UNLV CENTER FOR CRIME AND JUSTICE POLICY

September 2014, CCJP 2014-01

RESEARCH IN BRIEF

A Comparison of Different Online Sampling Approaches for Generating National Samples

By Miliaikeala S.J. Heen, M.A., Joel D. Lieberman, Ph.D., and Terance D. Miethe, Ph.D.

Online (internet) surveys provide a new and evolving method for measuring public opinion on a local, national, and international level. Compared to traditional survey approaches (e.g., face-to-face, mail, and telephone surveys), online surveys offer the advantages of speed, efficiency, and lower costs in data collection (see Dillman, Smyth, & Christian, 2008). To serve both consumer marketing and academic interests, various commercial venders now provide the technical means to collect online survey data from the general public and specific subgroups. However, an important question about these commercial sources for conducting survey research is the extent to which they provide representative samples of their targeted populations.

This *Research in Brief* summarizes the results of a comparative study of different commercial platforms for generating online sampling frames and the representativeness of their samples. Multiple national surveys were conducted using *Survey Monkey*, *Qualtrics*, and *Mechanical Turk* as the source for recruiting survey respondents. By comparing the sociodemographic profile of the sample respondents obtained from each platform with national census data, this study provides some evidence of the representativeness of different online strategies for recruiting and selecting potential survey respondents.

The final section of this report describes the limitations of the study and offers some particular ways that researchers may improve the generalizability of online survey results.

HIGHLIGHTS

- Numerous commercial sites provide national sampling platforms for conducting online surveys in the fields of consumer marketing and academic research.
- A comparison of the observed sample results from 3 sampling platforms indicates that the respondents' demographic profile from online surveys departs somewhat from their respective profile in the U.S. adult population.
- Depending on the purpose of the survey research, the average discrepancy rate of 5 to 10% between the particular demographic characteristics of online respondents and their known distribution in the U.S. population may or may not be problematic.
- The efficiency and affordability of online sampling approaches provide a practical alternative for surveys that require regional, national or international samples.

Commercial Platforms for Creating and Launching Online Surveys

The increased use of internet surveys for consumer marketing and academic purposes has led to a rise in commercial businesses that provide technical support for the design and implementation of these surveys. Numerous companies provide an array of formats for designing survey questions and, for a fee, offer specific "panels" of potential survey respondents who fit the particular target population of their client. The fee for these services often increases in direct proportion to the sample size and the level of specificity of the particular target population (e.g., a national sample of 500 male executives in specific professional fields would be more expensive than a simple national sample of 500 adults). Popular venders for internet survey construction and sample selection include *Survey Monkey* (www.surveymonkey.com), *Qualtrics* (www.qualtrics.com), and *Mechanical Turk* www.mturk.com). Similar

services are provided by a host of other companies (for a list of businesses and organizations that currently provide online survey projects, see the following website: www.surveypolice.com/countries/united-states).

For purposes of achieving representative samples from internet surveys, these companies employ various platforms and recruiting strategies to develop their sampling frames. For example, *Survey Monkey* creates a panel of eligible respondents by contacting individuals who have previously completed an electronic survey on their site. In contrast, *Qualtrics* outsources the process of recruiting participants to other firms. Created by Amazon.com in 2005, *Mechanical Turk* uses their labor workforce as a basis for their sampling frames. Their workforce-based pool of eligible respondents is composed of more than 500,000 individuals from 190 countries with most workers residing in the United States and India (Ipeirotis, 2010; Paolacci & Chandler, 2014.)[1]

Although systematic research is not available on the sampling frame panels provided by *Survey Monkey* and *Qualtrics*, some published studies have been conducted on the representativeness of samples generated through *Mechanical Turk*. These studies have found that the demographic profile of *Mechanical Turk*'s samples are "at least as representative of the U.S. population" and "at least as diverse and more representative of non-college populations" than those of typical internet and traditional samples (Buhrmester, Kwang, & Gosling, 2011, p. 5; Paolacci, Chandler, & Ipeirotis, 2010, p. 414). In terms of data quality, previous research has found (1) no evidence that survey data is of lower quality when collected on *Mechanical Turk* and (1) no effect of varying compensation levels on data quality (Buhrmester et al., 2011; Paolacci et al., 2010).

