An Inventory and Assessment of Sample Sources for Survey Research with Agricultural Producers in the U.S.

Jessica D. Ulrich-Schad  
*Utah State University, jessica.schad@usu.edu*

Shuang Li  
*Saint Ambrose University*

J. G. Arbuckle  
*Iowa State University*

Edem Avemegah  
*Utah State University*

Kathryn J. Brasier  
*The Pennsylvania State University*

Morey Burnham  
*Utah State University*

Additional works at: [https://digitalcommons.usu.edu/soca_facpub](https://digitalcommons.usu.edu/soca_facpub)

Part of the Anthropology Commons, and the Sociology Commons

See next page for additional authors


This Article is brought to you for free and open access by the Sociology and Anthropology at DigitalCommons@USU. It has been accepted for inclusion in Sociology and Anthropology Faculty Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
Authors

This article is available at DigitalCommons@USU: https://digitalcommons.usu.edu/soca_facpub/4
An Inventory and Assessment of Sample Sources for Survey Research with Agricultural Producers in the U.S.


To cite this article: Jessica D. Ulrich-Schad, Shuang Li, J. G. Arbuckle, Edem Avemegah, Kathryn J. Brasier, Morey Burnham, Anil Kumar Chaudhary, Weston M. Eaton, Wei Gu, Tonya Haigh, Douglas Jackson-Smith, Alexander L. Metcalf, Amit Pradhananga, Linda S. Prokopy, Matthew Sanderson, Emma Wade & Adam Wilke (2022) An Inventory and Assessment of Sample Sources for Survey Research with Agricultural Producers in the U.S., Society & Natural Resources, 35:7, 804-812, DOI: 10.1080/08941920.2022.2081392

To link to this article: https://doi.org/10.1080/08941920.2022.2081392
An Inventory and Assessment of Sample Sources for Survey Research with Agricultural Producers in the U.S.


Department of Sociology & Anthropology, Utah State University, Logan, UT, USA; Department of Criminal Justice and Sociology, St. Ambrose University, Davenport, IA, USA; Department of Sociology and Criminal Justice, Iowa State University, Ames, IA, USA; Department of Agricultural Economics, Sociology, and Education, Penn State University, University Park, PA, USA; Department of Sociology, Social Work, and Criminology, Idaho State University, Pocatello, ID, USA; School of Psychology, Sociology and Rural Studies, South Dakota State University, Brookings, SD, USA; National Drought Mitigation Center, University of Nebraska-Lincoln, Lincoln, NE, USA; School of Environment and Natural Resources, Ohio State University, Columbus, OH, USA; Department of Society and Conservation, University of Montana, Missoula, MT, USA; Department of Forest Resources, University of Minnesota, Minneapolis, MN, USA; Department of Horticulture and Landscape Architecture and Natural Resources and Environmental Science, Purdue University, West Lafayette, IN, USA; Department of Sociology and Department of Geography and Geospatial Sciences, Kansas State University, Manhattan, KS, USA; Water Resources Center, University of Minnesota, MN, USA

ABSTRACT

Researchers need probability samples to collect representative survey data about the behaviors and attitudes of agricultural producers they study in relation to the natural resources that they manage, yet obtaining accurate and complete sampling frames is challenging. We extract data from a publication database to identify the most commonly used sampling frame sources in survey research of agricultural producers in the U.S., finding that government program participant lists are used most often, while private vendor samples are increasingly being purchased. Based on our research experience, we find that for many projects, private vendors can provide the most rigorous samples. Given that survey methods remain a useful and popular method for studying the behaviors and attitudes of producers on a variety of topics, such an assessment and guide is needed for researchers and practitioners.

ARTICLE HISTORY

Received 3 December 2021
Accepted 12 May 2022

KEYWORDS

Agricultural producers, natural resources, probability samples, sample sources, survey research

Introduction

Surveys are often the primary data collection method used when researchers are studying agricultural producers’ natural resource management-related attitudes and behaviors (Prokopy et al. 2019). A probability sample, with a known probability of inclusion for
each member of a population, provides the opportunity to accurately represent that population (Dillman, Smyth, and Christian 2014). However, obtaining and maintaining high-quality sampling frames can be challenging and both the completeness and accuracy of the sampling frame are important factors to consider when administering probability sample surveys.

