

UCLA

UCLA Previously Published Works

Title

Use of Internet panels to conduct surveys

Permalink

<https://escholarship.org/uc/item/0jd059zk>

Journal

Behavior Research Methods, 47(3)

ISSN

1554-351X

Authors

Hays, Ron D

Liu, Honghu

Kapteyn, Arie

Publication Date

2015-09-01

DOI

10.3758/s13428-015-0617-9

Peer reviewed

Use of Internet panels to conduct surveys

Ron D. Hays¹ · Honghu Liu² · Arie Kapteyn³

© Psychonomic Society, Inc. 2015

Abstract The use of Internet panels to collect survey data is increasing because it is cost-effective, enables access to large and diverse samples quickly, takes less time than traditional methods to obtain data for analysis, and the standardization of the data collection process makes studies easy to replicate. A variety of probability-based panels have been created, including Telepanel/CentERpanel, Knowledge Networks (now GfK KnowledgePanel), the American Life Panel, the Longitudinal Internet Studies for the Social Sciences panel, and the Understanding America Study panel. Despite the advantage of having a known denominator (sampling frame), the probability-based Internet panels often have low recruitment participation rates, and some have argued that there is little practical difference between opting out of a probability sample and opting into a nonprobability (convenience) Internet panel. This article provides an overview of both probability-based and convenience panels, discussing potential benefits and cautions for each method, and summarizing the approaches used to weight panel respondents in order to better represent the underlying population. Challenges of using Internet panel data are discussed, including false answers, careless responses, giving the same answer repeatedly, getting multiple surveys from the same respondent, and panelists being

members of multiple panels. More is to be learned about Internet panels generally and about Web-based data collection, as well as how to evaluate data collected using mobile devices and social-media platforms.

Keywords Internet panels · Survey research

Baker et al. (2013) referred to the growth of panels for data collection as “one of the most compelling stories of the last decade” (p. 715). The use of Internet panels to collect survey data is increasing because it is cost-effective, enables access to large and diverse samples quickly, takes less time than traditional methods to get data back for analysis, and the standardization of data collection process makes studies easy to replicate. Internet panels for research can be traced back to Willem Saris (http://en.wikipedia.org/wiki/Willem_Saris), a sociology professor at the University of Amsterdam.

Probability-based Internet panels

When the first modems came on the market in 1985, Saris realized that one could create a computer-assisted data collection system without interviewers. Saris and de Pijper (1986) developed a working system for this purpose, the so-called *Telepanel*. In that time, a random sample of the population was provided with home computers and modems, and eventually with a telephone connection if it was not in the house. The experiences with this first system for interviewing without interviewers, before the Web existed, are summarized by Saris (1998). This system was bought by the Dutch Gallup organization and turned into the first nationwide computer-based panel for data collection in 1986. In 1991, the University of Amsterdam started, with the support of the

✉ Ron D. Hays
dr.ronhays@gmail.com

¹ Division of General Internal Medicine & Health Services Research, University of California, 911 Broxton Avenue, Los Angeles 90024, CA, USA

² Division of Public Health and Community Dentistry, School of Dentistry, University of California, Los Angeles, CA, USA

³ Center for Economic and Social Research, University of Southern California, Los Angeles, CA, USA

Dutch Science Foundation (NWO), a larger panel (of about 3,000 individuals). After 5 years this panel was taken over by the Tilburg University Center for Economic Research. That panel, *CentERpanel*, is the oldest academic Internet probability panel in the world.

A variety of other probability-based Internet panels were created following the success of the panel initiated by Saris. In 1999, Knowledge Networks (now GfK KnowledgePanel) created a panel of about 55,000 individuals using address-based sampling for recruitment: www.gfk.com/us/Solutions/consumer-panels/Pages/GfK-KnowledgePanel.aspx. Advances in methods to improve representation of cell phone numbers when conducting random-digit dialing have appeared (Hu, Pierannunzi, & Balluz, 2011; Voigt, Schwartz, Doody, Lee, & Li, 2011). But the GfK KnowledgePanel is recruited using address-based sampling from the US Postal Service's Computerized Delivery Sequence File, described as "essentially a complete list of all US residential addresses, including those that are cell phone only and often missed in RDD sampling": www.gfk.com/us/Solutions/consumer-panels/Pages/GfK-KnowledgePanel.aspx?gclid=CML0_t2Qy8ICFUmUfgodUIsAHg.

