Journal of Official Statistics, Vol. 18, No.2, 2002, pp 165–179

Election Night Estimation

Warren J. Mitofsky[1] *and Murray Edelman*[2]

1. What It's Like on the Front Line

Our talk is about election night. For the most part we will tell you about what we did at CBS News. We also will include something about how those activities changed when the networks pooled their money in a common effort. We are going to tell you something about why the networks first engaged in this enterprise. We will give you some understanding of what it takes to simultaneously conduct 51 surveys of over 100,000 respondents and present the results to millions of viewers – and almost never get the outcome wrong. And because ''almost never'' is not ''never'' we will tell you what happened in Florida in the last presidential election.

We thought the place to start was to get you in the mood. We want you to know what it feels like on a typical Election Day to be one of us. It starts the night before when we hope all the last-minute details have been dealt with. Everyone has his or her assignment. Everyone has been trained and rehearsed: the vote collectors at the precincts, the exit poll interviewers, the analysts, those entering and reviewing vote returns and data, those who use the results in the television and radio studios. There are literally tens of thousands of people in a myriad of jobs. All our reference materials are where they should be. All the phone numbers we need are handy. Every computer system works. We are ready for Election Day, the culmination of the last two years' work.

A good night's sleep and we can be ready for a day that will last from the time we get up on Election Day until after the next night's network news broadcasts – some 36 hours later. Unfortunately the night's sleep is not very restful. Usually the phone rings too early in the morning about yet another problem and a few more missed details cross our thinking. The area where we worked was in the studio during our CBS days. Now Murray is in a neutral site and is working for all the networks, the Associated Press and most of the large newspapers in the country as well as an assortment of television stations in cities around the country. Warren is in a similar place, but only has to worry about CNN and CBS.

Shortly after noon we are anxious to get our first look at the first wave of exit poll results. This is the time when we confirm that the computer system really works; that the interviewers are doing their jobs; and that no official has kept our interviewer away

[1] Mitofsky International, 1 East 53 rd Street, New York, NY 10022, U.S.A. Email: mitofsky@mindspring.com, http://www.mitofskyinternational.com
[2] Voter News Service, 45 Main Street, Suite 900, Brooklyn, NY 11201, U.S.A. Email: murray.edelman@-vnsusa.org
Acknowledgment: The authors thank Scott Dingman for his assistance with the analysis of the Florida errors in the 2000 Election.

from the exit to the polling place. We also want to see if the contests we were told were landslides really are landslides and the close races really are close. We do it when the first round of exit poll results comes in. We do it again in mid-afternoon when the second wave is available. We try not to forget that sometimes the results change over the day.

During the hour before the polls close in the first states we have to be ready for the third and last wave of exit poll results. Kentucky and Indiana close first. The best way to handle the buildup yet to come is to triage. Someone else on the decision desk can confirm the winner of an expected landslide as we get closer to poll closing. And the very close races can wait, too. They cannot be called from an exit poll. Exit polls are not that precise. It is those races that are in between, the ones with about eight point margins, that we must concentrate on in the half hour before the polls close.

We review detailed results for each race on our list. If we can make a call we do, and if not we wait for more sample precinct vote returns. The biggest problem in this time period is not to lose track of races that have still to be decided. Even though the computer screen lists all races in priority order, a race can slip between the cracks for a period of time. We try not to let this happen.

What we are trying to do is give the results and not make a mistake doing it. Forget those cynics who tell you that this is a reckless race to be first on the air with a winner. That is not the goal. We want to get projections on the air as soon as we can once we feel certain we have correctly identified the winner. Fear of being wrong is the overriding emotion, not racing through projections.

Once those first two states close we hurry to look at the nine states that close at 7:00 p.m. and 7:30 p.m. That number is manageable. Any states not called at poll closing are assigned to other members of the team to watch, unless they are races of national interest. Then we watch them ourselves. These are races where an incumbent senator might be defeated, like Ashcroft in Missouri or Abraham in Michigan in the last election, or states that are key to an electoral victory, like Florida.

