Original Article

Sociological Methods & Research
1–31

DOI: 10.1177/00491241241305095
journals.sagepub.com/home/smr
Sage

Using Google Maps to Generate Organizational Sampling Frames

Brad R Fulton[1] and David P King[2]

Abstract

Organizational researchers use a variety of methods to obtain sampling frames. The utility of these methods, however, is constrained by access restrictions, limited coverage, prohibitive costs, and cumbersome formats. This article presents a new method for generating organizational sampling frames that is cost-effective, uses publicly available data, and can produce sampling frames for many geographic areas in the U.S. The Python-based program we developed systematically scans the Google Maps platform to identify organizations of interest and retrieve their contact information. We demonstrate the program's viability and utility by generating a sampling frame of religious congregations in the U.S. To assess Google Maps' coverage and representativeness of such congregations, we examined two nationally representative samples of congregations and censuses of congregations in a small, medium, and large city. We found that Google Maps contains approximately 98% of those congregations—extensive coverage that ensures a high degree of representativeness. This study provides evidence that using Google Maps to generate sampling frames can improve the process for obtaining representative samples for organizational studies by

[1]O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, IN, USA
[2]Indiana University, Lilly Family School of Philanthropy, Indianapolis, IN, USA

Corresponding Author:
Brad R. Fulton, O'Neill School of Public and Environmental Affairs, Indiana University, 510.435.5576, 1315 E. 10th St., Bloomington, IN 47405, USA.
Email: fulton@iu.edu

Data Availability Statement included at the end of the article

reducing costs, increasing efficiency, and providing greater coverage and representativeness.

Keywords

organizations, sampling frames, samples, coverage, representativeness, efficiency

Researchers often seek to examine organizational fields for which there is no easily accessible sampling frame or for which no sampling frame exists (Knoke et al., 2017).[1] This limitation makes it challenging for organizational scholars to obtain sampling frames that are comprehensive, industry-specific, and affordable. For example, business lists maintained by the Census Bureau and the Bureau of Labor Statistics are difficult to obtain, not organized by industry, and limited to organizations with employees (Grønbjerg et al., 2010; Hidiroglou & Lavallée, 2009; Jarmin, 2019). In addition, nonprofit organizations with an annual revenue less than $50,000 are exempt from filing returns with the Internal Revenue Service (2023), and some organizations (such as religious congregations) are not required to report any information (Scheitle et al., 2016).

Other public records, like telephone directories and U.S. Postal Service databases, are not limited to organizations with employees and may contain low-revenue nonprofits and congregations (Blair & Blair, 2014; Dohrmann et al., 2014). However, these records often can only be acquired through third-party vendors, and the cost and data quality of those sources can vary substantially (Harter et al., 2016). Also, many of the listings are compiled at state and local levels, each constructed independently and with its own unique formatting specifications (Murphy, 2002). As a result, merging such databases across levels to create a uniform sampling frame can be very complicated and time-consuming.

Researchers can also use proprietary directories and establishment databases from vendors, including Dun & Bradstreet and Data Axle (formerly InfoUSA), to sample niche industries and hard-to-reach fields (Kalleberg et al., 1990), such as high-tech start-ups (Samagaio et al., 2018), microbreweries (Nilsson et al., 2018), and compounding pharmacies (Guharoy et al., 2013). However, obtaining those records can be expensive, many niche industries lack reliable, up-to-date data, and their coverage levels tend to be lower in rural areas (Drucker & Feser, 2008; Fleischhacker et al., 2013; Knoke et al., 2017).

In the absence of an adequate sampling frame, researchers can use hypernetwork sampling to generate a representative sample of organizations (McPherson, 1982). Hypernetwork sampling is based on the insight that "organizations attached to a random sample of individuals constitute a random sample of organizations" (Chaves et al., 1999, p. 460). Researchers can generate a representative sample of organizations by asking a representative sample of individuals about the organizations to which they are attached (Kalleberg et al., 1990). This method can work for any type of organization affiliated with individuals, such as hospitals, museums, and unions (McPherson, 2001), and it is commonly used to generate representative samples of business organizations (Kalleberg et al., 1994), voluntary associations (Popielarz, 1999), and religious congregations (Chaves et al., 2020; Fulton & King, 2018).

The hypernetwork sampling method, however, has three critical limitations. First, it is time-consuming and expensive to implement because it requires two waves of data collection. The first wave surveys a representative sample of individuals and asks them to provide the contact information of an organization they are affiliated with, such as the religious congregation they attend. The second wave surveys the organizations mentioned by the respondents (in this example, the congregations). Second, this method produces only one sample, and the cost of producing that sample increases as the size of the sample increases. Finally, samples generated via hypernetwork sampling also require complex weighting protocols. In particular, the weights must include the weighting for the sample of individuals and account for the likelihood of a particular organization being mentioned by a respondent, which depends on the number of individuals affiliated with that organization (Fulton et al., 2022).[2] Consequently, the hypernetwork sampling method, although feasible, is neither a viable nor sustainable option for most organizational researchers.

All of the aforementioned sampling frames and sample sources possess significant limitations that undermine their coverage, representativeness, and utility. To address these challenges, we developed a new method for generating organizational sampling frames that compile publicly available data contained on the Google Maps platform. Google Maps functions as a data repository that gathers information from multiple sources and continually organizes and updates its data. More specifically, Google Maps' "Places" feature obtains and curates information about every organization on its platform.[3] The program we developed extracts this information to generate organizational sampling frames. Our program uses Google Maps' application programming interface (API) to search for particular types of organizations in specific geographic areas and then compiles a list of organizations that meet the search criteria.

To demonstrate the viability and utility of this method, we used the program to generate a sampling frame of religious congregations in the U.S.—a population for which no sampling frame exists (Brauer, 2017; Chaves, 2002). To assess the coverage and representativeness of congregations listed on Google Maps, we used two nationally representative samples of congregations compiled in 2018 and censuses of congregations in a small city (Bloomington, IN), medium city (Indianapolis, IN), and large city (Chicago, IL) compiled respectively in 2023, 2022, and 2022. We found that approximately 98% of the congregations in those samples are on Google Maps. We then assessed Google Maps' representativeness by examining characteristics of the congregations not on Google Maps (e.g., size, religious tradition, online presence, type of meeting facility, and community setting). Apart from congregations with fewer than 100 members being marginally less likely to be on Google Maps, we found no other characteristic associated with not being on Google Maps. Moreover, Google Maps' extensive coverage of congregations ensures a high degree of representativeness.

Overall, using Google Maps to generate sampling frames addresses many limitations of other existing methods. Compared to using data from public records, Google Maps has greater coverage and returns the data in one consistent format. Unlike methods that use data from proprietary records, Google Maps' data include a broader range of organization types and cost less to obtain (Simsek & Veiga, 2001). Finally, compared to the hypernetwork sampling method, our method is less expensive and more efficient, can provide multiple samples and generate larger samples without increasing costs, and does not require complex weighting protocols.

Our method can be applied to several types of organizations across many geographic areas in the U.S. and is particularly beneficial for organizations for which a sampling frame is expensive, not easily accessible, or does not exist. Our case study demonstrates Google Maps' comprehensiveness for congregations, and additional case studies are needed to assess Google Maps' comprehensiveness for the other organization types listed on its platform. In addition to producing a comprehensive sampling frame from which random samples can be drawn, researchers can use this method to estimate the total number of organizations in a particular geographic area, conduct geospatial analyses, and test organizational ecology theories (Raghavan, 2014). In short, using Google Maps to generate sampling frames can significantly improve the process for obtaining representative samples for organizational studies by reducing costs, increasing efficiency, and providing greater coverage and representativeness.

Generating Sampling Frames from Google Maps

In this section, we describe the Python-based program we developed to generate sampling frames by extracting data from the Google Maps platform on particular organizations of interest within specified geographic areas (see Appendix C for the Python code used for this study). The program uses Google Maps' Places API, a service researchers can access using a software program that returns requested data from the platform in a structured uniform format ready to analyze (see Appendix A for links to the Google Maps' Places API features mentioned in this article).