The Longitudinal Internet Studies for the Social Sciences (LISS) in the Netherlands used population registry-based sampling and recruited respondents face-to-face and by telephone to obtain a panel of about 7,500 individuals in 2007: www.lissdata.nl/lissdata/. In the year before, the American Life Panel was developed in the US. This panel included approximately 6,000 adults recruited by random-digit dialing and by face-to-face and address-based sampling: <http://mmicdata.rand.org/alp/>. More recently, the Understanding America Study panel of about 2,000 individuals was created using address-based sampling: http://static.usc.edu/data_toolbox/understanding_america_study.

Despite the advantage of having a known denominator (sampling frame), the probability-based Internet panels tend to have low recruitment participation rates. About 6%–7% of the targeted Knowledge Networks panel respondents were excluded because they were not in the service area of a WebTV Internet service provider. Addresses were obtained for about 60% of the sampled telephone numbers. About 89% of those in the eligible random-digit dial sample were contacted for initial telephone interviews, 56% agreed to participate in the initial telephone interview and join the panel, and 72% of those installed the required WebTV device in their homes (Chang & Krosnick, 2009). Similarly, about 14% of eligible KnowledgePanel households stated a willingness to become a panel member, and of those, about three quarters followed through (see www.knowledgenetworks.com/accuracy/spring2010/disogra-spring10.html), for a net sign-up rate of about 10%. Also, the Understanding America Study achieves a sign-up rate of 15–20% (recruitment is still ongoing).

The highest recruitment participation rate, 48%, has been achieved by the LISS panel in the Netherlands. The recruiting procedure was based on drawing households from the national population registry and next following a process of initial invitation letters with an included prepaid incentive, follow-up phone calls, as well as face-to-face recruiting (Scherpenzeel & Das, 2011). The adopted process and incentive levels were based on an initial experiment aimed at maximizing the recruitment participation rate (Scherpenzeel & Toepoel, 2012).

Non-probability-based (convenience) Internet panels

Some have argued that there is little practical difference between opting out of a probability sample and opting into a nonprobability sample (Rivers, 2013). Indeed, a plethora of Internet panel vendors rely on non-probability-based recruitment, and many researchers use those panels. For example, the NIH Toolbox project developed a multidimensional set of brief measures assessing cognitive, emotional, motor, and sensory function from the ages of 3 to 85 years old. The study participants were part of the Delve, Inc., panel, assembled using online self-enrollment, enrollment through events hosted by Delve, and telephone calls from market research representatives (Gershon et al., 2013).

The composition of these nonprobability Internet (convenience) panels is known to differ from that of the underlying population. It is estimated that up to one third of the US adult population does not use the Internet on a regular basis (Baker et al., 2013). Panel members tend to be more educated and to have higher socioeconomic status than non-panel-members (Craig et al., 2013). The response rates for members of convenience panels tend to be low. Baker et al. (2013) suggested that response rates are often 10% or lower. As a result, many users of convenience panels utilize a quota-sampling approach by targeting respondents with particular demographic and other characteristics, and use poststratification adjustments (weights) to compensate for noncoverage and nonresponse. Panel respondents are weighted to match a target marginal distribution (e.g., the US Census).

An analysis by Schonlau, van Soest, Kapteyn, and Couper (2009) of 11,279 individuals age 55 and older in the 2002 Health and Retirement Study showed that the 30% who reported Internet access differed substantially in characteristics from those without Internet access (see Table 1). Propensity score weights were created by predicting Internet access from race/ethnicity, gender, education, age, marital status, income, owns house, and self-rated health. The weighted estimates were more similar to the underlying population, but there were still nontrivial differences. Yeager et al. (2011) concluded that probability sample surveys were consistently more accurate

Table 1 Health and Retirement Study 2002: 55-and-older full sample versus those with Internet access

	Full Sample	Internet Access	Weighted Internet Access
High blood pressure	55%	44%	52%
Depressed	19%	11%	15%
Difficulty dressing	9%	4%	7%
Difficulty walking several blocks	31%	15%	27%

Data are from Schonlau et al. (2009)

than non-probability-sample surveys, even after poststratification weighting of the data.

A recent study comparing responses to the Patient-Reported Outcomes Measurement Information System (PROMIS) global health items across four surveys found comparable estimates of physical and mental health, despite the differences in survey sampling (probability vs. nonprobability), although the National Health Interview Survey yielded more positive estimates of health due to the interview mode of data collection (Riley, Hays, Kaplan, & Cella, 2014). Chang and Krosnick (2009) found that nonprobability Internet data collection yielded the most accurate self-reports from the most biased sample, but that the probability Internet sample displayed the best combination of sample composition and self-report accuracy.