To further investigate the representativeness of samples derived from different internet sampling platforms, we conducted a set of national surveys through *Survey Monkey*, *Qualtrics*, and *Mechanical Turk*. We then compared the obtained demographic profile of respondents from each platform with estimates of these corresponding profiles provided by the 2010 U.S. Census.

Methodology

During the first week of June 2014, online surveys were administered to national samples of U.S. residents over 18 years of age, as part of a larger study on attitudes about aerial drone activities (Miethe, Lieberman, Sakiyama, & Troshynski, 2014). Survey samples were provided by three commercial survey platforms: (1) *Survey Monkey* (n = 154), (2) *Qualtrics* (n = 179) and (3) *Mechanical Turk* (n = 304). A total of 636 surveys were completed within this one-week period. Major demographic profiles of respondents across each sampling frame were compared to U.S. population estimates. These results are summarized below.

Comparison of Online Sampling Methods

Overall, the online samples in this study are comprised of a mid-aged (30-59 years old) and younger demographic group of respondents. A majority had some college education or a Bachelor's degree and identified with liberal perspectives. These online respondents typically resided in medium sized urban areas with populations that ranged from 50,000 to 1 million residents, and reported incomes ranging from $25,000 to $75,000 annually (see Tables 1-3).

However, none of the samples produced a group of respondents that uniformly approximated the U.S. population profile. Significant differences also emerged across samples on most of the sociodemographic characteristics that were examined.

Demographic Factors (see Table 1)

Gender. The samples from *Survey Monkey* and *Qualtrics* contained virtually equal proportions of female respondents (51% and 52%, respectively), but female respondents were less representative in the *Mechanical Turk* sample (44%). A close approximation to the gender distribution in the adult population is also found by combining the results of the three samples. The Combined Average (n = 636) was computed by adding each sampling frame proportion together, giving each equal weight, and dividing it by the total number of sampling frames (n = 3).

Age. *Mechanical Turk* provided younger survey respondents, with nearly half (47% being under 30 years old, compared to about 17% in the *Survey Monkey* sample and only 6% among *Qualtrics* respondents. Conversely, the proportion of respondents aged 60 or older was grossly underrepresented in the *Mechanical Turk* sample (5% vs. 24% in the U.S. population) and over

estimated among the *Survey Monkey* respondents (32% vs. 24%).

Race and Ethnicity. All three sampling panels produced samples that were predominantly White, and had disproportionately lower proportions of Hispanic participants. Across all categories for these variables, the *Mechanical Turk* sample most closely reflected the racial/ethnic distribution of the U.S. population. The *Survey Monkey* sample was the least representative of African-Americans (3% vs. 14% in the U.S. population) and *Qualtrics*' sample was the least representative of Hispanics (4% vs. 17% in the U.S. population). Even in the combined sample, a substantial gap existed between the sample proportions within these racial/ethnic categories and their population estimates.

Acquired Demographic Characteristics (see Table 2)

Education. All three samples were comprised of disproportionately well-educated survey participants. *Survey Monkey*'s respondents were especially overrepresentative of persons with post-graduate degrees (26% vs. 10% in the U.S. population). The *Qualtrics* sample was the most similar to U.S. Census estimates.

Income. Income ranges were more evenly distributed in the *Qualtrics* and *Mechanical Turk* samples compared to *Survey Monkey*. However, all three samples performed reasonably well at representing the proportion of U.S. population with incomes in the middle range of the spectrum ($25,000 to $100,000). The sample estimates were most discrepant from U.S. census data at the $100,000 or more level. The largest discrepancy was found among *Survey Monkey* respondents, with one-third (33%) reporting this level of annual income compared to an estimated 22% in this category among the U.S. population.

Table 1: Sample Estimates of Population Values for Intrinsic Demographic Factors by Survey Method

	Survey Monkey	Qualtrics	Mechanical Turk	Combined Average	2010 Census Estimates
Gender					
Female	**51.0%**	52.0%	43.8%	48.9%	50.8%
Male	**49.0%**	48.0%	56.3%	51.1%	49.2%
Age Range					
18 to 29	17.1%	6.1%	46.5%	**23.2%**	22.2%
30 to 59	51.3%	68.2%	48.9%	**56.1%**	53.9%
60 and older	31.6%	**25.7%**	4.6%	20.6%	23.9%
Race					
American Indian or Alaskan Native	2.1%	1.2%	1.1%	**1.5%**	1.6%
Asian	2.1%	**4.7%**	7.4%	**4.7%**	5.8%
Black or African American	2.8%	**9.9%**	9.5%	7.4%	13.7%
Native Hawaiian or Other Pacific Islander	0.7%	0.0%	**0.4%**	**0.4%**	0.3%
White or Caucasian	91.7%	84.3%	**79.9%**	85.3%	76.3%
Other	0.7%	0.0%	**1.8%**	0.83%	2.3%
Ethnicity					
Hispanic	4.7%	3.9%	**6.1%**	4.9%	16.9%
Non-Hispanic	95.3%	96.1%	**93.9%**	95.1%	83.1%

Notes: Red cell values represent sample estimates that are closest to U.S. population values.