Google Maps Places. Google Maps categorizes organizations as "places" and assigns each organization a unique *place ID*. Then, it assigns each organization a *place type* (selected from a list of 97 place types) that broadly describes the organization's type. The list includes terms such as school, hospital, and supermarket (see Appendix B for the complete list). In addition, each place (i.e., organization) can be assigned up to 10 *business categories* that provide more specific details. The list of 4,105 business categories includes terms such as charter school, animal hospital, and Asian grocery store.[4] Google Maps obtains information on places from multiple sources. Specifically, Google collects data on organizations from their websites, online postings, and public records. In addition, owners can list their organization as a place on Google Maps by registering it for free at Google My Business. Although this feature is named "Google My Business," any organizational entity—business, nonprofit organization, or public institution—can be registered on Google Maps using this feature. In addition, internet users can manually add an organization or update its information on Google Maps. The platform consolidates the information on each organization and stores it as *place details*. As a result, compared to other sampling frames, Google Maps can incorporate new organizations more quickly as well as organizations that do not have a fixed location.

Regrettably, Google Maps does not provide a list of all the places (i.e., organizations) on its platform. If such a list was available, researchers could simply draw samples from that list (Braun et al., 2018). Instead, Google offers the Google Maps Places API service that researchers can use to search for places of interest on Google Maps and request data on those places.

Finding Organizations and Obtaining Organizational Data from Google Maps. The program we developed finds organizations on the Google Maps platform using the API's Place Search request, which returns a list of places based on a specific search request and corresponding search criteria. Our program specifically uses the Nearby Search option because it allows researchers to search for a particular type of place within a specified geographic area. This search request returns the place IDs of organizations that meet the search criteria. Using the place IDs, the program then uses the Place Details request to obtain information about the organizations, including their contact information.[5] Each organization's name and corresponding data are added to the sampling frame database.

To ensure that the search includes the entire specified geographic area, the program uses a raster tile mapping schema that divides the world into small square tiles, each with a fixed geographic area and scale that corresponds with the assigned zoom level (Stefanakis, 2017). For example, a map at zoom level 15 creates square tiles that each represent 1.495 square kilometers (approximately 1 square mile), and those tiles form a two-dimensional array that covers the geographic area of interest.[6]

Our program begins by setting the geographic boundaries of the search area. It starts at zoom level 15 and creates a two-dimensional array of tiles that covers the geographic area. The program then selects the first tile in the array and sets the query point at the center of that tile. The API's Nearby Search request allows researchers to specify the radius of the query area. For a circular query area to encompass the entire area of a square tile, the radius parameter is set to $l(\sqrt{2})/2$ where l is the length of the tile. Because a portion of the query area extends beyond the boundaries of a tile, the query areas of adjacent tiles will overlap, and organizations in the overlapping part of the query areas will be returned by more than one search request. Knowing this possibility, we designed the program to reject a returned organization if it contains the same place ID as an organization already stored in the database.

A constraint of Google Maps' Places API is that it will return no more than 60 organizations with each search request. Our program accommodates this constraint by dynamically adjusting the size of the tile and radius of the query area. If a search request returns fewer than 60 organizations, then the search result contains all the organizations of interest from within that query area. If the search request returns exactly 60 organizations, it means the query area contains 60 or more organizations of interest. In such cases, the program deletes the results of that search request and divides the tile into four equally-sized square tiles, each of which is .374 square kilometers (zoom level 16). The program repeats the query process with each of these four tiles—selecting a query point, setting the radius of the query area, and submitting a search request. If the search request returns fewer than 60 organizations, the program moves to the next tile. If the search request returns exactly 60 organizations again, it repeats the process of deleting the results and dividing the tile into fourths. These steps repeat for each tile until the search request returns fewer than 60 organizations, thus ensuring all the organizations of interest in that tile have been identified. The purpose of this design feature is to confirm that no searched query area contains 60 returned organizations—an outcome where a theoretical 61st organization might not have been returned by Google Maps' Places API. The program continues until every tile has been searched and each tile contains fewer than 60 organizations of interest.

After all the organizations of interest for the specified geographic area have been returned, the program then obtains the contact information for each organization using its place ID and Google Maps' Place Details request. Creating such a database serves two main purposes: 1) to identify the entire population of organizations in a geographic area and estimate the total number of organizations in that area; and 2) to provide a sampling frame from which random samples can be drawn. Figure 1 displays a flowchart of the entire process, and Figure 2 displays a visual depiction of our program implemented in the state of Indiana.

Generating a Sampling Frame of U.S. Congregations

To demonstrate our program's functionality and assess its performance, we used it to generate a sampling frame of religious congregations in the U.S. by identifying every congregation on Google Maps. Although we could assess the program using any type of organization, congregations are a fitting test case because they exist in communities throughout the country—from densely populated urban centers to sparse rural areas and from affluent enclaves to low-income neighborhoods (Flórez et al., 2019; Fulton, 2011, 2016; Wong et al., 2018). The field of congregations also has characteristics similar to other organizational fields. Congregations vary substantially along several dimensions (e.g., size, religious tradition, online presence, and type of meeting facility) (Adler et al., 2020). Like other types of organizations, congregations can be located in a variety of venues, multiple congregations can operate out of the same building, and some congregations lack a permanent physical location (King et al., 2019). In addition, the birth and death rates of congregations are similar to those of other organizational populations (Anderson et al., 2008; Pitt, 2021). These characteristics make congregations an appropriate population for testing the viability and utility of our program.

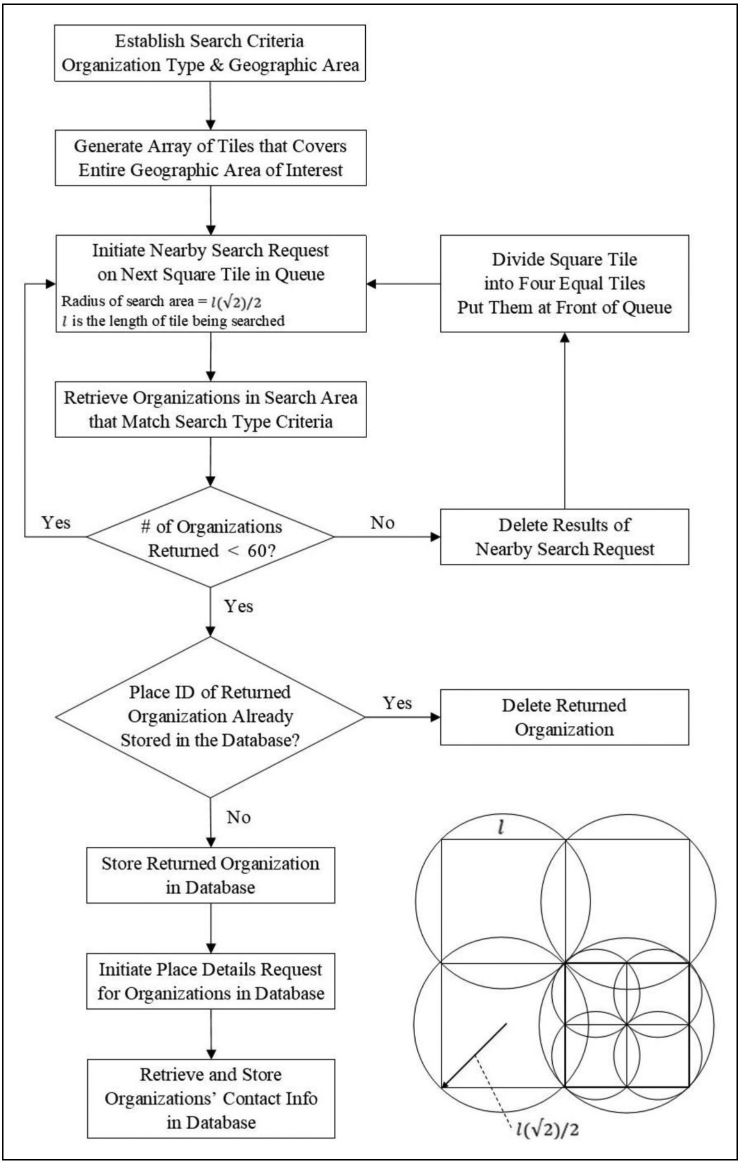

Figure 1. Algorithm for using a raster tile mapping schema to search for particular organizations of interest within specified geographic areas using the Google Maps platform and its places API.