But for the most part we want to get started on the 8 o'clock states. That is the big rush. Eighteen states close their polls in the 8 o'clock hour and another twelve the next hour. Those two hours will be the big crush. It will test our capacity to collect and process all that data. It also will test the organizing we have done so we, or the analysts working with us, can review everything carefully enough to be confident about what we project. If things go as usual, there will be lots of senate, governor and presidential states where the winner is known with enough precision from the exit poll. If we cannot make a projection at poll closing, we will have to keep monitoring the race until there are enough real vote returns in our sample precincts to enable a projection.

We have help doing this. First, there are fine statisticians working with us on Election Night. Next, there is a computer screen that lists all the races in states where the polls have closed that have not been called. As more sample precinct returns come in the color for each state indicates the current status of the results. Yellow means all criteria for a call have been met, and light blue means the results are getting close to call status. With any luck we will not see a state colored red. Red states go to the top of the list automatically. That means a race that has been called may no longer meet all the criteria for a projection. For the states in red we look to see if we have to make a retraction. Usually this early warning sign turns out to be of no consequence. One of many criteria for

making a call may have been marginal. One precinct more or less may turn this indicator red temporarily.

Occasionally, we have a mistake. The sooner we recognize it and announce on air that we are retracting a projection, the less trouble the errant projection will cause the news reporting. It is bad enough to have made a mistake, but it must be corrected as publicly as it was made. If a state is crucial to control of the Senate or to an electoral vote victory, the projection can seriously mislead the election reporting. If the state is not key to some trend, the wrong projection still has misled viewers. On the other hand we do not want to retract a call too quickly. We do not want to issue a retraction just because the leading candidate is not ahead by a big enough margin to satisfy our call criteria. We want to be reasonably sure we were wrong.

As the night progresses it becomes clear that some races are just too close to call from sample precinct returns. Our next source is the county vote tallies. In the New England states they report for towns or cities. These county and town returns trickle in after the polls close and eventually reach 100 percent of the precincts reporting in the early hours of the morning. We model these returns also. They can be used in combination with the precinct samples or they can be used separately. For the closest races this is what we rely upon for projections. It can be in the middle of the night or sometimes it takes until the next day. And sometimes the best call we can make when all the votes are counted is ''Too Close to Call.'' More about that when we get to Florida.

2. Why the Networks Do It

Curiously enough, the decision years ago to spend all that money on election night coverage was inspired at CBS by a mixture of pride, competition, embarrassment, and determination to overcome past mistakes.

In television's infancy following World War II, television news was fighting for legitimacy. It wanted the same credibility for news reported on television that newspapers had. The transfer of power in a democracy is probably the most important news event of the year. Certainly it is the most important event where news coverage can be planned in advance. We take this transfer of power for granted in this country, but we should not. If you want a feeling for its importance all one wants to do is follow the news of any emerging democracy when control of government switches hands. Any legitimate news organization will cover that story as best it can. Election news is a story made to order for television. Television has national coverage. And, it has the impact of immediacy that newspapers could never have.

Before television's domination of election night coverage the wire services gathered the votes on election night and the days following and sent them to the newspapers. Reports heard on radio were taken from the wire services or from the newspapers. Listeners would hear for example that according to the Baltimore Sun the next governor of Maryland was Smith. Or according to the New York Times the winner in the New York senate race was Jones. That is the way broadcasters reported elections up to 1948. After that, everything changed.

In the 1950s statisticians and computers got into the election night act. The computer makers got a plug and charged nothing. RCA, which owned NBC, provided its computers.

CBS made a deal with Remington Rand for UNIVAC. For a variety of reasons, including the egos of some network executives, the competition to decide elections first between NBC and CBS was important to news division chiefs. Beyond the ego satisfactions, there were also arguable gains to be made with respect to enhancing the network brand, generally, and that of its evening news broadcast, specifically. At a time, moreover, when presidential primary broadcasts were major network news undertakings, the ability to announce a winner first (and correctly) and to brag about it afterwards was also worth something in those early years.

It was a time, too, when all of the broadcast networks were highly profitable, and the costs of calling elections were within tolerable limits with respect to both the overall network and even the news division budgets.

