

# Drawing random samples from voter files

By Shellie Garrett

Survey research is a common tool used to help a campaign decide its central strategic question: who will vote for the campaign's candidate/issue and why they will vote for it. Reliable survey research depends upon three elements: drafting a questionnaire that, when used in the survey, gives the campaign the information it needs to develop its strategy; selecting a random sample which accurately reflects the voters of the district in question; and, conducting the survey interviews in a manner which reduces as much as possible any bias in survey question answers. The following article will discuss general considerations of random sample selection, as well as the utility of clustering in random sample selections.

The two most common sources for selecting random samples in a jurisdiction is from the file of registered voters for that district and random digit dialing of phone numbers within the district. Each method has advantages and disadvantages.

Random digit dialing is usually less expensive initially than drawing a sample from the registered voter file. It also allows the sampler the ability to reach unlisted phone numbers not available on most voter registration lists. Random digit dialing is not dependent upon the quality of voter files available in a particular area, and for that reason has become a popular method of most prominent national polling firms.

Nevertheless, this method of selecting random samples has had several serious drawbacks. Because nothing is known about the potential interviewee except his phone number, all necessary demographic information must be asked in the interview, including whether or not the person is a registered voter. When asked such "good citizenship" questions as "Are you a registered voter?" or "Did you vote in the last election?", respondents tend to answer in the affirmative, whether or not their answer is true. In a survey conducted several years ago for a candidate in a congressional special election runoff interviewers asked voters whether or not the respondent had voted in the primary election held a few days previously. Sixty percent of the respondents said they had voted in the primary, even though actual turnout was only thirty percent. Other information, such as party registration, age, or other voter characteristics must be asked in the interview, which adds to its cost or limits other questions which may be asked.

When the areas covered by the poll are large, such as statewide or on exact county boundaries, then selection of phone numbers by exchange can usually en-

sure that people generated all live in the desired area. When complex district lines are involved, however, (for instance, congressional districts in Los Angeles County) then prefix selection is not enough to ensure respondents who live the appropriate district. To determine whether an interviewee lives in the district then requires further screening questions. These questions may not only be confusing but the answers may actually be unknown to the interviewees in many urban districts whose boundaries wind through a particular city's streets.

Drawing random samples from voter files eliminates many of the disadvantages of random digit dialing. The sampler can be certain that any person contacted for an interview is a registered voter within the proper district. Also available will be any information registered voters carry on the file, such as party registration, vote history, ethnic surname coding, and in which sub-jurisdictions the voter resides. Some national polling firms still resist sampling from voter files because they do not feel that reliable voter files exist in enough areas of the country to justify converting their methodology from random dialing to voter file samples. Yet the emergence of national vendors such as Voter Contact Services offering standardized voter files nationwide should diminish this concern. If a voter files nationwide should diminish this concern. If a voter file does not yet exist in the district, it may be worth it to the campaign to create one for use in its first survey, as well as for later campaigning.

One refinement of sampling from voter files is the selection of "clusters" of voters from households in the neighborhood of the voter selected by the random sample process. For example, a pollster desires that his sample be composed of 400 clusters, with ten voters for each cluster. The sampling program would pass the file of registered voters in the district, sorted by address within precinct, and select at random 400 voters throughout the district. In turn, the first voter from the next nine voters on the file after the selected voter would be included in a cluster of ten with the voter originally selected. Interviewers would then work through the voters in each cluster until completing a successful interview. The cluster method of sampling insures the proper geographic mix of a completed survey better than an unclustered random sample. Voters who live in a given area of a district may be more likely not to be home when the survey is conducted and would be underrepresented in a survey based upon an unclustered sample. Such distortion would be overcome using a clustered sample, either by interviewing another voter in the cluster or by attempting to

reach one voter in the cluster at another time of day.

In states with partisan registration, a further refinement of clustering called "pure party" clusters can be created. Using the above example of cluster sampling to create pure party clusters, each voter selected by random sample would have nine other voters selected from his neighborhood who were of the same political party. Purity of clusters can also be designated by gender, so that the same voter, if a Republican female, would have nine other Republican females selected from her neighborhood. The utility of pure party and gender clusters is similar to that of ordinary clusters, except that correct party and/or gender mix, as well as the geographic mix, would be preserved in the survey. It should be remembered that when selecting pure clusters a great number of voters on the file must be passed to create the pure cluster. In our example, only ten households (the original selected plus nine others) must be passed to create ordinary clusters. The creation of pure party clusters, however, may require passing twenty-five to come up with the required voters for a pure party cluster. This number could go up to fifty voters if pure party and gender clusters were desired. A good rule of thumb in determining whether and how many degrees of purity for clusters may be used is that the voter population of the district must at least be four times larger than the voter list produced (number of clusters times voters per cluster) for clusters with one degree of purity, and for clusters with two degrees of purity the voter population must be ten times as large as the list produced. An alterna-

tive is to sort the voter file into the "clustering sequence" before drawing the sample. This always allows pure clusters to be drawn but significantly raises the cost.

A pollster could request ten households with all of the voters in the household to be displayed. If pure clusters are desired, only those voters with the desired characteristics would be displayed. For example, if the cluster were only to include Republican women, only Republican women in each of those households would be included in the cluster, regardless of the other voters. Similarly, only ten voters per cluster could be required, regardless of the number of households. Degrees of purity considerations would also prevail in this selection as above. Such options within the cluster allow the pollster to tailor the sample in accordance to the needs of his client and the district.

Sampling from voter files and clustering techniques allow for greater accuracy in the conduct of survey research. Voter file samples eliminate the question of whether or not an interviewee is really a registered voter in the district. Information already on the voter file can be used by the survey researcher to increase the value of his survey. As reliable voter files are created in more and more areas of the country, the value of selecting samples from voter files will be available to more survey researchers and their clients.

*Shellie Garrett is the national Republican Voter Contact Services Representative and an independent political consultant.*



**Season's Greetings**  
from  
all of us at  
**Voter Contact  
Services**

