

Article

Symposium 2008:
Data Collection: Challenges, Achievements and New Directions

Accuracy of Travel Data Samples: Utilizing Online vs. Mail Methodologies

by Nandini Nadkarni, Ph.D. and George Harmon

2009



Accuracy of Travel Data Samples: Utilizing Online vs. Mail Methodologies

Nandini Nadkarni, Ph.D. and George Harmon¹

Abstract

Online data collection emerged in 1995 as an alternative approach for conducting certain types of consumer research studies and has grown in 2008. This growth has been primarily in studies where non-probability sampling methods are used. While online sampling has gained acceptance for some research applications, serious questions remain concerning online samples' suitability for research requiring precise volumetric measurement of the behavior of the U.S. population, particularly their travel behavior. This paper reviews literature and compares results from studies using probability samples and online samples to understand whether results differ from the two sampling approaches. The paper also demonstrates that online samples underestimate critical types of travel even after demographic and geographic weighting.

Key Words: Online data collection, Travel studies, Accuracy, Consistency.

1. Introduction

1.1 Background

Many organizations in recent years have considered moving their tracking studies and custom research online in an effort to speed response times and to lower field or operational costs. While speed and costs savings are worthy goals, they are secondary to the primary goal of valid, representative measurements. For managers of private and public travel organizations with resource accountability and responsibility to shareholders or citizens, accurate and credible reports about travel behavior and visitors is the primary requirement.

Questions also arise regarding the differences between traditional sampling approaches and online panels. Traditional probability samples are structured from samples designed to represent the population. These samples are constructed from the bottom up by reaching out to geographic sections and balancing the panel by age, gender, income and other known characteristics at each level. Online samples are different in that they are constructed from volunteers – they are not recruited from geographic areas; rather they are recruited across the Web. After volunteers are assembled, online panel managers assign the volunteers back to geographic blocks and attempt to construct a representative sample. Identifying and maintaining volunteer samples balanced demographically and proportionate within state geographic levels is a challenge.

Online samples can be representative of online populations, but online populations may not be representative of the total U.S. population. According to the Pew Internet & American Life Project's December 2008 survey on Internet access, 73% of American adults in the U.S. use the internet or email (Madden and Jones, 2008). Another survey by Pew Internet & American Life Project found that 55% of adult Americans now have broadband internet connections at home, up from 47% who had high-speed access at home in the prior year. The two groups that report no change in growth include African Americans and lower income households. Non-internet users represent a large percentage of potential broadband users, but many are just not interested in getting online. About one-quarter (27%) of adult Americans are not internet users, and they tend to be older (the median age is 61) and have lower-incomes than online users. Non-internet users are more than twice as likely as users to live in low-income households (Horrigan, 2008).

Underscoring the challenges associated with online samples, leading professional research associations have issued strict guidelines for researchers using such non-probability samples with online volunteers and calling it inappropriate to even try to estimate sampling error for these studies. The American Association for Public Opinion

¹ D.K. Shifflet & Associates Ltd., 1750 Old Meadow Rd Suite 620, McLean, VA 22102 (nnadkarni@dksa.com).

Research (AAPOR) is the leading professional organization of public opinion and survey research professionals in the U.S., with members from academia, media, government, the non-profit sector and private industry.

AAPOR provides the following guidance to members when using opt-in polls like online samples, “For opt-in surveys and polls, responsible researchers and authors of research reports are obligated to disclose that respondents were not randomly selected from among the total population, but rather from among those who took the initiative or agreed to volunteer to be a respondent” And AAPOR further advises members: AAPOR recommends the following wording for use in online and other surveys conducted among self-selected individuals: “Respondents for this survey were selected from among those who have [volunteered to participate/registered to participate in (company name) online surveys and polls]. The data (have been/have not been) weighted to reflect the demographic composition of (target population). Because the sample is based on those who initially self-selected for participation [in the panel] rather than a probability sample, no estimates of sampling error can be calculated. All sample surveys and polls may be subject to multiple sources of error, including, but not limited to sampling error, coverage error, and measurement error” (American Association for Public Opinion Research).

