

Roles for Statisticians in Elections

John S. Gardenier, D.B.A.

When we think about elections, we tend to focus on what happens on Election Day and the subsequent time until the results have been certified. The reality is that there are crucial aspects of elections that span much longer time periods. It all starts with our inherent beliefs about the rights and responsibilities of citizens in a democracy. Who is a “citizen?” Do all citizens have an equal right to cast a ballot – what about convicted felons or persons who have been declared mentally incompetent? Do governments have a responsibility to try to ensure equal opportunity to vote to all qualified voters? What about the disabled, the illiterate, the homeless, the frail elderly without caregivers who could get them to the polls, absentees, the desperately poor, people who are social outcasts despite being legal citizens, those who only speak some exotic language and so forth? Even if we postulate that the government should try to give maximum opportunity to vote to all its citizens, is there no limit to the expense and effort the rest of us have to bear to support this noble intent? The statisticians who may be able to tell us how many people of each type live within a voting jurisdiction are called demographers. Those who are trained to measure the costs and benefits of governmental (and other) responsibilities often are called econometricians. In general, there are a large number of statistical specialties that students may aspire to. Many of those specialties have a potential role in assessing whether elections are handled fairly and responsibly.

The initial element in election management is the legislative component. Laws have to be written to address some of the issues above as well as to create the appropriate offices and mechanisms to manage elections, to authorize the use of facilities as polling places, to provide funding to buy the needed equipment, supplies and training, and to determine what post-election measures are appropriate for the initially apparent results to be certified as official. Do laws have to be fair? We may like to think so, but there are clear historical examples of unfairness in voting laws, such as the disenfranchisement of women and blacks into the twentieth century. Do all people tend to agree on what is fair? Clearly not, as illustrated by current court cases regarding what sort of proof of identity states may require as a prerequisite to being able to vote. How does one assess the fairness of voting laws? One of several methods may involve the use of statistical evidence. For a given standard of identity proof, how many people can be expected to claim a false identity, thereby casting a fraudulent vote? Is it enough to sway the electoral outcome? How many people may be unfairly prohibited from voting simply because it was too expensive or otherwise onerous for them to meet the standard of proof of identity? Would that level of disenfranchisement affect the election outcome?

Given the disagreements about what is fair, it is not uncommon for different groups within the voting public to feel that a legislative body may have deliberately set up laws that unfairly discriminate against people like them. Even if we assume the opposite, however, that the legislature in good conscience did its absolute best to treat all citizens equitably, how can we tell objectively whether or not that goal was achieved?

If the demographics change over time, can we be assured today that the laws written decades ago are fair to the current distribution of affected residents? There is no way to answer questions such as these without the use of sound statistical methods.

The American Statistical Association's *Ethical Guidelines for Statistical Practice* include in the Preamble a section on "Statistics and Society." The flavor of that section shows in its opening paragraph:

"The professional performance of statistical analyses is essential to many aspects of society. The use of statistics in medical diagnoses and biomedical research may affect whether individuals live or die, whether their health is protected or jeopardized, and whether medical science advances or gets sidetracked. Life, death, and health, as well as efficiency, may be at stake in statistical analyses of occupational environmental or transportation safety. Early detection and control of new or recurring infectious diseases depend on sound epidemiological statistics. Mental and social health may be at stake in psychological and sociological applications of statistical analysis."

One widespread application of statistics is quality control of routine processes. Food safety monitoring, industrial quality control, sales research, weather forecasting, and consumer protection generally tend to depend on meticulous record keeping and sound statistical analysis. Nonetheless, there is a vital area of our national well-being and sense of community as Americans where statistical quality control is seldom found and is never comprehensive. That is the quality monitoring of election processes. A terribly unfortunate result of the lack of quality control is selective deep distrust of election: laws, officials, organization, technology, registration procedures, voting procedures, polling place distribution, allocation of equipment to polling places, training of election officials and poll workers, security and safeguarding of records, tallying and reporting of vote counts and exit polls, and quality and integrity of communication with the voting public generally.

Fortunately, extreme distrust seems to be rare across states and across counties within the states. Still, deep cynicism does exist and may be growing. Comedians love it, but it is socially unhealthy. In the DVD, *The Best of the Colbert Report*, Steven Colbert was interviewing a Florida Congressman up for re-election. One remark was to the effect that since the election is in Florida, "you could receive all the recorded votes even if no one voted for you." Not all of such skepticism is unfounded. Scientific studies of many computerized voting systems have shown the vote tallies to be grossly unreliable. This has resulted in the top election officials of several states decertifying some or all of the voting technology they have purchased even fairly recently. There is widespread, although not universal, agreement among computer scientists that no computerized voting system can be trusted to report accurate vote tallies in the absence of an auditable paper trail. As a result, some states such as Maryland that previously relied entirely on computerized vote tallies have recently changed their laws

to require auditable paper trails. Typically, the fear cited is that “hackers” may break into the voting machines electronically and manipulate the vote counts. Apart from that, in highly partisan election situations, some opposition voters suspect that the party controlling the election officials and equipment may rig the software unfairly themselves. This suspicion is heightened when executives of the voting machine manufacturers function as campaign officials for the party controlling the election.

The term “auditable paper trail” in this context can be very ambiguous. Simplistically, it might mean that you just count paper records of the ballots to determine whether those counts match what the computers report. That simplicity is fraught with problems, however. Unless there is a means to allow each voter to certify the accuracy of her or his ballot, the equipment could be made to spew out paper “records” that match an inaccurate software count. The numbers of votes may run into the hundreds of thousands or even many millions; complete hand counts could be prohibitively expensive. It is far better to take a statistical sample, but how should we sample the ballots? How do we know that the paper records are complete and accurate? How can we ensure that the hand counting process is at least as reliable or more reliable than the computer tallies? Such questions cannot be answered rigorously without professionally designed statistical quality control.

