doi: 10.5191/jiaee.2019.26105

Using Non-Probability Sampling Methods in Agricultural and Extension Education Research

Alexa J. Lamm
Kevan W. Lamm
University of Georgia

Abstract

Understanding what the public thinks can guide how to target international agricultural and extension education interventions. Public opinion data can also provide insights into how the agricultural and natural resource industry communicates about emerging technologies and practices. However, the use of cellphones and the Internet have greatly reduced response rates to antiquated methods of public opinion data collection and the research world must alter its approach in response. The use of nonprobability sampling techniques has increased dramatically in public opinion research the past five years and more recently within agricultural and extension education research. To keep up with these trends, agricultural and extension education researchers must know the nuance associated with the use of nonprobability sampling techniques and how to mitigate some of the issues that may arise as a result.

Keywords: public opinion, non-probability sampling, weighting, quota sampling

Introduction

The world is changing with new technologies impacting the way people interact and communicate with one another (Ruggiero, 2000). Technological change has resulted in the use of new mediums for interaction, specifically the increased use of cell phones and ready access to the Internet. In 2018, 95% of Americans owned a cellphone of some kind and 77% owned smartphones with access to the Internet in their hand (Pew Center for Research, 2018). In addition, there has been a dramatic decline in the number of landlines in the home. The National Center for Health Statistics reported that as of December 2016 over half of American households did not have a landline (Blumberg & Luke, 2017). In addition, Americans are faced with increased demands on their time and have become more selective about when and how they can be contacted. An increase in the number of unsolicited telephone calls has resulted in people employing more sophisticated technology for screening their calls (e.g., voice mail, caller identification, call blocking and privacy managers).

Due to the ready access to the Internet, Americans use of mail services has also changed. The U.S. Postal Service has seen a dramatic decrease in First Class mail exhibited by a 28% decrease in First-Class mail revenue from 2007 to 2017 (United States Postal Service, 2018). The U.S. Postal Service projects the volume to continue to decline as Americans become less reliant upon, and responsive to, hard-copy mail versus digital communication methods such as e-mail and social media.

The changes in the way people communicate has dramatically impacted research sampling methods; especially probability-based methods that are reliant on random digit dialing and using randomly selected addresses to mail surveys. As early as 2013, Baker et al. (2013) reported that relying on landlines for research purposes may not be a true representation of a probability sample because of the differences in individuals who have a landline and those who do not. Now, pure access to a landline is impacting the use of random digit dialing as a sampling process. For those that do have landlines or if a researcher is able to access random cell phone numbers, Americans' avoidance of unsolicited calls has resulted in fewer people participating in telephone survey research. As a consequence, response rates have declined over the past decade to the point that nonresponse error and coverage error are a major concern (Baker et al., 2013). While researchers will not argue that probability sampling is ideal, we are now tasked with the challenge of collecting data from target audiences through new means (Baker et al., 2013; Hovland, 1959; Ruggiero, 2000).

The trends and observations associated with the United States are readily evident around the globe. From public perceptions of agricultural water use in a single state in the Southern United States (Lamm, Beattie, & Taylor, 2018) to public perceptions and support of locally grown food in Taiwan (Huang & Lamm, 2017), the importance of understanding how the public perceives issues is evident. Agricultural and extension educators are being increasingly called upon to not only describe and evaluate programs from a descriptive perspective, but also to describe the impacts of programs amongst participants and beyond (Lamm & Lamm, 2018). However, despite the need to show such impacts, one of the critical entry conditions to document change is absent, specifically, baseline measures. Baseline measures, particularly of a general public that is intended to be the ultimate beneficiary of educational interventions or training programs, is an invaluable tool for contemporary educators and Extension professionals (Rossi, Lipsey, & Henry, 2018).

Non-Probability Sampling Defined

Internet research has become increasingly popular due to researchers being able to reach millions of respondents online through the use of opt-in panels. Opt-in panels are groups of people that are recruited (typically by a marketing or research firm) to participate in studies through the Internet using computer-mediated survey software (Baker et al., 2013). They are often incentivized in some way for their participation. Most often this comes in the form of points an individual earns for participating in a study. Over time their points aggregate and they can cash them in for a prize, such as a gift card.