Table 2: Sample Estimates of Population Values for Acquired Demographic Factors by Survey Method

	Survey Monkey	Qualtrics	Mechanical Turk	Combined Average	2010 Census Estimates
Educational Attainment					
Less than high school	0.0%	**1.1%**	0.7%	0.6%	13.2%
High school graduate or the equivalent (e.g., GED)	12.8%	**24.0%**	11.2%	16.0%	30.0%
Some college	22.8%	36.9%	38.5%	**32.7%**	28.6%
College graduate	38.3%	**26.8%**	40.8%	35.3%	18.4%
Post-graduate degree (e.g., MA, MS, JD, MBA, MD, PhD)	26.2%	11.2%	**8.9%**	15.4%	9.8%
Annual Household Income					
Less than $25,000	8.8%	20.1%	**26.1%**	18.3%	24.4%
$25,000 to $50,000	**23.8%**	30.2%	35.3%	29.8%	24.2%
$50,000 to 75,000	**18.4%**	26.3%	19.5%	21.4%	18.0%
$75,000 to $100,000	16.3%	13.4%	10.2%	**13.3%**	11.9%
$100,000 or more	32.7%	10.1%	8.9%	**17.2%**	21.6%
Political Affiliation					
Democrat	**39.2%**	43.5%	49.0%	43.9%	34.0%[a]
Republican	**21.6%**	17.5%	11.9%	17.0%	39.0%[a]
Independent	39.2%	**39.0%**	39.1%	39.1%	26.0%[a]
Marital Status					
Married	59.3%	**53.1%**	40.7%	51.0%	56.1%
Unmarried	40.7%	**46.9%**	59.3%	49.0%	43.9%

Notes: Red cell values represent sample estimates that are closest to U.S. population values. [a]National estimates for political affiliation taken from Politico Voter Affiliation Poll from May 2, 2014.

Political Affiliation. Compared to U.S. Census estimates, all three samples produced substantially higher proportions of Democrat respondents and a lower proportion of Republicans and Independents. The *Survey Monkey* panel was the most representative estimate of the nation's political orientations (see Table 2).

Marital Status. There was considerable variability between the samples in their estimates of marital status. According to U.S. Census data, half (56%) of the U.S. population is married. The *Qualtrics* sample (53% are married) provided the best estimate of this population value.

Residential Characteristics (see Table 3)

Type of Home/Dwelling. Compared to other methods, *Mechanical Turk* respondents were most closely matched to U.S. population estimates on the type of home/dwelling. The *Survey Monkey* and *Qualtrics* samples were most representative of residents living in multi-unit dwellings, but these samples also contained a disproportionately higher number of respondents who lived in single-unit dwellings.

Urbanicity. Across all three samples, most of the respondents resided in medium sized urban areas that contained 50,000 to 1 million residents. However, the overall percentage of participants living in medium sized urban areas (defined as 50,000 – 1 million) and rural areas (less than 2,500) was lower than U.S. estimates. Compared to U.S. population estimates, the sample proportions of people living in "Urban Clusters" (2,500 to 50,000) and "Large Urban Areas" (> 1,000,000) were disproportionately higher for each sampling platform. Combining the sample results did little to provide a more representative sample of the U.S. population on this variable.

