Florida's District 13 Election in 2006:
Can Statistics Tell Us Who Won?

By Arlene Ash and John Lamperti

Elections seem simple. People go to the polls. They make choices about one or more contests or issues. The votes are counted. What can go wrong with that?

Unfortunately, many things can go wrong. In the United States voters are often confronted with bewildering numbers of issues to decide. Ballot choices and designs vary from election to election and from district to district—or even within a district. People may have trouble casting the votes they intend. Both machine and human issues affect how votes are recorded and counted. Especially in a close race, the official results may not reflect the actual choices of the voting public.

Florida's 13th Congressional District 2006 Election

The 2006 contest for the U.S. House of Representatives in Florida's District 13 is such a race. The Republican candidate Vern Buchanan was declared the winner by just 369 votes, triggering a "mandatory recount." Unsurprisingly, re-querying the same "touch-screen" machines that had delivered the vote the first time changed nothing. The Democrat, Christine Jennings, refused to concede and continues to challenge the result. The problem is not that the race was close. It is that in Sarasota County, an area of relative Democratic strength, some 18,000 people, almost 15% of those who went to the polls and cast ballots, had no choice recorded for their representative to Congress. A cast ballot with no recorded choice in a race is called an "undervote." The rest of the district contributed about half the total vote, but less than 3,000 undervotes. Jennings believes that the excess missing votes in Sarasota would have tipped the race to her. Can statistical analysis help evaluate that claim?

Congressional District 13 (CD-13) is geographically diverse (see Figure 1) including all of Sarasota, all or most of DeSoto, Hardee, and Manatee Counties, and a small part of Charlotte County. About half the district's population (a count of about 370,000 people) is in Sarasota . Manatee has a population of 310,000. DeSoto and Hardee together contribute 65,000 residents. Some issues and candidates are county-specific, so voters in different parts of the district faced different ballots. George Bush received 56% of the entire CD-13 vote in 2004. However, Sarasota County leans Democratic, and of course the broader political climate also shifted between 2004 and 2006.

In 2006, all voters in CD-13 participated in the House race plus 5 statewide elections—for U.S. Senate and 4 state offices: gubernatorial (for a combined governor/lieutenant governor slate), attorney general, chief financial officer, and commissioner of agriculture. They were also presented with numerous county-specific races and issues. Indeed, each District 13 voter faced a ballot presenting anywhere from 28 to 40 choices. Voting occurred in one of three ways: by

absentee ballot, early in-person voting, or the traditional election-day visit to the polls. Touch-screen voting machines (also known as Direct Recording Electronic, or DRE; http://www.fec.gov/pages/dre.htm) were used at all polling stations in Sarasota County for both early and same-day voting. Except for the absentee ballots, the machine totals are the only record of the vote.

What accounts for the 18,000 missing votes for U.S. representative? What would their effect have been?

Undervotes

Undervotes may be intentional—for example, in little-contested local races, where voters have no knowledge or preference. They may also be unintentional—the voters accidentally do not register a vote in a particular race. Finally, they may be entirely "false"—the voters choose, but no choice registers, as with the famous hanging chads of 2000. In well-publicized statewide or national races, undervoting is normally in the 1% to 3% range, with unknown contributions of intentional, unintentional, and "false." The campaign for this important, open U.S. House seat had been intense and, by many accounts, dirty. Yet in Sarasota County about 1 out of every 7 ballots cast by touch-screen recorded no vote in this race. Why?

State officials at first echoed the explanation offered by aides of the declared winner: voters must have abstained due to disgust at the nasty campaign. However, none of the other counties had unusual undervotes in the same race; Manatee County, for example, reported normal undervoting of only two percent. Why would "voter disgust" stop at the county line? Moreover, the undervote *on absentee ballots* was low everywhere; only ballots in Sarasota County that had been voted *on touch-screens* displayed abnormally high undervoting.