Opt-in study participant recruitment is a form of convenience sampling, also known as river sampling (Baker et al., 2013). It is important to immediately recognize that non-probability sampling is not random. People have to be willing to opt-in or sign up to be a part of a pool of individuals that may be contacted when a group needs respondents.

In recent years, the number of web-based surveys being employed to answer international agricultural and extension education research questions has increased dramatically with non-probability sampling becoming much more common. Access to the Internet, and the relatively low cost of conducting web surveys are contributing to the proliferation of surveys being conducted online. For example, Kumar Chaudhary, Lamm, and Warner (2018) used non-probability techniques to collect data from individuals using large amounts of water to gain an understanding of why they are not interested in conserving and how to deliver extension programs that will resonate and alter behavior. Beattie, Lamm, Rumble, and Ellis (2018) used non-probability sampling to gain an understanding of how the public thinks about and makes decisions regarding agricultural and natural resource technologies such as genetic modification. Ali, Ramey and Warner (2018) used the technique to explore the idea of personal norms and its effect on the conservation of water amongst the general public. Qu, Lamm and Rumble (2017) explored how to market blueberries to a local audience by testing different messages with a target population of those living in the eastern U.S. and Holt, Rumble, Telg and Lamm (2015) explored what messages are the best to reach specific audiences by integrating an experimental design into an online survey that generated a non-probability sample of the general public.

Non-probability online sampling techniques can be subject to significant biases resulting from under coverage and nonresponse. Not everyone has Internet access and there are significant demographic differences between those who do have access and those who do not. Individuals with lower incomes, less education, living in rural areas, or age 65 and older are underrepresented. For example, someone living in a rural area with dial-up Internet is probably not going to sign up or spend the time it would take them to complete a survey through a dial-up connection.

Non-probability sampling also allows for selection bias to occur based on the type of person that would opt-in to complete a survey online. As in any type of survey research there will be selection bias and it is up to the researcher to accept and/or minimize the effects. A similar effect is true for probability sampling using random digit dialing or mailed surveys. A certain type of person is willing to answer a phone survey or a mail survey just as specific individuals will be attracted to signing up for an online panel. Recognizing all social science sampling techniques introduce bias, it is the employment of the appropriate techniques to minimize its effect that can assist in getting the best population estimates possible to answer the research questions at hand.

Overcoming the Limitations of Using Non Probability Sampling

Despite concerns, research has shown that non-probability samples have yielded results that are as good as, or even better than, probability-based samples when the appropriate techniques are employed to overcome its limitations (Abate, 1998; Twyman, 2008; Vavreck & Rivers, 2008). Non-probability online sampling methods are most often used to make population estimates (Baker et al., 2013). When trying to make population estimates, it is imperative the population of interest is clearly defined demographically prior to data collection and the identification of respondents.

Using the United States (U.S.) voting public as an example, we would need to identify who makes up the voting public. We know they are 18 years of age or older and reside in all 50 states. The U.S. Census Bureau (2010) tells us there are 234,564,071 Americans over the age of 18, as of the last U.S. Census conducted in 2010. The U.S. Census also identifies that 48.5% of them are male and 51.5% are female. In addition to sex, we can use the U.S. Census data to identify demographic breakdowns for race, ethnicity, age, educational attainment and many other characteristics of our population of interest. In addition, we can use rural-urban continuum codes, based on Zip codes of respondents, to determine geographic location and the level of rurality within their environment. Questions pertaining to the demographic characteristics of interest must be included in the survey to obtain the data necessary to ensure your sample is representative.

Quota Sampling

In the best case scenario, quota sampling is used. Quota sampling is a non-probability sampling technique where the sample of individuals obtained matches the proportions of individuals for the entire population of interest (Moser & Stuart, 1953). Quotas are set up in advance to ensure the ability to obtain the best representation of your population. In the case of trying to reach the U.S. voting public, the questions pertaining to demographic breakdowns would be asked first so that, as your survey runs online, data is collected about who is entering the survey. Once a quota is reached, such as you have obtained enough white male respondents with a bachelors degree living in an urban environment, a new individual coming in to complete the survey who meets those same criteria would be automatically rejected from completing the survey.