Table 3: Sample Estimates of Population Values for Residential Characteristics by Survey Method

	Survey Monkey	Qualtrics	Mechanical Turk	Combined Average	National Composition
Type of Home/Dwelling Structure					
Single-Unit Dwelling	79.5%	78.5%	**67.7%**	75.2%	69.0%
Duplex	3.3%	4.0%	**6.0%**	4.4%	7.9%
Multi-Unit Dwelling	**17.2%**	17.5%	26.3%	20.3%	17.0%
Type of Living Area					
Large Urban Area (greater than 1 million population)	31.8%	**20.1%**	22.7%	24.9%	13.1%
Medium Size Urban Area (50,000 to 1 million population)	34.4%	**36.9%**	36.5%	35.9%	58.1%
Urban Clusters (2,500 to 50,000 population)	**24.5%**	27.4%	25.7%	25.9%	9.5%
Rural Area (less than 2,500 population)	9.3%	**15.6%**	15.1%	13.3%	19.3%

Notes: Red cell values represent sample proportions that are closest to U.S. population estimates.

Comparing Sampling Platforms Based on Discrepancy Scores

A simple overall measure of the population representativeness of different sampling platforms involves the computation of the average "discrepancy scores" between sample estimates and their known population values. In particular, for each of the 3 sampling frames, scores were computed by taking the sum of the differences between the sample frequencies and the population frequencies. These average discrepancy scores between sample and population values for each category of the demographic characteristics and by the type of sampling method are summarized in Table 4.

When examining their convergence with national estimates, Table 4 reveals that the sampling platforms provided by *Survey Monkey* (SM) and *Qualtrics* (QT) produced the most representative samples of the U.S. population's intrinsic demographic profile (i.e., gender, age range, race and ethnicity). The average discrepancy per category for these two sampling methods was 5.6 and 5.9% (see Table 4). *Qualtrics* yielded the lowest average discrepancy rate (8.3%) across categories of acquired demographic characteristics (i.e., educational attainment, annual household income, political affiliation and marital status).

Both *Qualtrics* and *Mechanical Turk* (MT) had the lowest average discrepancies with U.S. population estimates for residential attributes (i.e., type of dwelling, urban/rural living area). However, sample estimates of these residential attributes were generally the most discrepant with known population values across all three sampling methods, with average error rates ranging from 9 to 12%.

As shown in the last column of Table 4, the representativeness of sample estimates of U.S. population values was improved in some cases by combining the three separate methods. In particular, sample estimates of intrinsic demographic attributes varied from known population values by only 4% when the results of the three methods were combined. In contrast, the average error rate was 10% for estimating known population values for residential attributes regardless of the particular sampling platform (or combination of platforms) utilized.

Table 4: Average Discrepancy Scores by Method

	SM[b]	QT[b]	MT[b]	Combined
Attributes[a]				
Intrinsic	5.6	5.9	7.3	**4.0**
Acquired	8.8	**8.3**	12.4	9.0
Residential	11.8	**9.1**	9.2	10.0
Average Score	8.7	**7.3**	9.1	7.7

Notes: Red cell values represent sample proportions that are closest to U.S. population estimates. [a]See Tables 1-3 for specific characteristics within the particular categories of *intrinsic, acquired,* and *residential* attributes. [b]SM represents Survey Monkey, QT represents Qualtrics, and MT represents Mechanical Turk.

Implications for Future Research Studies

Based on three different sampling platforms used in the current study, two interrelated conclusions can be reached about online survey methods and population representation. First, even without post-stratification weighting and other statistical adjustments for potential sampling bias, the different sampling platforms (*Survey Monkey*, *Qualtrics*, and *Mechanical Turk*) provided samples with specific demographic attributes that are often within a 10% range of their corresponding values in the U.S. population. Second, compared to traditional survey methods (i.e., mail/telephone surveys, personal interviews), these online platforms provide an extremely efficient and inexpensive method for collecting national survey data. Depending upon the ultimate purpose of a study (e.g., consumer marketing, exploratory/ confirmatory research), concerns about a 10% error rate and sampling biases(e.g., due to distinct characteristics of internet respondents) may or may not be problematic.

However for many applications, the advantages of online surveys (e.g., efficiency of data collection, lower economic costs, and “acceptable” approximations to population profiles) far exceed their disadvantages in terms of external validity.

For improving the representativeness of online surveys, a number of statistical adjustments and alternative methods have been proposed in past research. These include the use of post-stratification weighting and propensity score matching to further enhance the potential representativeness of the selected samples (see Loosveldt & Sonck, 2008). In addition, the increased availability and use of specific survey panels for online research now offers a more comprehensive method for selecting representative samples of particular groups. As internet use becomes even more entrenched in contemporary society, the current problems with sampling biases due to differential access to this technology will likely dissipate over time. Under these conditions, well-designed online surveys will increasingly offer a valuable method for consumer marketing and academic research.