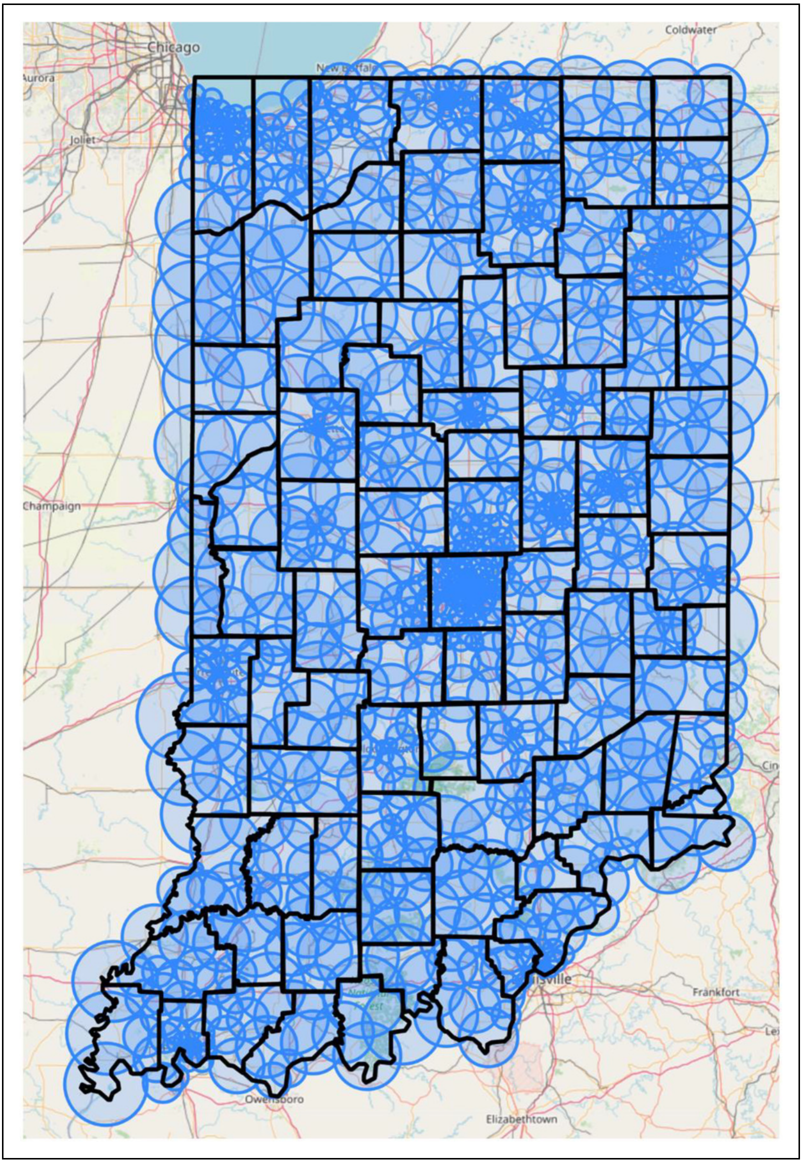

Figure 2. Map of raster tile mapping schema implemented to search for congregations within the state of Indiana using the Google Maps platform and its places API.

In the program, as part of the Place Search request, we directed the API to return every place within the U.S. identified by Google Maps as one of the following types: church, synagogue, mosque, or Hindu temple.[7] The program generated a database of congregations that includes the name and address of each congregation, and when available, its phone number and website address. In the following section, we examine the resulting sampling frame of U.S. congregations to assess the coverage and representativeness of congregations on the Google Maps platform.

Assessing Google Maps' Coverage and Representativeness

The comprehensiveness of places on Google Maps is not known. Specifically, researchers do not know the percentage of organizations on Google Maps or the characteristics of organizations *not* on Google Maps. Because most of Google Maps' internal data collection methods are proprietary and the biases associated with the crowdsourcing methods Google Maps uses to update its data are not overt (Adams & Brückner, 2015; Bright et al., 2018; Budhathoki & Haythornthwaite, 2013), the coverage and representativeness of organizational sampling frames generated from Google Maps need to be assessed. Below, we illustrate how to conduct such assessments using representative samples and censuses of a particular type of organization on Google Maps.

Sampling frames generated from Google Maps are susceptible to both undercoverage and overcoverage error. Undercovered units are in the target population, but not included in the sampling frame, while overcovered units are included in the sampling frame, but are not in the target population (Harter et al., 2016). Undercoverage exists when the sampling frame does not include an eligible organization because it is not listed on Google Maps or because Google Maps miscategorized the organization's type. For example, if Google Maps mistakenly assigns a particular mortuary the place type "cemetery" instead of "funeral home," it would not be included in a sampling frame of funeral homes. Overcoverage exists when the sampling frame includes an organization that is no longer operating or an organization that is out of scope because Google Maps miscategorized its type. For example, if Google Maps mistakenly assigns a particular pet store the place type "zoo" instead of "pet store," it would be included in a sampling frame of zoos. Overcoverage also exists when an eligible organization is listed more than once. For instance, a double listing occurs when an organization moves to a new location and Google Maps registers the new location before purging the old one. In such instances, the organization is listed twice on Google Maps, and if a sampling frame includes both the new location and the old location, it will contain a duplicate case of that organization.

To assess Google Maps' coverage of congregations throughout the U.S., we examined two nationally representative samples of congregations that were generated via the hypernetwork sampling method. These samples were created for the National Study of Congregations' Economic Practices (NSCEP) (Fulton & King, 2018) and the National Congregations Study (NCS) (Chaves et al., 2020). Both studies produced their hypernetwork sample of congregations by surveying a nationally representative sample of individuals—the NSCEP used the AmeriSpeak Panel (Dennis, 2022) and the NCS used the General Social Survey (Smith et al., 2019).[8] The samples in these surveys were used to generate a nationally representative sample of congregations. In both surveys, respondents who indicated attending religious services at least once a year were asked to provide the name and location of their congregation. Researchers then attempted to contact a key informant at each congregation (typically the head clergyperson) and ask them to complete a survey about the congregation. In total, 1,227 congregations participated in the NSCEP and 1,262 congregations participated in the NCS. Theoretically, the sampling frames from which the NSCEP and NCS samples were randomly drawn contain all of the congregations in the U.S. enabling us to use them to assess the coverage and representativeness of congregations on the Google Maps platform (Fulton et al., 2022).

Assessing the coverage involves determining the percentage of congregations in the NSCEP and NCS samples that are on Google Maps. For this analysis, we constructed the binary variable "on Google Maps" that indicates whether the congregation is on the text layer of Google Maps.[9] Table 1 displays the results of the coverage analyses. Analyses of the NSCEP and NCS samples indicate that respectively 98.14% and 98.75% of congregations in the U.S. are on Google Maps.[10]

In addition to examining Google Maps' coverage, we examined the accuracy of the geographic location, address, and contact information Google Maps provides. As part of a Place Detail request, Google Maps returns the latitude and longitude coordinates for every organization. For most organizations, Google Maps also returns a complete address: street address, city, state, and zip code. Among the records in our sampling frame of congregations, 90.88% include a complete address, and all of the records without a complete address include at least the city, state, and zip code. Among all of the records, 0.06% contain a mismatch, where the record's listed city differs from the city corresponding with the record's latitude and longitude coordinates.

Table 1. Percentages and Counts of Congregations on Google Maps by Sample Source.

Sample characteristics	National Study of Congregations' Economic Practices (NSCEP)[1]	National Congregations Study (NCS)[2]	Census of Congregations in Bloomington, Indiana[3]	Census of Congregations in Indianapolis, Indiana[4]	Census of Congregations in Chicago, Illinois[5]
Percentage of congregations on Google Maps	98.14%	98.75%	99.25%	98.51%	98.78%
Number of congregations not on Google Maps	3	5	1	12	39
Number of congregations on Google Maps, but not categorized as a congregation	0	0	0	2	4
Percentage of non-Christian congregations on Google Maps[a]	98.42%	99.48%			
Percentage of congregations located in a rural area on Google Maps[a]	99.20%	97.14%			

(continued)

Table 1. Continued

Sample characteristics	National Study of Congregations' Economic Practices (NSCEP)[1]	National Congregations Study (NCS)[2]	Census of Congregations in Bloomington, Indiana[3]	Census of Congregations in Indianapolis, Indiana[4]	Census of Congregations in Chicago, Illinois[5]
Percentage of congregations located in a low-income census tract on Google Maps[a]	NA	97.88%			
Percentage of congregations that lack a website on Google Maps[a]	98.88%	97.53%			
Percentage of congregations that rent (versus own) their building on Google Maps[a]	98.65%	98.77%			
Number of congregations in the sample/census	1,227	1,262	134	940	3,536

Data Sources: [1] Fulton & King, 2018; [2] Chaves et al., 2020; [3] Monroe County Community Organizations Database (https://mcpl.info/commorgs/all-organizations); [4] Center for Congregations (https://centerforcongregations.org/); [5] Chicago Congregations Project (https://chicagocongregations.org/).
[a]Data on this characteristic of the congregations in the Bloomington, Indianapolis, and Chicago datasets were not collected.