Quota sampling is especially important if you have a very specific targeted group of individuals for your research study and do not necessarily know their demographic breakdown in advance. For example, if you are interested in obtaining a sample of homeowners in a specific geographic region that use irrigation to water their landscape to ask them how they make their watering decisions you would need to have the following quotas set before your data collection begins: the geographic location they reside in, whether or not they own their home and if they have an irrigation system. You may even want to ask them if they control that irrigation system or if someone else controls it such as a homeowner's association or a landscape management company. Since you do not have a demographic breakdown for the target audience, you will be completely reliant upon the quota sampling and need to recognize that sampling bias may be introduced.

Weighting

Weighting is another method for fine-tuning your data to be sure it matches the characteristics of your population of interest as closely as possible. There will be times when demographic elements within your sample are overrepresented or underrepresented and weighting ensures their responses are taken into consideration at the appropriate level when inferential statistics are run on the dataset (Baker et al., 2013). Weighting provides a more accurate account for the perceptions and values associated with specific demographic characteristics within your population of interest (Pasek, 2015).

In an unweighted dataset every respondent has the same amount of weight when aggregated. In essence, an unweighted dataset assumes everyone is the same. Therefore, if a weight were assigned, everyone would have a weight of one. Each response would be multiplied by a one to obtain the weighted response, with an outcome of every respondent's original response matching their weighted response.

Unfortunately, we are not always able to obtain a dataset that matches our population of interest exactly. We will use our U.S. voting public scenario again as a fictitious example for illustrative purposes. In the hypothetical dataset we find we have a Hispanic male respondent. His answers, unweighted, count as a 1.00, just like any other respondent. However, we have found the Hispanic population is slightly overrepresented in our sample – we have more than needed. As a result, all Hispanic responses in the sample are weighted, or multiplied, by .93 instead of a 1.00. We then discover that male respondents were slightly underrepresented in our sample. This same respondent, because he is male, needs to have more representation. Therefore, his response is weighted with an additional .04, as is every other male respondent in the dataset. Consequently, once the .04 is added to the original .93 a total weighting score of .97 is calculated. The same over, or under, adjustments are made for any demographic variable that is measured, and of which we have census values to compare against. While weighting is based on simple math and can be done by hand, computer programs have systems for automatically setting up weights based on multiple demographic characteristics.

Weighting is a necessary step in ensuring your sample is truly representative of your population of interest and should be taken seriously to ensure the reliability of your results. Quota sampling and weighting should go hand in hand as it is ideal to get as demographically close as possible to your population of interest and then weight to make up for minor discrepancies. However, there are times when weighting is not possible, such as the homeowner example provided above. If there is no recognizable, reputable source of a demographic profile that can be weighted against, the researcher should acknowledge this limitation and discuss how quota sampling was used to ensure the sample was as representative as possible.

Conclusions, Recommendations & Implications

The agricultural industry is at the forefront of public opinion and discourse. There are few concerns more pressing and personal than the food we eat and the sustainability of the natural environment. While for some members of the global public the greatest concern may be making sure they have enough to eat; for others, the cultivation and management practices associated with how the food was grown and harvested is of more concern. Therefore, the perceptions of the general public around the globe related to the agricultural industry is of critical importance.

Agricultural educators and extension educators are uniquely positioned to contribute and add value to the interface between the agricultural industry and the public. Sitting at the nexus between these two groups, agricultural and extension educators are uniquely positioned to serve as a bridge between interests. From a public perspective, if agricultural and extension educators can accurately measure perceptions, trends, interests, and other indicators of interest they can better prepare educational interventions. Critical needs, such as effective science communication practices, should be informed by what the empirical needs of the public are, not what they are believed to be.

Similarly, from an agricultural perspective, agricultural and extension educators can serve as a reliable source of public perception data that might otherwise be unavailable. The intent of such data is to drive and inform the actual upstream practices that will be ultimately evaluated at the market level. The more the agricultural industry can use empirical data to drive decisions, the more potential the decisions will be viewed as adding value by the public.

Although general public surveying has been around for decades the rate of technological change has made many of the previous techniques basically irrelevant (Baker et al., 2013). However, along with technology changes new opportunities have emerged. Online nonprobability opt-in sampling constitutes a third-wave of technology breakthroughs in sampling following previous breakthroughs using postal mail and later the landline telephone. This approach is one that provides both flexibility and scale around the globe. Almost anywhere there is a group of individuals with sufficient access to the Internet, nonprobability opt-in sample data may be available. Using the data in a thoughtful manner, and making sure to follow recommendations such as quota sampling and weighting, can help to ensure agricultural and extension educators are well positioned to capitalize on this emerging technique and wrestle with research questions that would have otherwise gone unanswered.