Approaches to weighting convenience Internet panels

An example of a successful use of a convenience panel to represent the “general population” is the initial PROMIS data collection and weighting described by Liu et al. (2010). The study team set target quotas from the Polimetrix (now YouGov) convenience panel of over 1 million members: 50% female, 20% from each of five age groups (18–29, 30–44, 45–59, 60–74, and 75 and older), 12% African-American, 12% Hispanic, and 10% with less than high school education. The demographics of the resulting respondents versus the 2000 Census are shown in Table 2. The PROMIS sample had a greater percentage of females and was much more educated and a little older than the US general population (2000

Census). Poststratification adjustment (with analytic weights) was used to compensate for nonresponse and unequal selection probabilities. The PROMIS sample was weighted to have the same distributions on six demographic variables (gender, age, race/ethnicity, education, marital status, and income) using an iterative proportional fitting or raking method. Raking matches cell counts to the marginal distributions through cell-by-cell adjustments, repeated until there is convergence between the weighted sample and the US Census distributions.

Table 2 shows that the weighted PROMIS sample was similar to the US 2000 Census on the demographic characteristics. The mean scores on self-rated general health (“In general, how would you rate your health?” 5 = *excellent*, 4 = *very good*, 3 = *good*, 2 = *fair*, 1 = *poor*) for the PROMIS weighted sample was 3.42, as compared to 3.56, 3.50, and 3.52 for the 2004 Medical Expenditure Panel Survey, 2001–2002 National Health and Nutrition Examination Survey, and the 2005 Behavioral Risk Factor Surveillance System, respectively.

Although one can leverage the distribution of demographics of a sample to the target population, the weighting of convenience samples does not always yield complete comparability of the outcome measures to a target population. In the extreme, responses from convenience panels may differ so much from the target population that no adjustment can make them look similar. For example, the PROMIS 2010 recentering project collected data from members of the OP4G convenience panel ($n = 2,996$) who had demographic characteristics similar to the 2010 Census, but the respondents reported worse health by about a half a standard deviation on PROMIS domains as compared to the PROMIS Wave 1 general population sample. The weighted mean on the self-rated health item mentioned above (“In general, how would you

Table 2 Demographic characteristics of 2000 Census versus Polimetrix respondents in patient-reported outcomes measurement and information system (PROMIS) study

	2000 Census	PROMIS Sample	Weighted PROMIS Sample
% Female	52	55	52
% Hispanic	11	13	11
% African-American	11	10	11
% Less than high school	20	3	20
% High school/GED graduates	29	19	29
% greater than high school	51	78	51
Mean age	45	59	45

rate your health?” 5 = *excellent*, 4 = *very good*, 3 = *good*, 2 = *fair*, 1 = *poor*) was 3.24, versus the weighted mean of 3.42 observed for the PROMIS Wave 1 sample (noted above). Similarly, the OP4G sample had an average Health Utilities Index (HUI-3) score of only 0.54, whereas the median HUI-3 in the US noninstitutionalized population 35–89 years old was estimated to be 0.88 using random-digit dialing (Fryback et al., 2007). The telephone mode of data collection in the Fryback et al. study yielded HUI-3 scores about 0.10 higher than did mail sampling (Hays et al., 2009), but the mode effect cannot account for the much lower HUI-3 scores for the OP4G sample.

The PROMIS project also collected data from a sample of 640 adult Spanish-speaking Latinos in the Toluna Internet panel and found that only 2% selected Spanish as their language of preference, and they reported higher levels of education and lower levels of acculturation than the 2010 Census data for Latinos (Paz, Spritzer, Morales, & Hays, 2013). It is unlikely, given the characteristics and sample size, that weighting these data would produce compatible marginal distributions of health matching that of the US general Spanish population.

Challenges of using Internet panels

Data integrity is a concern when dealing with data collected from Internet panels. Respondents may engage in a variety of less-than-optimal strategies to get through surveys so that they can get whatever rewards or incentives are offered. This can lead to a variety of undesirable responses, such as false answers, answering too fast, giving the same answer repeatedly (also known as *straight-lining* or *satisficing*), and getting multiple surveys completed by the same respondent. To help improve the quality of the data, Liu et al. (2010) excluded respondents with high levels of missing data (e.g., who completed fewer than half of the items), who completed items quicker than 1 s per item, or who gave the same response to ten consecutive items. Panel companies often have procedures in place such as e-mail address and IP address verification to ensure the identities of individuals that join and to minimize duplicate representation on the panel. Another practice is to provide feedback to respondents who appear to be less than serious in responding to questions—for example, by noting that they are rushing through surveys or that they often seem to give the same answer.