Multiple sources of error can influence the representativeness of survey data. A holistic approach to measuring error considers how sampling procedures at each stage of the process can lead to “exclusion error” (Firebaugh 2008, 93; see also Dillman, Smyth, and Christian 2014). Exclusion error can occur at three stages of sampling a population: (1) when the sampling frame is an inaccurate or incomplete listing of population members (coverage error); (2) when the sample chosen differs substantially from all other potential samples (sampling error); and (3) when the final obtained sample data differ substantially from the initial chosen sample because of systematic differences between those who responded and those who did not (i.e., non-response error) (Groves and Lyberg 2010). Recent studies have outlined emergent issues that can contribute to exclusion error, including declining response rates and non-response bias, and have suggested research strategies to address them (e.g., Avemegah et al. 2021; Coon et al. 2020; Glas et al. 2019; Stedman et al. 2019). However, despite the importance of sampling in collecting survey data that accurately represents the population of interest, discussion about the accuracy and completeness of sampling frames is largely absent in research with agricultural producers.

The goal of this research note is to describe commonly used sources of sampling frames for survey research on agricultural producers in the United States (U.S.) and to assess potential sources of error. We analyze an existing database of relevant studies to quantify the use of various sample sources and examine some of the qualities of the sources. As researchers who frequently conduct these types of studies, we also draw upon our own research to assess these sources and provide guidance on when various sampling options are ideal.

What Sample Sources Are Being Used?

We used a database created by Floress et al. (2019) with data on articles published between 1982 and 2017 examining agricultural producers’ adoption of conservation practices or enrollment in government conservation programs in the U.S. to examine sources used to construct sampling frames for surveys. We extracted studies that used surveys of farmers as a data collection method (N=83). Of these, 77.1% used mail surveys, 6.0% used mixed or multiple modes, 3.6% used online surveys, 3.6% used drop-off/pick up, and 1.2% did not specify the mode. About two-thirds (65.1%) used probability samples (e.g., simple random, stratified, systematic) or censuses, whereas 21.7% used non-probability samples, and 13.3% did not describe their sampling method.

Table 1 shows the most commonly used sample source is “government program participants” (24.1%). In these studies, Freedom of Information Act (FOIA) requests are used to gain access to Farm Service Agency (FSA) lists of farmers who participate in government farming programs. Other common sources include working with National
Agricultural Statistics Service (NASS) to administer surveys or obtaining samples from non-governmental agricultural-related groups (e.g., land grant university Extension offices, women’s agricultural organizations). Notably, 14.5% of publications did not indicate the source of their sample. Private vendors, such as Farm Market ID (FMID; recently acquired by DTN) and CCI Marketing were used in 9.6% of studies. In 7.2% of studies, property tax parcels were used to create the sample; and “other” sources were used (e.g., county maps, field tour attendees) in 3.6% of studies. Table 1 also shows that sources, such as private vendors and property tax parcels have been used for a shorter duration than other sources and that response rates vary widely between sample sources.

**Assessment of Sampling Sources**

We draw upon our experiences with common sources to provide further information regarding coverage error, cost, time/effort to obtain, response rates, accuracy/bad addresses, and types of information provided for contacting those in the sample and conducting non-response bias checks. Our assessment is based upon studies the authors have conducted over approximately the past ten years (see Table 1 in the Supplementary Materials). We note that the data in the table does not account for differences in study design.

**Government Program Participants**

For this source, coverage error is dependent upon the degree that farmers participate in government programs. Since most federal government programs target producers of major crops (e.g., corn, soybeans), farmers who primarily raise livestock, grow specialty crops, or have diverse farm products may be underrepresented. In addition, a growing group of farmers in the U.S., the Amish, do not typically accept government financial aid for adopting conservation practices (Ulrich-Schad, Brock, and Prokopy 2017). Conversely, the expansion of federal crop insurance programs has led to more engagement of producers growing eligible crops. Federal regulatory programs (e.g., USDA organic certification or concentrated animal feeding operations) and the state department of agriculture license or certification programs (e.g., pesticide applicator certification) can provide relatively complete sampling frames for producers who are required
to obtain permits, licenses, or certifications. These lists tend to be updated frequently. These sample frames are generally free and accessible. Using best practices for survey design [e.g., the Tailored Design Method (Dillman, Smyth, and Christian 2014)], response rates for this source tend to be slightly below average [e.g., Stedman et al. (2019) found the average response rate for the most recent decade of surveys at their research center to be 43%]. Accuracy can be an issue with duplicate addresses common, particularly where lists are drawn from multiple government programs. It may be necessary to verify with survey respondents that they are currently farming given that government records can be out-of-date and often capture retirees and non-operator landowners. Typically, only names and mailing addresses are provided meaning that follow-ups using other modes and non-response bias checks are not possible.