Analyses of the NSCEP and NCS samples indicate that among the congregations on Google Maps that include a complete address, 98.96% of the congregations' address on Google Maps either matches or is within 100 meters of the congregations' address in the NSCEP and NCS samples. Among the congregations for which Google Maps does not provide a complete address, the geographic location of the congregation's address in the NSCEP or NCS sample is within 100 meters of the congregation's latitude and longitude coordinates provided by Google Maps. The results of these analyses provide evidence indicating that if a congregation is on Google Maps, its record is likely to include an accurate address and accurate latitude and longitude coordinates.

Compared to addresses, Google Maps is less likely to return a phone number for the organization. Among the congregations in our sampling frame, 78.20% include a phone number. Analyses of the NSCEP and NCS samples indicate that among the congregations on Google Maps that include a phone number, 98.12% of the phone numbers are associated with the correct congregation. The results of this analysis provide evidence indicating that if a congregation's record on Google Maps includes a phone number, it is likely to be accurate.

While the nationally representative samples enable us to assess Google Maps' coverage of congregations nationally, they do not allow us to assess the extent of its coverage for specific geographic areas. Thus, our coverage analysis also includes examining censuses of congregations in three cities. In order to assess Google Maps' coverage across U.S. cities of varying population sizes, we selected Bloomington, IN (population 79,107), Indianapolis, IN (population 880,621), and Chicago, IL (population 2,665,039).[11] We selected these three particular cities, because a census of the congregations in these cities are available, and we had researchers on-the-ground in each of these cities, who could locate and verify the presence of congregations at specific locations. For Bloomington, we used the Monroe County Community Organizations (MCCO) database maintained by the Monroe County Public Library and the Nonprofit Central Resource Center.[12] Of the 134 congregations in Bloomington, 133 (99.25%) are on Google Maps. For Indianapolis, we used the census of congregations in Indianapolis maintained by the Center for Congregations (CfC).[13] Of the 940 congregations in Indianapolis, 926 (98.51%) are on Google Maps (2 additional congregations are on Google Maps, but they are not categorized as a congregation). For Chicago, we used a dataset of congregations from the Chicago Congregations Project (CCP) (Beyerlein et al., 2022).[14] Of the 3,536 congregations in Chicago, 3,493 (98.78%) are on Google Maps

(4 additional congregations are on Google Maps, but they are not categorized as a congregation).

Analysis of these five independent datasets indicates that most of Google Maps' undercoverage of congregations is due primarily to congregations not being on Google Maps, rather than Google Maps miscategorizing congregations as something other than a congregation. Among the 66 congregations not included in the sampling frame, six were miscategorized (e.g., as a school or cemetery). The analysis demonstrates Google Maps' extensive coverage of congregations, and given Google Maps' multiple modes for acquiring data (Rhodes et al., 2015), its undercoverage error for organizational populations likely will continue to decrease (Tonidandel et al., 2018).

Sampling frames generated from Google Maps are also susceptible to overcoverage error, which can be caused by including inactive, out-of-scope, or duplicate organizations. We assessed Google Maps' overcoverage of congregations by examining the sampling frame of congregations in Bloomington, Indianapolis, and Chicago generated by our program.[15] Although our program ensures that the sampling frame will not include organizations outside the specified geographic scope, it could include organizations that are no longer operating, miscategorized by Google Maps, or listed on Google Maps more than once.

Table 2 displays the results of our overcoverage analysis. Of the 137 records in the Bloomington sampling frame generated from Google Maps, 133 were valid and unique congregations; 2 (1.46%) were inactive (i.e., no longer operating), 1 (0.73%) was out-of-scope (i.e., not a congregation), and 1 (0.73%) was a duplicate (i.e., listed more than once on Google Maps).[16] Of the 954 records in the Indianapolis sampling frame, 926 were valid and unique congregations; 13 (1.36%) were inactive, 4 (0.42%) were out-of-scope, and 11 (1.15%) were duplicates. Of the 3,595 records in the Chicago sampling frame, 3,493 were valid and unique congregations; 51 (1.42%) were inactive, 17 (0.47%) were out-of-scope, and 34 (0.95%) were duplicates.

Based on the analysis of congregations in Bloomington, Indianapolis, and Chicago, we estimate Google Maps' undercoverage error to be approximately 2% and its overcoverage error to be approximately 3%. When researchers purchase a sample, they typically do not have access to the sampling frame—making it impossible to calculate its coverage error (Biemer et al., 2017). Even when a sampling frame is available, it is often impractical to identify the erroneously excluded and included records needed to calculate its coverage error (Groves et al., 2009). Thus, for most studies, a frame's undercoverage remains unknown and its overcoverage is not detected until

Table 2. Characteristics of the Sampling Frames of Congregations Generated from Google Maps for a Small, medium, and Large City.

Sampling frame characteristic	Bloomington, IN (pop. 79,107)		Indianapolis, IN (pop. 880,621)		Chicago, IL (pop. 2,665,039)	
Number of records in the sampling frame	137		954		3,595	
Number of valid and unique congregations	133	97.08%	926	97.06%	3,493	97.16%
Number of inactive congregations	2	1.46%	13	1.36%	51	1.42%
Number of out-of-scope records	1	0.73%	4	0.42%	17	0.47%
Number of duplicate records	1	0.73%	11	1.15%	34	0.95%

Data Source: Google Maps

a sample is drawn and analyzed. Compared to studies for which the sampling frame's coverage error is calculated, Google Maps' estimated coverage error is among the lowest (Kölln et al., 2019; Weisberg, 2009), distinct in having both very low undercoverage and overcoverage error (Biemer et al., 2017). These analyses demonstrate that Google Maps' sampling frame of congregations has extensive coverage (Groves et al., 2009), and it is reasonable to expect similar coverage for other types of organizations and in other geographic areas within the U.S. (Ibarz & Banerjee, 2017; Schmidt & Weiser, 2012).

However, a sampling frame can have low coverage error and still not be representative of the target population. To assess the representativeness of congregations on Google Maps, we examined the five datasets (NSCEP, NCS, MCCO, CfC, and CCP) to ascertain whether particular congregational characteristics are associated with not being on Google Maps. Typically, it would be appropriate to conduct logistic regression analyses using the variable "not on Google Maps" as the dependent variable and various congregational characteristics as the independent variables. However, given the rarity of a congregation not being on Google Maps and the corresponding very small percentage of such cases in the NSCEP, NCS, MCCO, CfC, and CCP samples (less than 2%), conducting a standard logistic regression or even alternative estimation methods for rare events (e.g., Firth, 1993; Hirji et al., 1987; King & Zeng, 2001; Tonidandel & LeBreton, 2010) would not produce reliable results (van Smeden et al., 2019). The very small number of events (i.e., not being on Google Maps) would amplify the small-sample bias known to affect the maximum likelihood estimation of the logistic model and substantially undermine the reliability of the estimates produced by such analyses (Gart & Zweifel, 1967; Nemes et al., 2009).

As an alternative to conducting logistic regressions, we examined characteristics of the 60 congregations not on Google Maps.[17] We anticipated that the types of congregations less likely to be on Google Maps are those that are non-Christian, located in a rural area, and/or located in a low-income area, as well as those that have minimal online presence, meet in a non-traditional building (e.g., they rent space in a strip mall, school, or community center), and/or have few members. However, examination of the congregations not on Google Maps indicates that the distributions for the congregations' religious affiliation, urban/rural location, and community income-level are consistent with the distributions for congregations in the U.S. In addition, among the congregations not on Google Maps, 75% have a website and 87% meet in a building whose primary purpose is a place of worship; proportions that roughly correspond with the proportions for the entire field of congregations.