Other professions are also essential to the task; engineers, business equipment design and maintenance personnel, and professional election managers are just a few. Elections also rely on large numbers of volunteers; those people have to complete appropriate training as well doing the work at polling places. Some of these people have been known to put in twelve hours or considerably more on Election Day. The resulting fatigue can introduce additional error tendencies. For our current purposes, let us assume that all the right people have been selected and trained, the operations are well managed, and no errors are introduced by fatigue or other impairment. Can we be confident in the vote tallies under these circumstances? Maybe, but maybe not.

Some states may allow anyone to vote by mailed paper ballot if they choose to do so rather than go to a polling place. (In Switzerland, all citizens have that option now and a third option of online voting is currently being introduced.) Most states allow people to vote by mailed paper ballot only if they will be absent from the jurisdiction on Election Day. All absentee ballots tend to be in paper form. The paper ballots may not be verified, categorized, stored and safeguarded in the same manner as the computerized balloting at the polling places. This makes it difficult to sample from the mailed paper ballots in a manner identical to that used for the regular votes. Typically, at polling places, representatives of the major parties can observe the voting process and verify its fairness to the extent that is visible. That may not be true of the handling and tallying of paper ballots outside of polling places. Such issues tend not to be important unless the winning margins are very small. In heavily contested, highly partisan elections, the margins may be so tiny that mishandling of even a small percentage of the total votes cast could sway the election result. In such cases, voters

on the losing side are strongly tempted to suspect chicanery by the winners.

Issues with the vote tallies receive a lot of attention. Many states have made or are making detailed studies of these issues. At the federal level, the Help America Vote Act (HAVA) of 2002 created an Election Assistance Commission to help develop sound election system standards. The National Institute of Standards and Technology (NIST) is tasked and funded to conduct research toward that end. They have enlisted the help of many experts to serve on or to offer advice to a Technical Guidelines Development Committee. As a result, issues with voting machines, ballot safeguarding, and tallying processes are getting a lot of attention. It is not at all clear that sound statistical quality control is receiving the needed attention at the national level. It is starting to be addressed by some state legislatures and their top election officials. Volunteer members of the American Statistical Association (ASA) and of several private organizations specifically oriented to assuring election fairness and accuracy are working hard to make relevant professional quality advice available to any authorities willing to use it. A number of state election officials are participating in the same effort.

Still, we are only dealing with part of the problem; a much broader effort is needed. Vote tallies address those citizens who are aware of their rights and who are both willing and able to register and vote in person or by mail. In many elections, that covers only a minority of the citizens. It can be confusing and complicated to register; it may not be easy to find out how to vote by mail or where to find one's assigned polling place. The best way to determine whether or not all types of citizens have ready access to voting is to compare census data regarding the distributions of residents within a voting jurisdiction with the registration rolls. If, for example, such a comparison should show that far smaller percentages of rural citizens than urban or suburban citizens are registered, it may indicate insufficient outreach to those people. Similarly, with regard to people of minority ethnicities or religions, or people in low income families; there may be *de facto* or even *de jure*, discrepant access to voting. Whether there are significant discrepancies among such voter groups can best be determined statistically. For that to happen, public records must be maintained and made accessible regarding the registration rolls.

Mentioned above were especially problematic groups of citizens, including some of the disabled, the illiterate, the homeless, the frail elderly without caregivers who could get them to the polls, absentees, the desperately poor, people who are social outcasts despite being legal citizens, those who only speak some exotic language and so forth. Most of us know next to nothing about the lives and problems of such people. In many cases, of course, they have more immediate problems with food, shelter, clothing, health care and personal safety, thus leaving voting issues to a minor status. Still, if they are legally entitled to vote, should we not at least consider how best to allow them to do so? The best way to study this problem is to conduct statistical sample surveys of such people in which we ask them what problems they face as citizens trying to vote in

elections. How hard is it for them to register, to find out about coming elections, to find out how they are supposed to vote, to get access to mail ballots or polling places, to fill out the ballots and to cast their votes?

Speaking of statistical surveys, one of the serious professional issues in survey design is determining how to ask questions so they will be understood properly and will elicit sensible answers. When the survey is in pencil and paper form, rather than through personal or telephone interview, the issues in design of that form are very similar to the issues in design of ballots. Professional survey statisticians know that any such design must be tested either in a laboratory or in the field or both. Because of this expertise, they are exceptionally well equipped to assist election officials to avoid ballot design-induced voting errors, such as those of the infamous “butterfly” ballot in Florida in 2000.

As noted in the ASA President’s Invited Column in the June 2007 AMSTAT NEWS, “Official standardized data are also needed on how well each voting site works. Qualified poll observers should fill out forms showing how well each part of the election plan worked at each polling place. Plans for future elections should rely on statistical analyses of such data to clarify how well the election system served the voters and help identify needed improvements.” That is an essential element of process quality control.

Also from that same column, “Exit polls and post-election surveys are inherently statistical. As always, samples must be random and scientifically designed. Proper analysis and interpretation of such data can help explain results that may appear to be wrong on the surface. Maintaining checks on every election is the best way to make sure any errors or problems with an election process are found and fixed.”

To end this essay with the same thought that ended that column, “The public demands responsible quality control for consumer products. Our democracy deserves no less.”