**National Agricultural Statistics Service (NASS)**

The USDA NASS maintains a master list of farm operations for the periodic Census of Agriculture and other farm surveys. Their samples are drawn from the Census Mail List (CML), which is continually updated from numerous sources and methods (USDA NASS 2019). As a result, their database is assumed to be quite comprehensive and was considered the gold standard for surveys in the 1990s and early 2000s. Data linked to CML entries allows researchers to set selection and/or stratification criteria (e.g., farm size and type). However, there is little transparency in how some NASS lists are compiled, and experts have suggested that some segments of producers are underrepresented (e.g., women, beginning, small, or transitioning farms) (USDA NASS 2017). Conducting surveys through NASS is increasingly challenging for social science researchers. Current NASS policy requires surveys to be implemented by NASS staff (for a fee), meaning the agency conducts the mailing and data entry, which can be an expensive and lengthy process. To protect the identity of respondents, names are not shared, and NASS de-identifies continuous variables, such as farm size or farmer age by reporting them in categories, and may provide only general location data, which greatly reduces analytical possibilities. By policy, surveys conducted for external stakeholders must also be reviewed by NASS administrators and potentially by the Office of Management and Budget (OMB), adding time and potential instrument modification that may be an issue for researchers. Response rates tend to be a bit above average, but list accuracy is high. Using this source typically allows researchers to compare results to Census of Agriculture data.

**Organizational Memberships/Institutional Lists**

The quality and types of samples obtained from non-governmental organizations or institutions with producer contacts are highly variable. For instance, grower organizations may not have membership lists considered representative of the agricultural producing population in general, and, rather, over-represent specific sub-populations (e.g., larger, specialized, and more engaged producers). This source tends to be free to obtain, but the effort needed to obtain the sample can depend upon the relationship between the researcher and the organization. Organizations may be more likely to work with
researchers when the research topic is highly relevant to membership and they have a previous and positive working relationship. There can also be a cost to the organization in terms of the goodwill of its members and willingness to open mail/emails from them, so researchers should consider that they are “banking” on that goodwill. Studies using this method generally have response rates that are somewhat higher than normal, and bad addresses are typically more limited than government lists. The availability of multiple types of contact information and data about the producer/operation varies by organization. Some organizations provide only limited access to their membership [e.g., they will only send out surveys themselves, allow researchers to send via a listserv, or promote organizational materials via opt-in options (which are non-probability samples)]. This limited access has implications for the quality of the data collected, including response rates, if multiple contacts are not made. In addition, accurate response rates can only be calculated when opt-in methods are not used to contact members.

**Private Vendors**

As NASS access has declined, the use of private vendors has increased. In the last decade, private vendors (particularly FMID) have mined public and private datasets to build lists of farm operations (plat maps can also be purchased through private vendors to create sampling frames). These lists are provided for a charge to researchers and marketing organizations. Private vendors often generate their lists through publicly-available government lists, with resulting accuracy and completeness of information similar to those obtained using the USDA FSA records. List costs vary based upon study design, information provided, contact information requested/available, and whether the mapping is needed to identify those in the sample by watershed, for example. Private vendors tend to be relatively quick and the effort for researchers lower (e.g., less work deleting duplicates, etc.). Several of our authors have relied on private vendors to draw representative farmer samples, but have observed coverage errors associated with the under-coverage of smaller and newer farmers. Vendors do not always consistently update or check their lists, which leads to the inclusion of many retired and/or deceased farm operators and relatively high proportions of undeliverable addresses. While response rates can be relatively good, depending upon study design, some have also been on the lower end of normal. One of the primary benefits of using such sources is the availability of multiple types of contact information (e.g., emails and mailing addresses) and data for conducting non-response bias checks.

**Property Tax Parcels**

Property tax parcel ownership records are a useful public data source for surveys targeted at rural property owners and allow for spatial analysis given that parcel data is often available as a geospatial layer. Plat maps can also be purchased from private vendors. However, sampling frames generated from parcel data are not limited to agricultural producers and include all types of landowners. One potential way to address this issue is to select only parcels above a particular size, or to overlay parcel data with land use/land cover data, or to use local land use designations or zoning categories.
However, there are differences in how localities define agricultural land in their parcel data. Another limitation is that parcel ownership records do not include information on agricultural producers who exclusively rent land from others, though this is a small portion of farmers (Jackson-Smith and Petrzelka 2014). The cost of obtaining parcel data varies: it can be free (e.g., Minnesota; Pradhananga and Davenport 2019) as well as purchased (e.g., Eaton et al. 2019). Parcel data may require extensive time and effort to obtain and process. This source often relies on the capacity of local land use planning agencies to collate and share their data—which can vary greatly across rural spaces and states. Duplication in lists will occur where owners own multiple parcels; further, lists can include parcels owned by non-households (trusts, LLCs, corporations, etc.) or non-residential owners. Typically, only mailing address information is provided, and limited data (e.g., parcel size) is available for conducting non-response bias tests. Response rates tend to be low to average.