Put differently, among the congregations in the NSCEP and NCS samples that are non-Christian, located in a rural area, located in a low-income census tract, lack a website, and/or rent (versus own) their building, greater than 97% are on Google Maps (see Table 1). Although most of the congregations not on Google Maps have fewer than 100 members, the median size of congregations in the U.S. is approximately 50 members (Adler et al., 2020). Thus, apart from having fewer than 100 members, we did not identify any specific characteristic associated with a congregation not being on Google Maps. In summary, the low risk of congregations not being on Google Maps combined with no major differences between eligible congregations on and off the frame, demonstrates the representativeness of congregation sampling frames generated from Google Maps.

Limitations

Our analysis identified three key limitations that affect the validity of the sampling frames generated from Google Maps. The first limitation concerns overcoverage error: including organizations outside the scope of the sampling frame parameters (ineligible cases). The types of ineligible cases include 1) organizations that are no longer active, but are still listed on Google Maps and 2) organizations that Google Maps has miscategorized as the target population type. For example, the congregation sampling frame we generated contains a bookstore, a school, and a funeral home that Google Maps miscategorized as congregations. Our examination of Google Maps' sampling frame of congregations in Bloomington, Indianapolis, and Chicago indicates that 66 (1.41%) cases were no longer operating and 22 (0.47%) cases were not congregations.

Although overcoverage will inflate the population count for target populations, researchers can estimate a sampling frame's overcoverage error caused by ineligible cases by drawing a sample and identifying the percentage of ineligible cases it contains. They can then revise the estimated target population size accordingly. For sampling purposes, however, it is not necessary for researchers to search the entire sampling frame for inactive or miscategorized cases (i.e., assess overcoverage). Instead, if researchers discover an ineligible case among their sample, they can remove it from the sample and replace it with a new case drawn at random from the sampling frame (see Harter et al., 2010:173–4). Alternatively, researchers could address overcoverage by oversampling (Kalton, 2021).

The second limitation concerns duplication error: when multiple cases in a sampling frame are linked to the same organization in the target population (duplicate cases).[18] A sampling frame generated from Google Maps will contain a duplicate case of an organization if the organization has moved and is listed on Google Maps at both its old and new location. Our examination of Google Maps' sampling frame of congregations in Bloomington, Indianapolis, and Chicago indicates that 46 (0.98%) cases were duplicates.

Duplicate cases are more concerning because they will inflate the population count for target populations, and they are twice as likely to be selected in random samples drawn from the sampling frame. Researchers can attempt to identify duplicate cases in the sampling frame by flagging cases that have the same name or phone number. However, not all such cases represent duplicates, as it is possible for two unique organizations to have the same name or phone number. For example, they might share a phone line, or both might list the phone number of the same regional office or national headquarters.

When it is not feasible to purge duplicates from the sampling frame prior to sample selection, researchers can address this concern during subsequent data collection. They can include an item in their study that asks respondents whether their organization has moved locations in the past year, and if so, request the address of its previous location. Then, researchers can search the sampling frame for duplicate cases of those organizations. If they discover a duplicate in the sampling frame, they can make the population count more accurate by removing the duplicate; specifically, the case with the less-current location information. They can also correct the duplicate case's doubled likelihood of being selected by applying a weight of .5 to the sampled case, thereby reducing by one-half its relative contribution to the sample estimates (see Groves et al. 2009 and Kalton 2021 for other ways to handle such cases). In addition, the duplication error of a sampling frame generated from Google Maps is likely to be very small given that Google Maps works continuously to detect and eliminate duplicate listings (Manasse, 2009).

The third limitation concerns undercoverage error: not including organizations that are in the target population (eligible cases not included in the sampling frame). The types of eligible but erroneously excluded cases include 1) organizations that are not on the text layer of Google Maps (omitted cases) and 2) organizations that Google Maps miscategorized as not being the target population type. For example, the congregation sampling frame we generated does not include a particular congregation because Google Maps miscategorized it as a school. Our analysis indicates that Google Maps' sampling frame of congregations has an undercoverage error of approximately 2%, and it is reasonable to expect a similar percentage for other organizations (Ibarz & Banerjee, 2017; Schmidt & Weiser, 2012). Furthermore, we expect the percentage of organizations not on Google Maps to decrease over time given Google Maps' multiple modes for populating its database and because it allows users to add organizations to its platform. We also expect Google Maps to increase its accuracy in categorizing organizations as its artificial intelligence and natural language processing algorithms improve (Hickman et al., 2020; Kobayashi et al., 2018), and because it allows users to correct organizations' information on Google Maps.

Undercoverage will create an undercount of the target population. However, researchers can conduct analyses similar to those discussed above to estimate the percentage of eligible organizations that were omitted or miscategorized, and then revise the estimated target population size accordingly. In addition, Google Maps' very low undercoverage error and our analysis of its representativeness suggest that undercoverage bias will not significantly impact analyses that use sampling frames generated from Google Maps.

Overall, the limitations associated with using Google Maps to generate sampling frames are minimal, especially compared to the limitations of alternative methods, and can be mitigated through supplemental analyses. In addition, sampling frames generated by this approach can be enhanced by applying it to other online data repositories that obtain their data from a variety of different sources (Bhutta, 2012). For example, Yelp is an online directory of businesses that crowdsources its data from user-supplied information and reviews. Alternatively, YellowPages.com gathers its information on organizations primarily from phone companies and business owners. Furthermore, an increasing number of organizations, such as SafeGraph and Esri, are building open data platforms and compiling places datasets. Although the data on the Google Maps platform are the most comprehensive, using multiple repositories that obtain their data from complementary sources allows for cross-checking and could further increase the coverage and representativeness of the sampling frames that are generated.

Moving forward, we can further improve the quality of the sampling frames our program produces by making our program available to other researchers. As researchers use Google Maps to generate more sampling frames for more types of organizations and conduct studies on samples drawn from those frames, estimates of overcoverage and undercoverage error will improve and be calculable for multiple types of organizations. Furthermore, coverage error is far from unique to sampling frames generated from Google Maps. Every database is susceptible to this error and its severity is inversely related to how often the databases are updated and the accuracy of their categorizations. For Google Maps, we expect its error rates to continue to decrease as it updates its databases more frequently and increases the accuracy of its categorizations (Ibarz & Banerjee, 2017; Schmidt & Weiser, 2012).

Conclusion

Most researchers obtain sampling frames from sources such as government records, public databases, and proprietary directories. However, those sources tend to have access restrictions, limited coverage, prohibitive costs, and cumbersome formats. By comparison, our method of generating sampling frames from Google Maps mitigates many of those shortcomings and offers several advantages. It can be applied to several types of organizations and is particularly beneficial for organizations where a sampling frame is expensive, not easily accessible, or does not exist. From the sampling frames it produces, researchers can draw multiple samples of varying sizes without increasing cost. In addition, because the method allows researchers to specify the geographic area and search criteria, researchers can save time by not having to identify and remove organizations outside their scope of study.

Furthermore, the information that traditional sampling frame sources contain is often outdated because they are typically updated and published no more than once per year. In contrast, because the information on Google Maps is continuously updated, obtaining sampling frames using this method is not bound by structured reporting and publication schedules. Researchers can run the program at any time and generate a sampling frame with the most up-to-date information available.

In addition to generating sampling frames from which random samples can be drawn, researchers can use this method to estimate the total number of organizations in a particular geographic area and conduct geospatial analyses. Also, by running the program at regular intervals, researchers can obtain more refined measures of field-level changes over time and use these longitudinal data to test organizational ecology theories. In short, using Google Maps to generate sampling frames can improve the process for obtaining representative samples for organizational studies by reducing costs, increasing efficiency, and providing greater coverage and representativeness.

Authors' Note

The data from Google Maps are publicly available via Google Maps' API. In Appendix C in the online supplement, we include the Python code used to extract the data from Google Maps. In Appendix D in the online supplement, we include the Stata code used to analyze the coverage of Google Maps using the NSCEP and NCS data.