The Canadian Market Research and Intelligence Association (an association representing the recently merged Canadian Association of Market Research Organizations (CAMRO), the Canadian Survey Research Council (CSRC) and the Professional Marketing Research Society (PMRS) also note the limits of non-probability sampling in their 2007 Code of Conduct Integrity of Reporting section stating that researchers should “refrain from making statements about margins of sampling error on population estimates when probability samples are not used” (Canadian Market Research and Intelligence Association, 2007).

To evaluate new methodologies such as online sampling, comparability studies are developed to assess whether the new methodology delivers equally valid results as the established methodology. If results from the new methodology equal those of the traditional methodology then the new methodology could be a viable substitute. For 25 years D.K. Shifflet & Associates has conducted large mail panel surveys among representative samples of the U.S. population (50,000 a month for the last 15 years). These monthly studies have consistently been used to track travel behavior, travelers, visitors and the travel expenditures of U.S. residents. Results from these studies are used to estimate total travel, market share for commercial brands and travel volume for the total U.S., individual states and cities. D.K. Shifflet & Associates’ research systems and methods produce an accuracy that closely replicates key metrics in various travel vertical sectors. Their decades’ rich database of travel behavior enables comparison to periodic, large scale government studies that validate the accuracy of D.K. Shifflet & Associates systems.

During 2006 and 2007 D.K. Shifflet & Associates tested more than 200,000 online households in parallel, using the same monthly timeframes and questions as the standard D.K. Shifflet & Associates’ mail panel studies, thus enabling a direct comparison of online samples to a proven traditional probability sample. Further, D.K. Shifflet & Associates has explored whether online samples can be reweighted to appear demographically representative of the U.S. population to determine whether their attitudinal and behavioral questions were representative.

D.K. Shifflet & Associates is clearly in a unique position to compare results gathered from online samples with traditional probability based samples. With D.K. Shifflet & Associates’ validated methodology as a benchmark, the results obtained from online samples from parallel tests can reveal whether online samples accurately represent U.S. residents’ demographics, travel behavior and attitudes.

2. Analyses

The senior staff of D.K. Shifflet & Associates has decades of data collection experience utilizing the NPD, NFO and Market Facts/Synovate panels as well as in-depth experience with a number of well established commercial online sample providers. For the current analyses, D.K. Shifflet & Associates’ first compared the composition of the final sample, that is those responding to the survey both as travelers and non-travelers, from both the mail and online samples. Both samples were designed to be geographically representative and balanced by age, gender, income, education and household composition.

Below are a few examples of the major differences between the parallel online samples and the benchmark mail studies. (Table 4-1) The final sample composition of online samples generally:

- under represent those with less education defined as less than a high school education;
- under represent those with lower incomes defined as those households making less than \$25,000 per year;
- over representative of households making more than \$75,000 per year; and
- over represent ages 45-65 and under represent ages 65+.

While respondent reweighting can address many of the demographic differences, the critical question remains; do online panel members have different attitudes or behave differently once demographic differences are factored out? The following differences among online samples are noted after reweighting to make both samples comparable in terms of demographics (Table 4-2). Online samples:

- report far fewer business trips;
- under represent attendance at business group meetings trips;
- take more day trips;
- have far more travel to visit friends and relatives;
- report shorter stay durations; and
- have lower spending visitors.

3. Conclusion and implications

This analysis illustrates that travel studies using online data alone, do not provide consistent and accurate representations of the total U.S. population and their travel behavior, even after reweighting the data on the basis of known geographic and demographic characteristics. Inaccurate and highly variable online results are seen for attitudinal and behavioral measures including market share. Furthermore, unless online samples are substantially weighted, they provide an understatement of travel behaviors such as visitor levels, business and convention travel volume, overnight stays and all visitor spending.