In Sarasota County the highest undervote rate occurred in early voting. Thus the huge undervote in Sarasota was: specific to that county, applied to in-person voting but not to absentee ballots, and moderated, somewhat, between early and election-day voting. As we will see below, there is at least one obvious explanation for this pattern—a ballot design (in Sarasota County only) that made it more difficult to vote for U.S. Representative there than elsewhere in CD-13. Indeed, the Sarasota *Herald-Tribune* cited contacts from "more than 120 Sarasota County voters" reporting problems, mainly with ballot screens that "hid the race or made it hard to verify if they had cast their votes." This alone would hurt Christine Jennings, since Sarasota County voters were more favorable to her than were voters in the other counties.

The ballot design in Sarasota County certainly caused problems. Computer Screen 1 was devoted entirely to Florida's U.S. senatorial race, with 7 lines of choices presented, immediately beneath a bright blue banner labeled "Congressional." The undervote rate in this race was normal (that is, low). But Screen 2 presented the House race at the top with only two voting lines and no special banner. The bulk of the page, following a second bright blue banner ("State") listed seven choices on 13 lines for the gubernatorial election. See Figure 2.

Laurin Frisina and three collaborators believe that the CD-13 undervote in Sarasota County was due to the ballot screen layout. They point out that abnormally high undervote rates (ranging from 17 to 22%) were also found in the attorney general's race, and just in one part of CD 13—

Charlotte County. On that ballot (only) it was the AG race with only two candidates that shared a screen with 13 lines of choices for the gubernatorial election.

Other factors likely contributed as well. For example, there were abnormally slow machine response times that could lead people to "unvote" while trying to ensure that their vote registered. This was flagged as a problem by the voting-machine supplier the previous August but not fixed prior to early voting. Furthermore, there are strong patterns in the undervote within Sarasota County (see below), despite the fact that all Sarasota voters faced the same ballot. Walter Mebane and David Dill, after extensive study, believe that the cause of "the excessive CD-13 undervote rate in Sarasota County is not yet well-understood, and will not be understood without further investigation." Regardless of the cause(s), warnings to precinct captains about problems, which the Supervisor of Elections issued after the early voting, are consistent with observed lower rates of election-day undervotes. Even so, it continued to be far more difficult to vote in the CD-13 race in Sarasota County than elsewhere. This much is beyond dispute.

Consequences of the undervote

But did it *matter* that 18,000 Sarasota voters had no recorded votes in the House race? Assuming a normal rate of intended undervotes, the choices of some 15,000 voters were not counted. What inferences can be made about how those votes would have divided between the candidates if they had been recorded? Would they have changed the outcome? There are several ways to tackle this question, and we'll describe perhaps the simplest one. Imagine a group of N voters, with R of them intending to vote for the Republican candidate and D for the Democrat, so that. R+D = N. Suppose a random group of N-*n* votes are "lost" creating an undervote. Thus, *n* votes are actually counted, *r* Republican votes and *d* Democratic ones ($d = n - r$). Let's think of these *n* recorded votes as a random sample taken without replacement from the population of N would-be voters. Of course, we often make inferences from samples to the whole population. Usually, the sample size, *n*, is a small fraction of the population size, N. Here we have a very large sample; *n* is over 85% as large as N! Never mind, the calculations are the same.

The *r* Republican votes in the sample are viewed as the result of *n* "trials," draws without replacement from a population of size N, where the "success" probability is p = R/N, here $\approx 1/2$. Thus, the expected value of *r* and its variance are computed in the familiar way:

$$\mathrm{E}(r) = n \cdot \mathrm{p};$$

$$\mathrm{Var(r)} = np(1-p)\frac{N-n}{N-1} \approx \frac{n(N-n)}{4N}.$$

The multiplier (N–n)/(N–1) is the familiar "finite population correction factor" for sampling without replacement, found in any survey sampling text. It can often be neglected—but not here! Both N–*n* and *n* are large, so the distribution of *r* is nearly normal. In this case, all we need do to estimate the Republican advantage (possibly negative) in the whole population is "inflate" *r*–*d*, the Republican advantage in the counted votes, by N/*n*, the fraction by which the whole population exceeds the counted vote. Thus a statistically unbiased estimator of R–D is