In 1952 and 1956 the way computers were used in election coverage was pretty laughable. Incomplete returns from selected counties were used in a regression model. At CBS, UNIVAC and Charles Collingwood gave the ''odds'' on a winner. The odds that John Kennedy would beat Richard Nixon were 3-to-2. That sort of thing. That was embarrassing. CBS was eager to usher in the computer age with some accuracy and sophistication and this was not going to do it.

All was well at CBS until the 1960s. By then Huntley/Brinkley was dominating the evening news ratings. The prevailing CBS attitude at the time was ''We gotta do something!'' Among the things that were done was to replace Douglas Edwards with Walter Cronkite. The political conventions, starting with 1948, were really big deals. Cronkite had been anchoring the political conventions and doing quite well. It was taken, as a matter of faith that whoever won the ratings at the conventions would fairly quickly rise to the top of the evening news ratings.

Planning for a new type of election night coverage began in 1961. CBS got into a partnership with the pollster, Louis Harris. Harris had worked for John Kennedy during the 1960 presidential campaign. CBS also got IBM into the act as a partner. IBM was anxious to promote computers in those days. By early 1962 they had the Harris version of election projections. It was called Vote Profile Analysis. When NBC got around to it in 1964 they called it EVA, Election Vote Analysis, and ABC called their system Research Selected Voter Profile (RSVP). Harris's innovation was to sample precincts within states, albeit with quota samples. He successfully projected the winners in 13 senate and gubernatorial races in seven states.

Many people believe the vote count they hear on election night comes from some official state agency. It does not! Let us repeat that – the vote counts on election night do not come from an official state agency. They come from the media. Until 1964 the wire services were the only ones tabulating the vote count. But because the wire services were very slow at collecting the vote CBS and the other networks each counted it during the presidential primaries in 1964. There were many different counts of the vote during the primaries. The last primary of the year was the hotly contested California Republican primary. When CBS News president, Fred Friendly, got off the plane from California in New York the next morning he was greeted with the newspaper headline, ''Networks Take Pratfall.'' That morning the wire services were still reporting Nelson Rockefeller leading in the vote count. Apparently the newspaper believed the wire services and not the networks. The wire services' still incomplete tallies were based on the votes

they collected at California's 58 county election offices. Meanwhile, ABC, NBC, and CBS each had a reporter at each of California's 25,000 precincts. In some places phones were installed to speed the reporting. These 75,000 reporters produced a more advanced count than the wire services. The networks' own vote tallies up the networks' projections of Barry Goldwater's victory. The News Election Service was born later that day.

The networks were prepared to do something similar for the general election in order to count the vote, but the confusion caused by the conflicting counts made it clear that a combined effort would spare the public from the confusion of independent counts while it saved the members a lot of money.

Thereafter, the counting of votes has been a pooled operation. By most data processing standards the quality control on the incoming data has been very unsophisticated. There has never been an assessment of the accuracy of the incomplete count at different time points on Election Night and there has been no meaningful statistical quality control of the vote tally while the tally is in progress. The only assessment of the accuracy of the media's counting has been whether the final tally agrees with the official tally some days or weeks later. From experience we know that a final media tally of a state's votes can vary by up to half a percentage point from the certified result.

ABC, NBC, and CBS continued through the 1960s and 1970s doing essentially the same thing on election night. The cutthroat competition for the advantage of their news divisions' long-term interests faded somewhat. Election night coverage eventually matured into sound news reporting. The networks still competed by selecting their own sample precincts and using them to make state projections. There were congressional hearings following the 1964 election about the effect of projections on west coast voters, but there were no controversies from 1966 through 1979.

CBS continued using the Harris quota samples through 1965 and a slightly improved method in 1966. The significant problem that year was not the sample or the estimator. They had what is called a quality control problem. The Harris quality control used two tests to decide whether the precinct votes were correct. First, the Harris staff looked to see if a precinct's percentage vote for a party differed more than some fixed amount from the vote in a past election. They made a similar test on the total vote. If the precinct failed either of these tests its votes were dropped from the estimate. This turned out to be a mistake in two states in 1966. In both the Maryland and Georgia governors' contests, the Democratic candidates were more conservative than their opponents. As a result, precincts with lots of African-American voters did not vote for the Democrats as they had in the past. In Maryland, for example, they voted almost 90 percent for the moderate Republican, Spiro Agnew. The Harris quality control procedure eliminated these deviant precincts from the samples in both states, which led to erroneous projections. Those precinct votes were correct.