Without known proportions it is impossible to adjust online samples. To generate known proportions requires costly parallel studies with representative samples available through traditional offline methodologies. Given the historical monthly and seasonal variations in travel, a large scale representative sample is required each month. Online samples simply do not provide consistent and accurate measures of travel at this time.

Online samples can be used, but only as a complimentary supplement to frequently recurring (monthly), representative samples that generate benchmarks for reweighting the online responses. These analyses show that online samples are not as representative of the U.S. population as traditional sampling methodologies such as mail panel studies. This is not surprising given that a significant portion of the U.S. population does not use the Internet. At some point in the future online samples may accurately estimate key travel behaviors, but for the present and near future, studies based on online sampling alone are likely to be inconsistent, inaccurate and misleading. For managers of private and public travel organizations with accountability to shareholders or citizens, credible travel volume and profile reports must be based on representative probability samples to ensure valid and reliable numbers.

4. Tables

The following tables illustrate compositional differences between respondent final samples (samples of persons responding to the survey as travelers and non-travelers) from online surveys contrasted as with probability samples. Differences at 0 indicate comparability to probability results.

Negative (-) differences indicate that the online respondents are less represented than the mail sample.

No negative sign indicates that the online respondents are more highly represented.

Table 4-1
Profile of final samples

	Percentage Difference	
	2006	2007
Age		
18-34 years	2	-16
35-44 years	7	13
45-54 years	14	28
55-64 years	21	31
65 years +	-33	-43
Education		
Up to high school	-17	-22
Some college (1-3 yrs)	13	9
At least a bachelor degree	9	15
Post-graduate degree	-12	1
Occupation		
Managerial, professional	2	17
Technical, sales, administrative	3	3
Service	-30	-26
Farming, forestry, fishing	-58	-57
Craftsman, repairman	-43	-19
Operator, laborer	-13	-16
Retired, student, armed forces	7	-2
Sex		
Male	-14	-8
Female	8	4
Household Income		
Less than \$15,000	-31	-46
\$15,000 - \$24,999	-18	-29
\$25,000 - \$34,999	-12	-20
\$35,000 - \$49,999	-3	-2
\$50,000 - \$74,000	19	26
\$75,000 - \$99,999	36	14
\$100,000 - \$149,999	12	53
\$150,000+	-7	43

Table 4-2
Behavioral difference in final samples - Demographics controlled

	Percentage Difference	
	2006	2007
Purpose		
Business	-26	-20
Group business	-30	-20
Transient business	-24	-20
Leisure	11	8
Vacation	-14	-15
Visit friend/relative	49	40
Special event/other	8	11
Length of Stay		
Day	23	34
1 Night	0	2
2 Nights	-1	-6
3-5 Nights	-12	-16
6-7 Nights	-29	-34
8-14 Nights	-34	-44
15 or more Nights	-35	-45
Transportation Type		
Airplane	-20	-17
Train	110	58
Bus	-34	-38
Auto	10	11
Camper/RV	120	146
Large truck or semi	-100	-100
Other	-46	-59
Accommodation Type		
Home/apt/condo(not mine)	118	115
My second home	-99	-98
Hotel	-34	-31
Timeshare	-59	-60
Bed and breakfast	-60	-52
Camping	-51	-57
Ship	-95	-97
Other non-hotel	74	111
Corporate apartment	71	-27
Travel Spending		
Total Spending	-16	-16
Food	-15	-17
Transportation	-6	-4
Entertainment	-15	-18
Other Expenditures	-9	-16
Shopping	-10	-12

References

American Association for Public Opinion Research. Opt-In Surveys and Margin of Error. unpublished report.

Canadian Market Research and Intelligence Association (2007). Code of Conduct and Good Practice for Members of the Marketing Research and Intelligence Association, 10.

Horrigan, J.B. (2008), Home Broadband Adoption 2008, Pew Internet & American Life Project, Washington D.C., USA.

Madden, M. and Jones, S. (2008), Networked Workers. Pew Internet & American Life Project, Washington D.C., USA.