Another issue to confront is the fact that the panelists on convenience panels participate on average in 2.7 panels (Tourangeau, Conrad, & Couper, 2013). Indeed, Miller (2006) estimated that 30% of Internet surveys are completed by 0.25% of the US eligible population. A study that recruited US adults from seven panel vendors using identical quotas found variability in the response rates and estimated that

different panel vendors appeared to draw 15%–25% of their samples from a common pool (Craig et al., 2013).

Conclusions and future study

Whether panels (convenience or probability-based) represent the underlying population is not a concern unless the research project needs precise estimates of population values or unbiased estimates of relationships between the variables of interest (although associations are typically not as affected). When the objectives of the study are different, the use of panels to select samples is similar to a large body of research based on undergraduates, to patients receiving care at select sites of care, or to samples that are not representative of a true underlying population. For these purposes, the use of convenience panels has the advantages of relatively low cost, greater speed of data collection, and the ability to obtain large numbers of respondents in subgroups of interest. Similarly, methodological and psychometric research that requires a diverse but not necessarily representative sample can benefit greatly by the use of Internet panel data sources.

When there is value in representing a defined underlying population, then convenience Internet panels may be useful if the data can be weighted to compensate adequately for coverage errors and selection bias. As we described above, in some cases even convenience Internet panels can be used as the basis of population norms. But there is no guarantee that any particular convenience Internet panel will be suitable for this purpose. Probability-based panels have the major advantage of having a known denominator, but the recruiting rate for these panels is often low. Chang and Krosnick (2009) compared a convenience panel (Harris Interactive) with a probability-based panel (Knowledge Networks) and concluded that “probability samples were more representative of the nation than the nonprobability sample in terms of demographics . . . even after weighting” (p. 641). But the average errors of estimates of demographic variables relative to the 2000 Current Population Survey were actually very similar for Knowledge Networks and Harris Interactive (Table 3).

Table 3 Average errors for Harris Interactive (convenience panel) and Knowledge Networks (probability-based panel) versus 2000 current population survey estimates

	Harris Interactive	Knowledge Networks
Education	6%	3%
Income	5%	6%
Age	2%	2%
Race	1%	2%
Gender	0%	1%

The weighted data are reported in Table 3 of Chang and Krosnick (2009)

This is consistent with the suggestion by Rivers (2013) that there is little practical difference between opting out of a probability sample and opting into a nonprobability sample.

No hard-and-fast rules determine when convenience panels are adequate for use in population inference or when response rates to probability Internet panels will be high enough to assume unbiased estimates. For instance, bias in the estimate of a simple mean is a function of the covariance between the propensity to respond and the variable of interest, as well as the response propensity of the sample members (Bethlehem, 2002). Meta-analysis suggests that the relation between response rate and bias is not very strong in most cases (Groves & Peytcheva, 2008). Gutsche, Kapteyn, Meijer, and Weerman (2014) used the American Life Panel (with a recruitment participation rate of 10%–15%) to forecast the popular vote in the 2012 presidential election. Their forecast of the final tally was one of the very best among some 25 US polling firms, which may suggest that response propensity and one's political preference were at most weakly correlated.

Survey research has entered a new era, with less emphasis on interviews and increasing use of new technologies for data collection (Link et al., 2014). More needs to be learned about the strengths and disadvantages of probability-based and convenience Internet panels, as well as the use of Web-based data collection in general (Bergeson, Gray, Ehrmantraut, & Hays, 2013; Brown, Serrato, Hugh, Kanter, & Hays, submitted). There will also be future opportunities to evaluate data collected using mobile devices and social-media platforms.

Author note This article was presented at the 2014 Society for Computers in Psychology Meeting, Long Beach, CA. R.D.H. was supported in part by grants from NCI (No. 1U2-CCA186878-01), the NIA (No. P30-AG021684), and the NIMHD (No. P20-MD000182). H.L. was supported in part by a grant from NIDA (No. R34-DA031643). A.K. was supported by the NIA (Grant No. R01-AG20717).