$$\text{Estimated(R–D)} = \frac{N}{n}(r-d) = \frac{N}{n}(2r-n).$$

The associated standard error is SE = $\sqrt{N(N-n)/n}$. This translates easily into a 95% confidence interval for R-D:

$$\frac{N}{n}(r-d) - 2\sqrt{N\frac{(N-n)}{n}} \le \text{R} - \text{D} \le \frac{N}{n}(r-d) + 2\sqrt{N\frac{(N-n)}{n}}.$$

How does this result apply to the District 13 election? First, let's imagine that, say, 20,000 non-voters had been randomly chosen from the whole voting population of the district, which was roughly N = 240,000 in 2006. The counted ballots gave Republican Buchanan an edge of 369 votes; that's the value of (*r–d*). By the above formula, the 95% confidence interval for R–D ranges from a low of just over 100 to a high of nearly 700. Since the interval contains only positive numbers, we conclude with (greater than) 95% confidence that there would not be enough Democratic votes among the missing 20,000 to shift the outcome. Thus, despite the tiny winning margin (less than 1/6 of 1%) and the huge number of missing votes—*if the missing votes were distributed just like the whole population*—random error due to their loss would not threaten the outcome!

Of course, the missing votes were *not* chosen randomly from the whole district. For starters, the vast majority came from Sarasota County where Jennings had an advantage. Suppose that there was a "normal" intentional undervote of 2.5% among the 120,000 voters in that county, so that only 15,000 (of the 18,000) undervotes were unintentional. Assume the 15,000 uncounted votes were chosen randomly from the county. Would that matter? Indeed it would! In Sarasota, the recorded votes gave Jennings an edge of 6,833, so *r–d* = –6,833. If R–D now stands for the true Republican advantage among 117,000 would-be voters in Sarasota County, the point estimate for R-D is –7,838, with a 95%-confidence interval ranging from about –8,100 to –7,575. Elsewhere in the district Buchanan had an advantage of 7,202 votes. If we treat the votes in the other parts of the district as error-free, the estimate indicates a win for Jennings by 636 votes, with a 95% confidence interval for R–D ranging from –898 to –373.

Again the interval does not cross zero, and so with more than 95% confidence we conclude that Jennings should have won. In fact, had we used ±4 SE instead of ±2 SE, the confidence interval still would not include zero; this raises the confidence level to 99.9%. Moreover, in the context of a one-sided question—did Buchanan really get more votes than Jennings? —1-sided confidence bounds could be used, raising the level of certainty even higher.

Refining the estimate

In making this estimate, we assumed that 15,000 unintentional undervoters in Sarasota County differ from those who did vote only in the fact that their votes were not recorded. Can this assumption be tested? Table 1 and Figure 3 are based on "ballot image" data from Walter Mebane, that show the *sets of choices* for the 104,631 Sarasota County ballots with touch screen votes recorded in all 5 statewide contests. The data are arranged by early versus election-day

voting and by the number of Democrats chosen in the 5 statewide contests. We'll soon see how useful such data can be.

First, in both early and election-day balloting, there is a steep gradient associating partisan voting in the other races and the preference of voters—those whose choices were captured—in the House race. For example, in early voting among otherwise "straight-ticket" Democrats only 1.4% of votes for the House race went to Buchanan, as opposed to 94.9% of recorded votes among early-voting Republican stalwarts.

Second, it was far easier to "lose" Democratic votes than Republican ones in this race. For example, the straight-ticket Democrats had 18% uncounted votes in early voting as opposed to "only" 10% for their early-voting Republican counterparts. Understanding what caused these differences is crucial for the legal challenge to this election, and for avoiding future voting debacles. For our purposes we merely note that—in contrast to our previous assumption—not all Sarasota voters were equally at risk for unintentional undervotes. We'll return in a minute to the more refined calculation of the expected effect of the lost votes that these data allow.