The data used to assess the coverage and representativeness of Google Maps were obtained from the National Study of Congregations' Economic Practices, National Congregations Study, Monroe County Public Library and Nonprofit Central Resource Center, Center for Congregations, and Chicago Congregations Project. Restrictions apply to the availability of these data because they contain personal identifying information (i.e., the name, address, and contact information of the congregations that participated in those studies) and/or because the data was used with special permission for this article. The data can be obtained with permission from the principal investigators of the respective studies.

National Study of Congregations' Economic Practices
(https://www.nscep.org/)
National Congregations Study
(https://www.nationalcongregationsstudy.org/)
Monroe County Public Library and Nonprofit Central Resource Center
(https://mcpl.info/npc/nonprofit-central)
Center for Congregations
(https://centerforcongregations.org/)
Chicago Congregations Project
(https://chicagocongregations.org/)

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Lilly Endowment (grant number 2017-0740).

ORCID iDs

Brad R Fulton https://orcid.org/0000-0001-5082-4022
David P King https://orcid.org/0000-0001-6725-0053

Data Availability Statement

Example text of a Data statement, as provided by the author.

Supplemental Material

The supplemental material for this article is available online.

Notes

1. For example, studies of organizational fields that often rely on representative samples of organizations include health care organizations Lasater, K. B., Jarrín, O. F., Aiken, L. H., McHugh, M. D., Sloane, D. M., & Smith, H. L. (2019). "A Methodology for Studying Organizational Performance: A Multistate Survey of Front-Line Providers." *Medical care* 57 (9): 742–749. , retail food outlets Jones, K. K., Zenk, S. N., Tarlov, E., Powell, L. M., Matthews, S. A., & Horoi, I. (2017). "A Step-by-Step Approach to Improve Data Quality When Using Commercial Business Lists to Characterize Retail Food Environments." *BMC research notes* 10 (1): 1–12. nursing homes Fielding, E., Beattie, E., O'Reilly, M., & McMaster, M. (2016). "Achieving a National Sample of Nursing Homes: Balancing Probability Techniques and Practicalities." *Research in gerontological nursing* 9 (2): 58–65. , arts organizations McPherson, M. (2001). "Sampling Strategies for the Arts: A Hypernetwork Approach." *Poetics* 28 (4): 291–306. , social movement organizations Andrews, K. T., Edwards, B., Al-Turk, A., & Hunter, A. K. (2016). "Sampling Social Movement Organizations." *Mobilization: An International Quarterly* 21 (2): 231–246. , and nonprofits in general Grønbjerg, K. A. (2002). "Evaluating Nonprofit Databases." *American Behavioral Scientist* 45 (11): 1741–1777.
2. The probability that an organization appears in a sample generated by the hypernetwork sampling method is proportional to the number of people affiliated with the organization. Because organizations are nominated by individuals affiliated with them, organizations with more people are more likely to be in the sample than organizations with fewer people. Although larger organizations are overrepresented in the sample, they are overrepresented by a known degree and their overrepresentation can be accounted for with weights.
3. As of November 2024, Google Maps reports that there are over 200 million places on its platform and that over 50 million updates to the places' information are made daily (https://cloud.google.com/maps-platform/places).
4. The complete list of business categories can be viewed at https://pleper.com/index.php?do=tools&sdo=gmb_categories
5. The base price per call for a Nearby Search request is $.032 and the base price per call for a Place Details request is $.017. Google Maps offers tiered pricing with discounts applied when requests exceed 100,000.

6. The zoom levels on Google Maps range from 1 (zoomed all the way out), where the entire world is contained on one tile, to 20 (zoomed all the way in), where each tile is 1,461 square meters (equivalent to a mid-sized building).
7. In almost every case, Google Maps assigns congregations one of these four place types regardless of their religious tradition. For example, it is not uncommon for Google Maps to assign a Buddhist temple the place type "church." In addition, Google Maps tends to assign non-Christian congregations the place type "church." This type of (mis)categorization is consistent with the common practice of referencing all places of worship as churches irrespective of their religious tradition Fulton, B. R. (2020). Religious Organizations: Crosscutting the Nonprofit Sector. In W. W. Powell & P. Bromley (Eds.), *The Nonprofit Sector: A Research Handbook, 3rd Edition* (pp. 579–597).Stanford: Stanford University Press. https://doi.org/10.1515/9781503611085-035.
8. Using hypernetwork sampling to generate a representative sample of U.S. congregations first requires identifying a representative sample of U.S. adults. Research has shown that the AmeriSpeak Panel Bilgen, I., Dennis, J. M., & Ganesh, N. (2018). *Nonresponse Follow-up Impact on Amerispeak Panel Sample Composition and Representativeness*. NORC at the University of Chicago retrieved from https://amerispeak.norc.org/content/dam/amerispeak/research/pdf/Bilgen_etal_WhitePaper1_NRFU_SampleComposition.pdf. and the General Social Survey Morgan, S. L. (2020). *Response Rates and Representativeness: A Benchmark Comparison of the General Social Surveys to the American Community Surveys,* 2012-2018. retirieved from https://osf.io/7q58d/download. are representative of U.S. households. Both use the NORC National Frame, a probability-based sample frame constructed by NORC that uses the U.S. Postal Service Delivery Sequence File and provides a sample coverage for over 97% of the U.S. households Dennis, J. M. (2022). *Technical Overview of the Amerispeak Panel Norc's Probability-Based Household Panel*. NORC at the University of Chicago retrieved from https://amerispeak.norc.org/content/dam/amerispeak/research/pdf/AmeriSpeak%20Technical%20Overview%202019%2002%2018.pdf. In addition, the NCS is recognized as containing the most representative sample of congregations in the U.S. Bartkowski, J. P. (2006). "Congregations in America." *Sociology of Religion* 67 (1): 111-113., Bender, C. (2005). "Congregations in America." *Social Forces* 83 (4): 1785-1786. , Wuthnow, R. (2009). *Saving America? Faith-Based Services and the Future of Civil Society*. Princeton University Press. The NSCEP uses the same methodology as the NCS, and comparisons of the demographic composition of two samples fall within the sample estimates' margin of error for most characteristics.
9. The Google Maps platform has a visual layer and a text layer. Because the search program scans the text on Google Maps, we do not consider a congregation to be

"on Google Maps," if only its image appears on the visual layer of Google Maps (i.e., via satellite or street view) Alvarez León, L. F. (2016). "Property Regimes and the Commodification of Geographic Information: An Examination of Google Street View." *Big Data & Society* 3 (2). https://doi.org/10.1177/2053951716637885. We consider a congregation to be *on Google Maps* only if its name appears on the text layer of Google Maps.

10. The hypernetwork sampling method used to create the NSCEP and NCS samples yields a probability-proportional-to-size sample (i.e., the probability of a congregation being nominated corresponds with the number of adults that attend its services), which means that larger congregations are more likely than smaller congregations to be included in the samples. All of the analyses account for the overrepresentation of larger congregations by applying congregation-level weights that undo the probability-proportional-to-size bias in the NSCEP and NCS samples.
11. Population estimates based on 2022 U.S. Census Bureau data (https://www.census.gov/quickfacts/fact/table).
12. https://mcpl.info/commorgs/all-organizations
13. https://centerforcongregations.org/
14. The specific CCP dataset we used was the one that researchers compiled by traversing every street in each of Chicago's 77 Community Areas (virtually using Google Street View) and documenting every congregation they observed (https://chicagocongregations.org/).
15. Resource limitations prevented us from assessing Google Maps' overcoverage of congregations for the entire U.S. Additional studies are needed to perform such assessments for congregations and other types of organizations.
16. Although there are only 134 congregations in Bloomington, the original sampling frame generated from Google Maps by our program contained 137 records, 4 of which were either inactive, out-of-scope, or duplicate records. Thus, the final sampling frame contained 133 valid and unique congregations. The counts for the Indianapolis and Chicago sampling frames of congregations are calculated in the same way.
17. This analysis does not include the 6 congregations that were on Google Maps, but were not categorized as a congregation.
18. Duplication error is a special instance of overcoverage error, and because its impact on sampling frames is distinct, we cover it separately.

References

Adams, J. and H. Brückner. 2015. "Wikipedia, Sociology, and the Promise and Pitfalls of Big Data." *Big Data & Society* 2(2):1-5.

