
Non-Google Surveys										
Internet Probability	1,995	1,165	2.6%	\$8,100.00	8 days	4.70%	Jul 20–26, 2011	Y	N	Points; free Internet access; sweepstakes
Internet Non- probability	10,085	2,017	N/A	\$6,900.00	23 days	5.87%	Aug 16–19, 2011	N	Y	Points; sweepstakes
Average	6,040	1,591	0.026	\$7,500.00	15.5 days	5.29%				
Google Consumer Surveys										
Sample 1	7,712	1,916	25.23%	\$1,340.90	1 day	3.56%	Nov 6–7, 2011	-	N	Access to online content
Sample 2	29,320	3,001	10.66%	\$2,100.80	1 day	3.97%	Dec 14–15, 2011	-	N	Access to online content
Sample 3	40,763	3,010	7.78%	\$2,106.90	1 day	3.09%	Jan 10–12, 2012	-	N	Access to online content
Sample 4	20,191	3,262	16.46%	\$2,283.50	12 hours	4.49%	Jan 23, 2012	-	N	Access to online content
Sample 5	20,805	2,556	23.63%	\$1,789.16	4 hours	3.71%	Jan 30, 2012	-	N	Access to online content
Average	23,758	2,749	16.75%	\$1,924.25	1 day	3.76%				

[†] For Consumer Surveys samples, an impression of the survey is equivalent to an invitation.

Table 2. *Benchmark measurements.*

Type	Question	Yes	No	Source
Media	Does your household have a satellite dish?	28.7%	71.3%	Scarborough RDD telephone Poll (Scarborough 1H 2011, 208,274 respondents)
Media	Have you watched any programs or events on Video-On-Demand in the past 12 months?	20.6%	79.4%	Scarborough RDD telephone Poll (Scarborough 1H 2011, 208,274 respondents)
Media	Does your household own or use a digital video recorder (DVR)?	43.1%	56.9%	Scarborough RDD telephone Poll (Scarborough 1H 2011, 208,274 respondents)
Health	Have you smoked 100 or more cigarettes in your lifetime and now smoke every day or occasionally?	20.6%	79.4%	http://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm#national
Health	Have you been diagnosed with asthma at any point in your lifetime?	13.8%	86.2%	http://apps.nccd.cdc.gov/brfss/list.asp?cat=AS&yr=2010&qkey=4417&state=All
Health	In the past year have you been in a car accident while driving that involved property damage, an injury or fatality?	4.0%	96.0%	http://www.distraction.gov/research/PDF-Files/Distracted-Driving-2009.pdf
Health	Because of a physical or mental health condition do you have difficulty with daily life activities, walking, or working around the house or at a job?	20.8%	79.2%	http://apps.nccd.cdc.gov/brfss/list.asp?cat=DL&yr=2010&qkey=4000&state=All

Table 3. Overall accuracy metrics.

Evaluative Criteria	Non-Google Surveys			Google Consumer Surveys					
	Internet Probability	Internet Non-probability	Average	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Average
Average absolute error									
Media Benchmarks	4.70%	5.87%	5.29%	3.40%	4.73%	1.77%	3.27%	3.07%	3.25%
<i>Without Post-Stratification</i>	-	-	-	5.67%	5.07%	3.10%	4.97%	4.50%	4.66%
Health Benchmarks	-	-	-	3.68%	3.40%	4.08%	5.63%	4.20%	4.20%
<i>Without Post-Stratification</i>	-	-	-	5.20%	4.03%	3.95%	4.80%	4.60%	4.52%
Combined Average Absolute Error	4.70%	5.87%	5.29%	3.56%	3.97%	3.09%	4.49%	3.71%	3.76%
<i>Without Post-Stratification</i>	-	-	-	5.40%	4.47%	3.59%	4.89%	3.99%	4.47%
Rank (Average Absolute Error)	5	7	-	2	3	1	4	6	-
Rank <i>Without Post-Stratification</i>	3	7	-	6	2	1	4	5	-
Largest absolute error									
All	9.20%	11.40%	10.30%	7.00%	7.70%	7.00%	8.40%	5.90%	7.20%
<i>Without Post-Stratification</i>	-	-	-	11.70%	10.30%	7.20%	7.40%	8.00%	8.92%
Rank	6	7	-	3	4	2	5	1	-
Rank <i>Without Post-Stratification</i>	4	6	-	7	5	1	2	3	-
% of answers with 3.5 pp of benchmarks									
All	33.30%	33.30%	33.30%	57.14%	42.86%	57.14%	42.86%	42.86%	48.57%
<i>All Without Post-Stratification</i>	-	-	-	14.29%	42.86%	42.86%	28.57%	42.86%	34.29%
Rank	6	7	-	1	3	2	4	5	-
Rank <i>Without Post-Stratification</i>	4	5	-	7	1	2	6	3	-