A third important fact that emerges (Figure 3) is that the undervote declined substantially within all categories of voters between early voting and election–day voting. Apparently, many voters were helped by actions taken to mitigate the problems seen in early voting. A study exploring associations between corrective actions taken at individual precincts and undervote rates could be very informative. We do not have such data.

What we *do* have in the ballot image data leads to a sharper estimate of the likely disposition of most of the missing Congressional votes. First, it is hard to imagine that many of the 12,000 voters who expressed a choice in all 5 statewide races (including Commissioner of Agriculture and Chief Financial Officer) but had no vote recorded in the House race had intentionally undervoted. Let's suppose that they all intended to vote. How would they have voted? A good guess is that the people with missing House votes in each of the 12 strata in Table 1 would have voted in the same proportions as those in the same stratum whose votes were recorded. That is, we perform the same calculations as above, this time within each stratum of Table 1. Then we sum the estimates of the "full" vote across the strata, leading to a new estimate of R–D, representing the Republican advantage after imputing values for the undervote among these 12,000 people. This calculation suggests that Jennings' advantage among these lost votes alone was almost certainly greater than 3,000! It swamps Buchanan's original 369 vote winning margin.

For whatever reasons, it was harder to cast a successful vote for Jennings than for Buchanan in Sarasota County. The higher observed undervote among presumed Democrats means that our previous confidence interval calculation was conservative; the conclusion that Christine Jennings was the real winner in CD 13 becomes even surer.

The study by Frisina *et al* mentioned earlier uses two different methods to analyze the CD 13 undervote. Both infer undervoters' choices from their votes for other candidates. One uses precinct-level data from Sarasota County. The other involves matching Sarasota voters with

counterparts in Charlotte County. Both show that Jennings was almost certainly the preferred choice among the majority of CD-13 voters.

These different estimates may seem confusing. However, the key point is that all plausible models of what the lost votes would have been point to the same conclusion. Furthermore, the more carefully we examine the data, the more support we see for that conclusion. While poor ballot design may or may not fully account for the Sarasota undervote, it is clear that those missing votes switched the outcome of the Congressional race from Jennings to Buchanan.

What happens now?

Finally, two questions. How should Florida—and other states—fix their flawed electoral processes? Requiring a paper record is useful but not enough, since recounting such a record in District 13 might have simply confirmed that 18,000 Sarasota County voters recorded no choice for their U.S. Representative. The paper record therefore must, at least, be confirmed by each voter. We favor *paper ballots,* plus optical scanners to read them—the method familiar to us all from grading tests and used now for elections in many states. It is relatively inexpensive and foolproof. It does not require new, possibly fragile, technology or big capital investments. It provides an independent check on what is going on inside the machines that tally the votes. Optical scan ballots are also easier to read and less prone to the design problems that disfigured the CD-13 House race. Indeed, optical scanning was used in 2006 in Sarasota County for the absentee ballots, and it worked well.

Just how easy it *should* be to register and vote in the United States is controversial, as is the extent to which voters deserve redress for problems that they could conceivably have overcome, had they tried hard enough. But electoral outcomes due to extra burdens being imposed on certain voters and not others in the same race violate the basic fairness that Americans have a right to expect in their elections.

The second question, of course, is what to do about that dubious 2006 election. The statistical evidence shows, beyond any reasonable doubt, that more voters wanted Jennings than Buchanan. However, there is—as yet—no precedent for a court overturning an electoral "count" based on a statistical analysis. We have recommended doing this election over—and doing it right! For the future, statisticians and voting experts should work together to develop guidelines for the appropriate use of statistical evidence to confirm, or overturn, elections.

Further Reading

Adams, Greg (2001). Voting irregularities in Palm Beach, Florida. *Chance,* 14, 22-24.

Frisina L, Herron M, Honaker J, and Lewis J. Ballot Formats, Touchscreens, and Undervotes: A Study of the 2006 Midterm Election in Florida. (2008.) *Election Law Journal*. To appear. Draft at www.dartmouth.edu/~herron/cd13.pdf.