Adler, G. J. Jr., B. R. Fulton and C. Hoegeman. 2020. "Survey Data Collection Methods and Discrepancy in the Sociological Study of Religious Congregations." *Sociology of Religion* 81(4):371-412.

Alvarez León, L. F. 2016. "Property Regimes and the Commodification of Geographic Information: an Examination of Google Street View." *Big Data & Society* 3(2):1-13.

Anderson, S. L., J. H. Martinez, C. Hoegeman, G. Adler, and M. Chaves. 2008. "Dearly Departed: how Often Do Congregations Close?" *Journal for the Scientific Study of Religion* 47(2):321-8.

Andrews, K. T., B. Edwards, A. Al-Turk, and A. K. Hunter. 2016. "Sampling Social Movement Organizations." *Mobilization: An International Quarterly* 21(2): 231-46.

Bartkowski, J. P. 2006. "Congregations in America." *Sociology of Religion* 67(1):111-3.

Bender, C. 2005. "Congregations in America." *Social Forces* 83(4):1785-6.

Beyerlein, K., R. Martinez-Schuldt, B. Loskota, M. Emerson, and M. Sisk (2022). "*Finding Faith in Urban Areas: The Chicago Congregations Project* Paper presented at the annual meeting of the Society for the Scientific Study of Religion, Baltimore, MD".

Bhutta, C. B. 2012. "Not by the Book: facebook as a Sampling Frame." *Sociological Methods & Research* 41(1):57-88.

Biemer, P. P., E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, and B. T. West. 2017. *Total Survey Error in Practice. Hoboken, NJ:* John Wiley & Sons.

Bilgen, I., J. M. Dennis, and N. Ganesh. 2018. "*Nonresponse Follow-up Impact on Amerispeak Panel Sample Composition and Representativeness*." NORC at the University of Chicago retrieved from https://amerispeak.norc.org/content/dam/amerispeak/research/pdf/Bilgen_etal_WhitePaper1_NRFU_SampleComposition.pdf.

Blair, E. and J. Blair. 2014. *Applied Survey Sampling*. Sage Publications.

Brauer, S. G. 2017. "How Many Congregations Are There? Updating a Survey-Based Estimate." *Journal for the Scientific Study of Religion* 56(2):438-48.

Braun, M. T., G. Kuljanin, and R. P. DeShon. 2018. "Special Considerations for the Acquisition and Wrangling of Big Data." *Organizational Research Methods* 21(3):633-59.

Bright, J., S. De Sabbata, and S. Lee. 2018. "Geodemographic Biases in Crowdsourced Knowledge Websites: do Neighbours Fill in the Blanks?" *GeoJournal* 83(3):427-40.

Budhathoki, N. R. and C. Haythornthwaite. 2013. "Motivation for Open Collaboration: crowd and Community Models and the Case of Openstreetmap." *American Behavioral Scientist* 57(5):548-75.

Chaves, M. 2002. "Religious Organizations: data Resources and Research Opportunities." *American Behavioral Scientist* 45(10):1523-49.

Chaves, M., S. Anderson, A. Eagle, M. Hawkins, A. Holleman, and J. Roso. 2020. *National Congregations Study: Cumulative Datafile and Codebook*. Durham: Duke University, Department of Sociology: Duke University.

Chaves, M., M. E. Konieczny, K. Beyerlein, and E. A. Barman. 1999. "The National Congregations Study: background, Methods, and Selected Results." *Journal for the Scientific Study of Religion* 38(4):458-76.

Dennis, J. M. (2022). *Technical Overview of the Amerispeak Panel Norc's Probability-Based Household Panel*. NORC at the University of Chicago retrieved from https://amerispeak.norc.org/content/dam/amerispeak/research/pdf/AmeriSpeak%20Technical%20Overview%202019%2002%2018.pdf.

Dohrmann, S., T. D. Buskirk, A. Hyon, and J. Montaquila. 2014. "Address Based Sampling Frames for Beginners." *American Statistical Association Proceedings of the Survey Research Methods Section*: 1009-18.

Drucker, J. and E. Feser. 2008. *Matching Census Business Data to Dun & Bradstreet Marketplace Records*.

Fielding, E., E. Beattie, M. O'Reilly, and M. McMaster. 2016. "Achieving a National Sample of Nursing Homes: balancing Probability Techniques and Practicalities." *Research in Gerontological Nursing* 9(2):58-65.

Firth, D. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1):27-38.

Fleischhacker, S. E., K. R. Evenson, J. Sharkey, S. B. J. Pitts, and D. A. Rodriguez. 2013. "Validity of Secondary Retail Food Outlet Data: a Systematic Review." *American Journal of Preventive Medicine* 45(4):462-73.

Flórez, K. R., B. R. Fulton, and K. P. Derose. 2019. "National Trends in Food Insecurity and Congregation-Based Food Provision Between 1998 and 2012." *Ecology of Food and Nutrition* 58(3):265-80.

Fulton, B. R. 2011. "Black Churches and Hiv/Aids: factors Influencing Congregations' Responsiveness to Social Issues." *Journal for the Scientific Study of Religion* 50(3):617-30.

Fulton, B. R. 2016. "Trends in Addressing Social Needs: a Longitudinal Study of Congregation-Based Service Provision and Political Participation." *Religions* 7(5):51-67.

Fulton, B. R. (2020). "Religious Organizations: Crosscutting the Nonprofit Sector." Pp. 579-98 in *The Nonprofit Sector: A Research Handbook*, 3*rd Edition*, edited by W. W. Powell and P. Bromley. Stanford: Stanford University Press.

Fulton, B. R., I. Bilgen, V. Pineau, L. Liebert, D. P. King, and M. Dennis. 2022. "Evaluating Nonresponse Bias for a Hypernetwork Sample Generated from a Probability-Based Household Panel." *Journal of Quantitative Methods* 6(2):1-25.

Fulton, B. R. and D. P. King. 2018. *National Study of Congregations' Economic Practices: Survey Datafile*. Bloomington: Indiana University.

Gart, J. J. and J. R. Zweifel. 1967. "On the Bias of Various Estimators of the Logit and Its Variance with Application to Quantal Bioassay." *Biometrika* 181-7.

Grønbjerg, K. A. 2002. "Evaluating Nonprofit Databases." *American Behavioral Scientist* 45(11):1741-77.

Grønbjerg, K. A., H. K. Liu, and T. H. Pollak. 2010. "Incorporated but Not Irs-Registered: exploring the (Dark) Grey Fringes of the Nonprofit Universe." *Nonprofit and Voluntary Sector Quarterly* 39(5):925-45.

Groves, R. M. and F. J. Fowler JrM. P. Couper, J. M. Lepkowski, and E. SingerR. Tourangeau. (2009). *Survey Methodology* (Vol. 561). John Wiley & Sons.

Guharoy, R., J. Noviasky, Z. Haydar, M. G. Fakih, and C. Hartman. 2013. "Compounding Pharmacy Conundrum: "We Cannot Live Without Them but We Cannot Live with Them" According to the Present Paradigm." *Chest* 143(4):896-900.

Harter, R., M. P. Battaglia, T. D. Buskirk, D. A. Dillman, N. English, M. Fahimi, M. R. Frankel, T. Kennel, J. P. McMichael, and C. B. McPhee. 2016. *Address-Based Sampling* (prepared for the AAPOR Council by the Task Force on Address-based Sampling. 7(25): 2018.

Harter, R., S. Eckman, N. English, and C. O'Muircheartaigh. (Eds). 2010. *Applied Sampling for Large-Scale Multi-Stage Area Probability Designs*.

Hickman, L., S. Thapa, L. Tay, M. Cao, and P. Srinivasan. 2020. "Text Preprocessing for Text Mining in Organizational Research: review and Recommendations." *Organizational Research Methods*. (published online before print).

Hidiroglou, M. A. and P. Lavallée. 2009. "Sampling and Estimation in Business Surveys." In pp. 441-70. *Handbook of Statistics*. Elsevier.

Hirji, K. F., C. R. Mehta, and N. R. Patel. 1987. "Computing Distributions for Exact Logistic Regression." *Journal of the American Statistical Association* 82(400):1110-7.