**Direct Comparison of Samples**

Studies where multiple sources were used in samples for the same target population allow for a more direct comparison of quality indicators. In studies in two Indiana watersheds, Big Pine and St. Marys, FSA and FMID were used as sample sources. While both sources included duplicates and unique addresses (see Supplementary Materials, Table 2), there were a few notable differences in quality between the samples which were apparent for both watersheds: (1) the FSA lists included more duplicate addresses that needed to be removed than the FMID lists (Big Pine 79.6% for FSA and 0% for FMID; St. Marys 26.1% for FSA and 0% for FMID); and (2) FSA lists excluded many addresses that lie within study area boundaries. For instance, in the Big Pine Watershed, 157 out of 180 FMID unique addresses were found in the watershed boundary but were not on the FSA list. If these farmers had received subsidy or conservation payments from FSA, they should have been identified with the FSA address source.

In a study conducted with South Dakota livestock producers who graze their cattle, FSA and FMID were also used for sampling frames (see Supplementary Materials, Table 3). The team also found duplicates within the sampling frame provided by FSA (none on the FMID list), though a much smaller percentage than in the Indiana studies (5.1%). This may be a result of the different and more specialized target population. There were also duplicates between the two sources, but each also provided a substantial number of unique addresses (57.8% of FMID and 85.8% of FSA). Additional comparisons between these sources including response rates, accuracy, and bad addresses are provided in Avemegah et al. (2021).

**Discussion**

Analysis of commonly used sample frames for survey research with farmers in the U.S. leads us to a few suggestions. First, published studies must be clear on the source of their sample frame and sampling methods used, so that assessments can be made of the quality of the data collected. In our dataset, 14.5% of studies did not indicate the source of their sample and many left out important details. Second, there is a clear need for
improving agricultural sample frames. We see opportunities for federal and state governments to collaborate with researchers and with each other to improve existing sampling frames by reducing coverage error, for example. Despite the need to keep certain identifying information private, this might include data sharing/privacy agreements where some data is blurred or combining lists of participants from various agencies. These types of collaborations would be useful for researchers, but also can ensure efforts to assess the effectiveness of government policies or interventions and provide valid and complete information.

Probability samples are likely to remain the gold standard for making valid inferences about target farm populations. However, each potential sampling frame has potential sources of error to be considered. Meanwhile, non-probability samples, including online panels, are increasingly being used and advocated for in the conservation social sciences (see Wardropper et al. 2021). While there are limitations with non-probability samples (see Baker et al. 2013), for some target populations (e.g., new women organic part-time farmers) there are few good public or private sample frames and non-probability samples (e.g., online panels) may be the best available option. With the rise of online survey modes (probability and non-probability based), emerging issues include the absence of email addresses from most publicly available farmer sample frames, duplication of emails for individuals in the lists that do exist, low online response rates, and the lack of broadband accessibility in many rural areas.

Choosing a sample source should ideally be based upon the research design while balancing the resources available for the study, the quality of available options, and the type(s) of farmers studied. There are limitations with any sample source, so being aware of them, and working to minimize error will improve future research on agricultural producers’ attitudes and behaviors. Overall, we find that private vendors are currently the most rigorous and useful sampling source given that they are more likely to produce generalizable data (e.g., more complete and accurate sampling frames), and allow for contacting potential respondents in multiple modes as recommended (see Dillman, Smyth, and Christian 2014; e.g., helps to increase response rates), and provide data that can be used for non-response bias checks (e.g., can compare respondents to non-respondents). We recommend that private vendors are used when the project resources allow and it makes sense for the particular population being studied.

Acknowledgments

First, we would like to acknowledge all of the producers who participated in the studies over many years that we draw upon here. We also appreciate the support provided by Jackie Getson and Laura Esman at Purdue University in helping compare addresses from Farm Market ID and FSA. This work was partially funded by NSF award 1739191. This paper benefited from the support provided to many authors through a National Institute of Food and Agriculture (NIFA) of the U.S. Department of Agriculture (USDA) project, NC1190: Catalysts for Water Resources Protection and Restoration: Applied Social Science Research.

ORCID

Jessica D. Ulrich-Schad http://orcid.org/0000-0002-2437-5419
Shuang Li http://orcid.org/0000-0002-7991-4694
References


USDA NASS. 2017. NASS expert panel publication of agriculture census data on farm operator demographics. USDA NASS. https://www.nass.usda.gov/Education_and_Outreach/Reports_