Marker, D., Gardenier, J., and Ash, A. (2007). Statistics Can Help Ensure Accurate Elections. *Amstat News*, President's Invited Column. On-line at http://www.amstat.org/publications/AMSN/index.cfm?fuseaction=pres062007

McCarthy J, Stanislevic H, Lindeman M, Ash AS, Addona V, Batcher M. (2008). Percentage-Based Versus Power-Based Vote Tabulation Statistical Audits. *The American Statistician.* Special Section: Statistics for Democratic Processes. 62(1): 1-6. (A more detailed version of this paper is available on the web as "Percentage-based versus SAFE Vote Tabulation Auditing: A Graphic Comparison." (2007). http://www.verifiedvotingfoundation.org/auditcomparison).

Mebane W Jr, Dill DL. Factors Associated with the Excessive CD-13 Undervote in the 2006 General Election in Sarasota County, Florida. Draft of January 18, 2007. http://www-personal.umich.edu/~wmebane.

Meyer, Mary C. (2002). Uncounted votes: Does voting equipment matter? *Chance,* 15, 33-38.

Wallace, J. (2006). Political operatives gather for recount. *Sarasota Herald Tribune,* http://www.heraldtribune.com/apps/pbcs.dll/article?AID=/20061111/NEWS/611110643 November 11, 2006. See links to

- "Special Section: District 13 election": http://www.heraldtribune.com/apps/pbcs.dll/section?CATEGORY=NEWS0521&template=ovr2
- "Precinct-by-precinct map of the undervotes": http://www.heraldtribune.com/apps/pbcs.dll/section?CATEGORY=Election03.

Wolter K, Jergovic D, Moore W, Murphy J, O'Muirheartaigh C. 2003. Reliability of the Uncertified Ballots in the 2000 Presidential Election in Florida. *The American Statistician*. 57(1):1-14.