References

Abate, T. (2008). *Accuracy of online surveys may make phone polls obsolete*. The San Francisco Chronicle, D1.

Ali, A., Ramey, & Warner, L. A. (2018). Exploring the effect of personal norms and perceived cost of water on conservation. *Journal of Agricultural Education, 59*(3), 169-184. doi: 10.5032/jae.2018.03169

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., et al. (2013). *Report of the AAPOR task force on non-probability sampling.* American Association for Public Opinion Research. Retrieved from http://www.aapor.org/AM/Template.cfm?Section=Reports1&Template=/CM/ContentDisplay.cfm&ContentID=5963

Beattie, P. N., Lamm, A. J., Rumble, J. N., & Ellis, J. D. (2018). Identifying generational differences to target extension programming when discussing genetic modification. *Journal of Agricultural Education, 59*(3), 154 - 168. doi: 10.5032/jae.2018.03154

Blumberg, S. J., & Luke, J. V. (2017). *Wireless substitution: Early release of estimates from the national health interview survey July – December 2016*. National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201705.pdf

Huang, P. & Lamm, A. J. (2017, April). *Factors influencing public perceptions and support of locally grown food in Taiwan.* Presentation given at the Association for International Agricultural and Extension Education annual meeting, Minneapolis, MN.

Holt, J., Rumble, J. N., Telg, R. & Lamm, A. J. (2015). The message or the channel: An experimental design of consumers' perceptions of a local food message and the media channel used to deliver the information. *Journal of Applied Communication, 99*(4), 6 – 19.

Hovland, C. I. (1959). Reconciling conflicting results derived from experimental and survey studies of attitude change. *American Psychologist, 14*(1), 8-17.

Kumar Chaudhary, Lamm, A. J., & Warner, L. A. (2018). Using cognitive dissonance theory to theoretically explain water conservation intentions. *Journal of Agricultural Education, 59*(4), 194-210. doi: 10.5032/jae.2018.04194

Lamm, A. J., Beattie, P. N., & Taylor, M. R. (2018). Evaluating public perceptions of agricultural water use by regions to guide extension programming. *Journal of Southern Agricultural Education Research,* 68. Retrieved from http://www.jsaer.org/pdf/Vol68/2018_010%20formatted%20to%20print

Lamm, A. J., & Lamm, K. W. (2018). Considering an appreciative approach to international extension evaluation. *Journal of International Agricultural and Extension Education, 25*(4), 74-82. doi: 10.5191/jiaee.2018.25406

Moser, C. A. & Stuart, A. (1953). An experimental study of quota sampling. *Journal of the Royal Statistical Society. Series A, 116*(4), 349-405. Doi: 10.2307/2343021

Pasek, J. (2015). When will nonprobability surveys mirror probability surveys? Considering types of inference and weighting strategies as criteria for correspondence. *International Journal of Public Opinion Research, 28*(2), 269-291. doi: 10.1093/ijpor/edv016

Pew Research. (2018). *Mobile phone ownership over time*. Mobile Fact Sheet. Washington DC: Pew Research Center. Retrieved from http://www.pewinternet.org/fact-sheet/mobile/

Qu, S., Lamm, A. J., & Rumble, J. N. (2017). Marketing power berries: An importance-performance analysis of blueberry attributes. *Journal of Applied Communications, 101*(3). doi: 10.4148/1051-0834.1842

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.

Ruggiero, T. E. (2000). Uses and gratifications theory in the 21st Century. *Mass Communication & Society, 3*, 3-37.

Twyman, J. (2008). Getting it right: Yougov and online survey research in Britian. *Journal of Elections, Public Opinions and Parties, 18*, 343-354.

United States Postal Service. (2018). *A new reality: Correspondence mail in the digital age.* Office of Inspector General Report number RARC-WP-18-004. Retrieved from https://www.uspsoig.gov/sites/default/files/document-library-files/2018/RARC-WP-18-004.pdf

Vavreck, L., & Rivers, D. (2008). The 2006 cooperative congressional election study. *Journal of Elections, Public Opinion and Parties, 18*(4), 355-366.