References

- Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., . . . Tourangeau, R. (2013). Summary report of the AAPOR Task Force on Non-probability Sampling. *Journal of Survey Statistics and Methodology*, 1, 90–143. doi:10.1093/jssam/smt008
- Bergeson, S. C., Gray, J., Ehrmantraut, L. A., & Hays, R. D. (2013). Comparing Web-based with mail survey administration of the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Clinician and Group Survey. *Primary Health Care*, 3(1), 132. doi:10.4172/2167-1079.1000132
- Bethlehem, J. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. Groves, D. Dillman, R. Eltinge, & L. Little (Eds.), *Survey nonresponse* (pp. 275–288). New York: Wiley.
- Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73, 641–678. doi:10.1093/poq/nfp075
- Craig, B. M., Hays, R. D., Pickard, A. S., Cella, D., Revicki, D. A., & Reeve, B. B. (2013). Comparison of US panel vendors for online surveys. *Journal of Medical Internet Research*, 15, e260. doi:10.2196/jmir.2903
- Fryback, D. G., Dunham, N. C., Palta, M., Hanmer, J., Buechner, J., Cherepanov, D., . . . Kind, P. (2007). U.S. norms for six generic health-related quality-of-life indexes from the national health measurement study. *Medical Care*, 45, 1162–1170.
- Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N., Cook, K. F., & Nowinski, C. J. (2013). NIH Toolbox for the Assessment of Neurological and Behavioral Function. *Neurology*, 80(Suppl. 3), S2–S6.
- Groves, R. M., & Peytcheva, M. (2008). The impact of nonresponse rates on nonresponse bias. *Public Opinion Quarterly*, 72, 167–189. doi:10.1093/poq/nfn011
- Gutsche, T., Kapteyn, A., Meijer, E., & Weerman, A. (2014). The RAND Continuous 2012 Presidential Election Poll. *Public Opinion Quarterly*, 78, 233–254.
- Hays, R. D., Kim, S., Spritzer, K. L., Kaplan, R. M., Tally, S., Feeny, D., . . . Fryback, D. G. (2009). Effects of mode and order of administration on generic health-related quality of life scores. *Value in Health*, 12, 1035–1039.
- Hu, S. S., Pierannunzi, C., & Balluz, L. (2011). Integrating a multimode design into a national random-digit-dialed telephone survey. *Preventing Chronic Disease*, 8, A145. Retrieved from www.cdc.gov/pcd/issues/2011/nov/10_0230.htm
- Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Childs, J. H., & Tesfaye, C. L. (2014). *Mobile technologies for conducting, augmenting and potentially replacing surveys: Report of the AAPOR task force on emerging technologies in public opinion research* [Technical Report]. Washington, DC: American Association for Public Opinion Research. Retrieved from https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/REVISED_Mobile_Technology_Report_Final_revised10June14.pdf
- Liu, H. H., Cella, D., Gershon, R., Shen, J., Morales, L. S., Riley, W., & Hays, R. D. (2010). Representativeness of the PROMIS internet panel. *Journal of Clinical Epidemiology*, 63, 1169–1178.
- Miller, J. (2006). Online marketing research. In R. Grover & M. Vriens (Eds.), *The handbook of marketing research: Uses, misuses, and future advances* (pp. 110–131). Thousand Oaks: Sage.
- Paz, S. H., Spritzer, K. L., Morales, L. S., & Hays, R. D. (2013). Evaluation of the Patient-Reported Outcomes Information System (PROMIS®) Spanish physical functioning items. *Quality of Life Research*, 22, 1819–1830.
- Riley, W., Hays, R.D., Kaplan, R. M., & Cella, D. (2014). Sources of comparability between probability sample estimates and nonprobability web samples estimates. In *Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference*. Retrieved from www.fcsm.gov/13papers/B4_Riley_20
- Rivers, D. (2013). Comment. *Journal of Survey Statistics and Methodology*, 1, 111–117.
- Saris, W. E. (1998). Ten years of interviewing without interviews: The telepanel. In M. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls, & J. O'Reilly (Eds.), *Computer assisted survey information collection* (pp. 409–429). New York: Wiley.
- Saris, W. E., & de Pijper, W. M. (1986). Computer assisted interviewing using home computers. *European Research*, 14, 144–152.
- Scherpenzeel, A., & Das, M. (2011). True longitudinal and probability-based internet panels: Evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmarek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and research strategies* (pp. 77–104). New York: Taylor & Francis.

- Scherpenzeel, A., & Toepoel, V. (2012). Recruiting a probability sample for an online panel: Effects of contact mode, incentives, and information. *Public Opinion Quarterly*, 76, 470–490. doi:[10.1093/poq/nfs037](https://doi.org/10.1093/poq/nfs037)
- Schonlau, M. A., van Soest, A., Kapteyn, A., & Couper, M. (2009). Selection bias in web surveys and the use of propensity scores. *Sociological Methods and Research*, 37, 291–318.
- Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). *The science of Web surveys*. Oxford: Oxford University Press.
- Voigt, L. F., Schwartz, S. M., Doody, D. R., Lee, S. C., & Li, C. I. (2011). Feasibility of including cellular telephone numbers in random digit dialing for epidemiologic case-control studies. *American Journal of Epidemiology*, 173, 118–126.
- Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, 75, 709–747.