Figure 1. Map of Congressional District -13

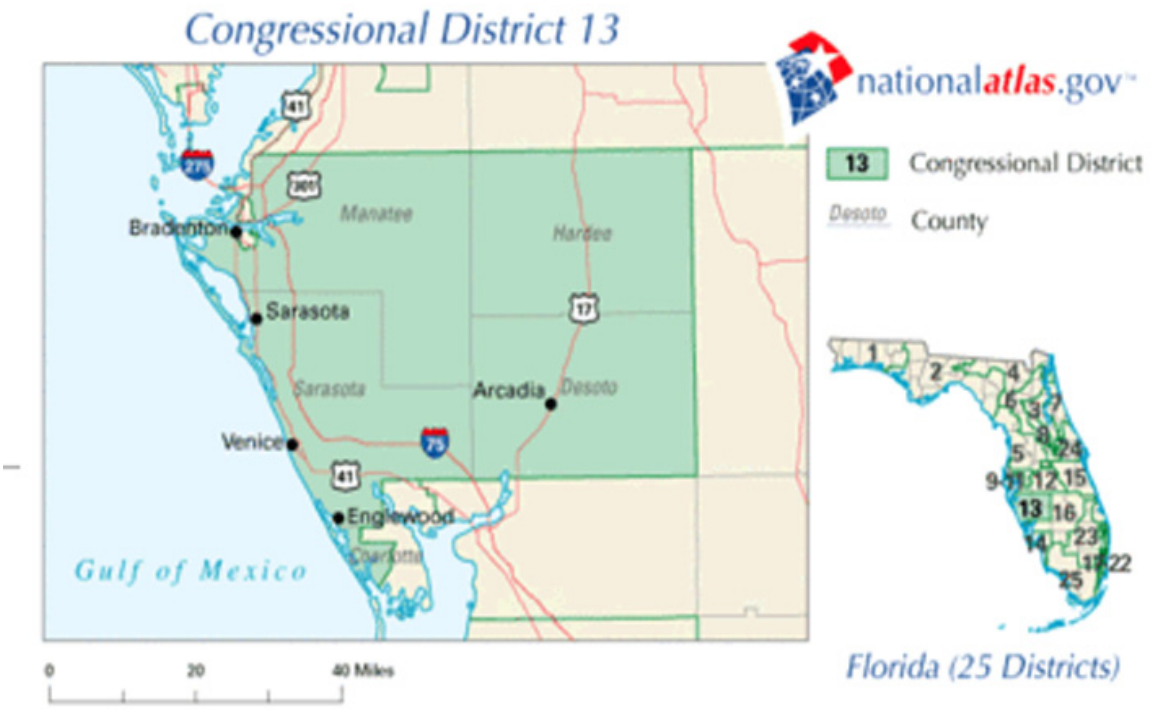

Figure 2. Screen shots of the 1st two (of 21) pages of the Sarasota County 2006 touch screen ballot.

OFFICIAL GENERAL ELECTION BALLOT
SARASOTA COUNTY, FLORIDA
NOVEMBER 7, 2006

CONGRESSIONAL

UNITED STATES SENATOR
(Vote for One)

Katherine Harris	REP	☐
Bill Nelson	DEM	☐
Floyd Ray Frazier	NPA	☐
Belinda Noah	NPA	☐
Brian Moore	NPA	☐
Roy Tanner	NPA	☐
Write-In		☐

Page 1 of 21
Public Count: 0

Next Page

U.S. REPRESENTATIVE IN CONGRESS
13TH CONGRESSIONAL DISTRICT
(Vote for One)

Vern Buchanan	REP	☐
Christine Jennings	DEM	☐

STATE

GOVERNOR AND LIEUTENANT GOVERNOR
(Vote for One)

Charlie Crist Jeff Kottkamp	REP	☐
Jim Davis Daryl L. Jones	DEM	☐
Max Linn Tom Macklin	REF	☐
Richard Paul Dembinsky Dr. Joe Smith	NPA	☐
John Wayne Smith James J. Kearney	NPA	☐
Karl C.C. Behm Carol Castagnero	NPA	☐
Write-In		☐

Previous Page

Page 2 of 21
Public Count: 0

Next Page

Table 1. Florida's CD13 Race in Sarasota County For All With Votes in 5 Out of 5 Statewide Contests

Number of Democratic Votes in the Other 5 Contests	Total # of ballots	Recorded and Missing Votes in the CD-13 Contest for the U.S. House of Representatives					Proportional Allocation of the Undervote		Change in Buchanan Minus Jennings Tally From Including the Undervotes
		Buchanan	Jennings	No vote	Recorded % for Buchanan	% Under-vote	Buchanan	Jennings	
Early Voting									
5	10,764	122	8,655	1,987	1.4%	18.5%	28	1959	-1,932
4	2,789	151	2,250	388	6.3%	13.9%	24	364	-339
3	1,170	174	831	165	17.3%	14.1%	29	136	-108
2	1,167	346	664	157	34.3%	13.5%	54	103	-49
1	2,173	1,227	657	289	65.1%	13.3%	188	101	87
0	9,455	8,059	435	961	94.9%	10.2%	912	49	863
Election Day									
5	25,326	468	21,541	3,317	2.1%	13.1%	71	3246	-3,176
4	7,637	561	6,261	815	8.2%	10.7%	67	748	-681
3	3,629	691	2,529	409	21.5%	11.3%	88	321	-233
2	3,847	1,387	2,022	438	40.7%	11.4%	178	260	-82
1	7,305	4,402	2,116	787	67.5%	10.8%	532	255	276
0	29,364	25,676	1,359	2,329	95.0%	7.9%	2212	117	2,095
Sum of Both	**104,631**	**43,264**	**49,320**	**12,042**	**46.7%**	**11.5%**	**4,383**	**7,661**	**-3,279**

Figure 3. Undervotes in the House Race by Voting Venue and Partisanship of Other Votes: Among 104,631 Ballots with Votes Recorded in All 5 Statewide Contests