In 1967, our first year at CBS, we changed things significantly. Working as a team we had guidance from Joe Waksberg, who had been our boss at the U.S. Census Bureau. Our partnership continued without interruption through 1993.

At this point we should point out that CBS tried to hire Morris Hansen as their first choice to run their new in-house election unit. Some years after we worked there we discovered a memo listing the names of prominent sampling statisticians that they contacted. All came from solid statistical backgrounds. We never found out why CBS

pursued such people at that time, but we always thought they deserved credit for going in that direction. We like to think it was due to the influence of Frank Stanton, who was the president of corporate CBS and a former survey researcher.

The sample selection we introduced was probability based. It was a stratified sample and precincts were selected in a single stage with probability proportionate to the total vote in a recent past election. The IBM team that had been around during the Harris days asked if we were going to find faster reporting precincts than the ones we selected and swap them. They said that is what Harris did in 1966. They were convinced we would do this to speed up the reporting on election night. Needless to say we did not. This has not been done in all the years since, even though we understand it was common practice at the other networks.

In the summer of 1968 Dick Scammon, who had been director of the U.S. Census Bureau when we first worked there, asked Warren if he had visited each of the sample precincts. He said if we did not go out and look over the neighborhoods we would not be able to call an election with enough assurance. We did not do that either.

We got some good advice from Joe Daly, the chief mathematical statistician at the U.S. Census Bureau in those days. When we told him what we were going to do for CBS he said your main problem is quality control of the vote returns that come pouring in on Election Night. And he was right.

To handle this problem we adapted an idea developed by Max Bershad, who was also a U.S. Census Bureau statistician. Bershad's paper was about controlling the mean squared error in the Current Population Survey when unexpected new construction yielded a very large number of housing units in a segment that had only the minimum number of housing units when it was selected. The principle was to reduce the weight on the segment, which reduced the sampling error and added a bias component. The net result was a reduction in the mean squared error.

We adapted that principle in order to control the size of the total vote in a precinct and to control the variation in a party's vote. What we wanted to avoid was a report of an unusually large precinct vote. This could come from reversing the digits in the vote, or reporting the votes from two precincts combined rather than one, or some other error in vote reporting. The party vote might differ from a precinct's party vote in an earlier election by much more than other precincts. This could be the beginning of a trend, an outlier, or an error. The reason was not important.

Here is how we handled the problem. To control the effect of a large total vote what we did was calculate the contribution of each precinct to the total sum of squares. Any precinct that exceeded a threshold had its weight reduced so that the precinct would not exceed the threshold. We used an iterative procedure for a maximum of six precincts. The control on party vote was similar.

Estimation was another matter. On Election Night we were going to make estimates after the polls closed. While the full sample for a state might be between 40 and 120 precincts, we did not plan to wait for all of them to report their votes. As each new precinct in a state reported its vote we were going to make a new estimate. We tried simulating Election Night using a variety of estimators: ratio estimators, regression estimators and difference estimators. We expected regression estimators to work best. The problem was estimating the coefficients from the relatively small samples. The sample size,

especially early in the night, is not ideal for estimating the regression coefficients. Under these circumstances ratio estimators were more stable.

The ratio estimates we used had a past party vote in the denominator rather than any other characteristic. No other characteristic is as highly correlated with current vote as past vote. On the occasions when party vote is not correlated, there is another estimator to turn to that uses the past total vote from the sample selection race as the denominator. Under these circumstances the estimator reduces to a simple unbiased estimator once all the precincts have reported.

Before making these ratio estimates there is a poststratification of the precincts. There are two different poststratifications. One is based on geographic areas within a state. This comes from dividing a state, usually along county lines, into about five strata. The second poststratification divides all precincts within a state into five homogeneous groups with respect to the percentage vote for one of the political parties in a recent election. In states with a substantial number of African-American voters there is a sixth stratum composed of precincts where the residents are mostly black. They usually vote for the Democratic candidate, but as we saw in Maryland and Georgia in 1966 they sometimes move as a block in a different direction than other Democratic voters. Early on Election Night the strata are collapsed. As more precincts report, the strata unfold to their maximums.