END NOTES

i. The level of compensation (or "reward") for individual's participation in these internet surveys varies widely across companies and topics and ranges from $0.01 to $40 per survey (Paolacci et al., 2010; Goodman, Cryder & Cheema, 2013.)

References

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science, 6, 3-5.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2008). Internet, Mail, and Mixed-Mode Surveys, The Tailored Design Method. Hoboken, NJ: John Wiley & Sons.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. Journal of Behavioral Decision Making, 26, 213-224.

Ipeirotis, P. G. (2010). Demographics of Mechanical Turk.

Loosveldt, G. & Sonck, N. (2008). An evaluation of the weighting procedures for an online access panel survey. Survey Research Methods, 2, 93-105.

Miethe, T. D., Lieberman, J. D., Sakiyama, M. & Troshynski, E. I. (2014). Public Attitudes about Aerial Drone Activities: Results of a National Survey. [Monograph]. State Data Brief Center for Crime and Justice Policy: Las Vegas, NV. CCJP June 2014, 02.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. Current Directions in Psychological Science, 23, 184-188.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon's Mechanical Turk. Judgment and Decision Making, 5, 411-419.

Politico Voter Party Affiliation Poll - May 2, 2014. (2014). Retrieved September 12, 2014.

UNLV | CENTER FOR CRIME AND JUSTICE POLICY

CENTER FOR CRIME AND JUSTICE POLICY

STATE DATA BRIEF SERIES

This report is part of the "Research in Brief" series produced by the Center for Crime and Justice Policy at the University of Nevada, Las Vegas. The Center is housed in the Department of Criminal Justice, which is located in the Greenspun College of Urban Affairs. Research in Briefs are modeled after the Bureau of Justice Statistics' Special Reports and Bulletins.

The Briefs provide summaries of various criminal justice system practices in Nevada over time, and highlight differences between Nevada and other states. These reports cover all aspects of the criminal justice system, including trends in crime and arrests, police practices, prosecution, pretrial activities, adjudication, sentencing, and corrections. Although Research in Briefs typically focus on criminal justice issues within Nevada, these reports may focus on national issues as well.

Research in Briefs are designed to provide members of the general public, local officials, community organizations, and media outlets a concise and objective profile of current crime and criminal trends in Nevada and elsewhere. These briefs may serve as a foundation for informed discussions of future crime control policies and practices.

CONTACT INFORMATION

Questions of comments about the information contained in this report, data used to generate this report, or about other resources available related to this topic should be addressed to:

Terance D. Miethe, Ph.D.
Research in Brief Project Coordinator
Center for Analysis of Crime Statistics
University of Nevada, Las Vegas
4505 Maryland Parkway - Box 5009
Las Vegas, NV 89154-5009

Phone: 702-895-0236
Fax: 702-895-0252
Email: miethe@unlv.nevada.edu

Previous Research in Briefs

A Comparison of Different On-Line Sampling Approaches for Generating National Samples

Aerial Drones, Domestic Surveillance, and Public Opinion of Adults in the United States

Arrest-Related Deaths in Nevada, 2009-2011

Arson Trends in Nevada, 1997-2006

Auto Theft in Nevada, 1994-2008

Burglary Trends in Nevada, 1990-2007

Capital Punishment in Nevada, 1977-2008

Clearance Rates in Nevada, 1998-2009

Communication Intercepts Authorized in Nevada, 1997-2008

Comparison of Different On-Line sampling Approaches for Generating National Samples

Criminal Victimization in Nevada, 2008 Criminal Victimization in Nevada, 2011

Deaths in Custody in Nevada, 2001-2006

Impact of Foreclosures on Neighborhood Crime in Nevada, 2006-2009

Justice Assistance Grant (JAG) Program in Nevada, 2005-2010

Nevada vs. U.S. Residents Attitudes Towards Surveillance Using Aerial Drones

Patterns in School Violence in Nevada

Public Attitudes about Aerial Drone Activities: Results of a National Survey

Rape and other Sex Offenses in Nevada, 1990-2007

UNLV | CENTER FOR CRIME AND JUSTICE POLICY

Center for the Analysis of Crime Statistics
Department of Criminal Justice
University of Nevada, Las Vegas
4505 S. Maryland Pkwy - Box 5009
Las Vegas, NV 89154-5009

POSTAGE REQUIRED