Ibarz, J. and S. Banerjee. 2017. Updating Google Maps with Deep Learning and Street View. https://ai.googleblog.com/2017/05/updating-google-maps-with-deep-learning.html

Internal Revenue Service. (2023). *Exempt Organizations Annual Reporting Requirements - Overview - Annual Return Filing Exceptions*. retrieved from https://www.irs.gov/charities-non-profits/exempt-organizations-annual-reporting-requirements-overview-annual-return-filing-exceptions.

Jarmin, R. S. 2019. "Evolving Measurement for an Evolving Economy: thoughts on 21st Century Us Economic Statistics." *Journal of Economic Perspectives* 33(1):165-84.

Jones, K. K., S. N. Zenk, E. Tarlov, L. M. Powell, S. A. Matthews, and I. Horoi. 2017. "A Step-by-Step Approach to Improve Data Quality When Using Commercial

Business Lists to Characterize Retail Food Environments." *BMC research Notes* 10(1):1-12.

Kalleberg, A. L., D. Knoke, P. V. Marsden, and J. L. Spaeth. 1994. "The National Organizations Study: an Introduction and Overview." *American Behavioral Scientist* 37(7):860-71.

Kalleberg, A. L., P. V. Marsden, H. E. Aldrich, and J. W. Cassell. 1990. "Comparing Organizational Sampling Frames." *Administrative Science Quarterly* 35(4):658-88.

Kalton, G. 2021. *Introduction to Survey Sampling*. Sage Publications.

King, D. P., B. R. Fulton, C. W. Munn, and J. Goodwin. 2019. "The National Study of Congregations' Economic Practices Executive Report." *Lake Institute on Faith & Giving*.

King, G. and L. C. Zeng. 2001. "Explaining Rare Events in International Relations." *International Organization* 55(3):693-715.

Knoke, D., P. V. Marsden, and A. L. Kalleberg. 2017. "Survey Research Methods." *In The Blackwell Companion to Organizations*: 781-804.

Kobayashi, V. B., S. T. Mol, H. A. Berkers, G. Kismihók, and D. N. Den Hartog. 2018. "Text Classification for Organizational Researchers: a Tutorial." *Organizational Research Methods* 21(3):766-99.

Kölln, A.-K., Y. P. Ongena, and K. Aarts. 2019. "The Effects of Sampling Frame Designs on Nonresponse and Coverage Error: evidence from the Netherlands." *Journal of Survey Statistics and Methodology* 7(3):422-39.

Lasater, K. B., O. F. Jarrín, L. H. Aiken, M. D. McHugh, D. M. Sloane, and H. L. Smith. 2019. "A Methodology for Studying Organizational Performance: a Multistate Survey of Front-Line Providers." *Medical Care* 57(9):742-9.

Manasse, M. S. (2009). "Method for Duplicate Detection and Suppression." US Patent 7,603,370. Google Patents.

McPherson, J. M. 1982. "Hypernetwork Sampling: duality and Differentiation among Voluntary Organizations." *Social Networks* 3(4):225-49.

McPherson, M. 2001. "Sampling Strategies for the Arts: a Hypernetwork Approach." *Poetics* 28(4):291-306.

Morgan, S. L. (2020). *Response Rates and Representativeness: A Benchmark Comparison of the General Social Surveys to the American Community Surveys, 2012–2018*. retirieved from https://osf.io/7q58d/download.

Murphy, G. B. 2002. "The Effects of Organizational Sampling Frame Selection." *Journal of Business Venturing* 17(3):237-52.

Nemes, S., J. M. Jonasson, A. Genell, and G. Steineck. 2009. "Bias in Odds Ratios by Logistic Regression Modelling and Sample Size." *BMC Medical Research Methodology* 9(1):1-5.

Nilsson, I., N. Reid, and M. Lehnert. 2018. "Geographic Patterns of Craft Breweries at the Intraurban Scale." *The Professional Geographer* 70(1):114-25.

Pitt, R. N. 2021. *Church Planters: Inside the World of Religion Entrepreneurs*. Oxford University Press.

Popielarz, P. A. 1999. "Voluntary Association: a Multilevel Analysis of Gender Segregation in Voluntary Organizations." *Gender & Society* 13(2):234-50.

Raghavan, P. 2014. "It's Time to Scale the Science in the Social Sciences." *Big Data & Society* 1(1). doi:

Rhodes, B. B., A. E. Kim, and B. R. Loomis. 2015. "Vaping the Web: crowdsourcing and Web Scraping for Establishment Survey Frame Generation." Proceedings of the 2015 Federal Committee on Statistical Methodology Research Conference. https://fcsm.sites.usa.gov/files/2016/03/H3_Rhodes_2015FCSM.

Samagaio, A., N. F. Crespo, and R. Rodrigues. 2018. "Management Control Systems in High-Tech Start-Ups: an Empirical Investigation." *Journal of Business Research* 89: 351-60.

Scheitle, C. P., E. J. Dollhopf, and J. D. McCarthy. 2016. "Exploring Religious Congregations' Registration with the Irs." *Nonprofit and Voluntary Sector Quarterly* 45(2):397-408.

Schmidt, M. and P. Weiser. 2012 "Web Mapping Services: Development and Trends." Pp. 13-21 in *Online Maps with Apis and Web Services*, edited by M. P. Peterson. Springer.

Simsek, Z. and J. F. Veiga. 2001. "A Primer on Internet Organizational Surveys." *Organizational Research Methods* 4(3):218-35.

Smith, T. W., M. Davern, J. Freese, and S. L. Morgan. 2019. *General Social Surveys, 1972–2018. Data File and Codebook*. Chicago: NORC at the University of Chicago. retrieved from. https://gss.norc.org/documents/codebook/gss_codebook.pdf.

Stefanakis, E. 2017. "Web Mercator and Raster Tile Maps: two Cornerstones of Online Map Service Providers." *Geomatica* 71(2):100-9.

Tonidandel, S., E. B. King, and J. M. Cortina. 2018. "Big Data Methods:leveraging Modern Data Analytic Techniques to Build Organizational Science." *Organizational Research Methods* 21(3):525-47.

Tonidandel, S. and J. M. LeBreton. 2010. "Determining the Relative Importance of Predictors in Logistic Regression: an Extension of Relative Weight Analysis." *Organizational Research Methods* 13(4):767-81.

van Smeden, M., K. G. M. Moons, J. A. H. de Groot, G. S. Collins, D. G. Altman, M. J. C. Eijkemans, and J. B. Reitsma. 2019. "Sample Size for Binary Logistic Prediction Models: beyond Events Per Variable Criteria." *Statistical Methods in Medical Research* 28(8):2455-74.

Weisberg, H. F. 2009. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. University of Chicago Press.

Wong, E. C., B. R. Fulton, and K. P. Derose. 2018. “Prevalence and Predictors of Mental Health Programming among Us Religious Congregations.” *Psychiatric Services* 69(2):154-60.

Wuthnow, R. (2009). *Saving America? Faith-Based Services and the Future of Civil Society*. Princeton University Press.

Author biographies

Brad R. Fulton is an associate professor of nonprofit management and data science at Indiana University. His research draws on organizational theory and network analysis to examine the social, political, and economic impact of nonprofit organizations. Fulton leads Project 990, which is analyzing data on over one million nonprofit organizations to construct the first-of-its-kind, longitudinal funder-grantee network of the entire U.S. nonprofit sector. He also directs the National Study of Community Organizing and co-directs the Observing Civic Engagement Lab. Fulton is the co-author of *A Shared Future* (UChicago Press), and his book *Bridging Social Divides* is under contract with Oxford Press. His research has received 18 national awards from academic associations spanning six disciplines and is regularly covered by media outlets including the *New York Times*, *Washington Post*, and *NPR*. Episodes from Fulton's podcast—Statistics Made Simple—have been played over 150,000 times by people from 147 different countries.

David P. King is the Karen Lake Buttrey Director of the Lake Institute on Faith and Giving and Associate Professor of Philanthropic Studies within the Indiana University Lilly Family School of Philanthropy. His first book, *God's Internationalists: World Vision and the Age of Evangelical Humanitarianism* (UPenn Press, 2019) won the Peter Dobkin Hall Prize for the best book in the history of philanthropy. Trained as an American religious historian, his research interests broadly include exploring the practices of twentieth and twenty-first century American and global faith communities as well as more specifically investigating how the religious identity of faith-based nonprofits shapes their motivations, rhetoric, and practice.