3. Why Exit Polls

In 1967 we introduced exit polls at CBS. We were led to interviewing voters at the polling places because of a story we heard from George Fine. Fine ran a market research company. Through him we hired a field staff that did our research in county election offices all over the country. Fine told us about how he interviewed moviegoers after they had seen a motion picture. His purpose was to get feedback for a movie company so they could revise the film before distributing it nationally.

We interviewed voters at the polling place after they voted. We did it so we would have a better idea of what to expect. We were concerned that we would be misled by the results from the earliest sample precincts as they trickled in with their vote returns on Election Night. We thought that these interviews at the polling place would be better than what we would learn from pre-election polls. We interviewed voters at the precincts for the first time in Kentucky in 1967 for a lone governor's race. In 1968 we used them for the six primaries CBS covered and in 21 states for the general election.

Roy Wetzel, who headed the NBC political unit from the mid-1970s and through the 1980s, told us that the first exit poll he was aware of at NBC, was done in 1973. Others at NBC confirmed this. ABC did its first exit poll in 1980. It was NBC that introduced the name exit polls.

An exit poll is not much different than having vote returns in sample precincts. But instead of real vote counts from precinct officials after the polls close, in an exit poll an interviewer stands outside the polling place and hands a questionnaire to a sample of voters throughout the voting day. One of the questions asks for whom the person voted. One other difference is that the exit poll is conducted in only a subsample of the sample precincts that are used to collect vote returns.

This is mostly due to cost. It costs almost four times as much to have an interviewer at a polling place all day as it does to collect the vote from a precinct official after the polls close. An estimate based on exit poll data is not nearly as accurate as one based on the same size sample of real vote returns. The exit poll has within precinct sampling error and bias due to nonresponse and response error. A typical projection based on an exit poll might have a margin between the two leading candidates of eight percentage points or more. With real votes in the same number of precincts the margin could be about three points less for the same level of assurance.

By 1970 exit polls became an analytic tool. Early estimates based on real vote returns proved to be reasonably reliable. As a result we did not need to spend the money on an exit poll for that purpose, as they were not used for projections. Once we added questions to the exit poll we were able to know which constituencies voted for which candidates. What the issues were that mattered. What voters wanted from their newly elected officials. No longer would the reporting be dependent on analyzing precinct votes or the spin of the campaigns.

In 1980 everything changed. The pre-election polls had promised a close race, but they were misleading. Ronald Reagan won a resounding victory by about ten percentage points. He won 45 states and 489 out of 538 electoral votes.

NBC projected Ronald Reagan's victory over Jimmy Carter at 8:15 Eastern time, almost three hours before the West Coast polls closed. They announced Reagan victories in 22 states that had closed their polls. Those states gave Reagan exactly 270 electoral votes. NBC made eleven of those projections based on exit polls.

For that election CBS and ABC used actual vote returns for their projections and not exit polls. ABC went on the air about 9:50 with their presidential victory announcement. CBS was the laggard. We did not have enough electoral votes for Reagan until 10:20, after Carter's ill-timed concession.

From 1982 on all three networks used exit polls along with vote returns in sample precincts. This drove the cost of elections up a notch.

By the end of the 1980s the networks were anxious to change the way they covered elections. The cost of election night was staggering and the competitive advantages that drove the early years no longer existed. Also, opponents of early calls became a constituency that harassed the networks' other business interests in the halls of Congress. NBC disbanded its election unit following the 1988 election. It was planning to join a joint network operation whenever one got organized. Price pressures at CBS made it willing to cooperate with the other networks in a joint election coverage consortium. CNN had no election coverage up to that time and was eager to join a consortium with a cut-rate price rather than take up the burden unilaterally. Of the two proposals to organize a pool CBS's election unit won out over ABC's. We became the source for projections and analysis on election night for all the networks. ABC, not altogether happily, fell in line. A consortium, of course, enormously reduced everyone's cost.

The new group was formed in 1990 after a year of negotiation. It was called Voter Research and Surveys. The News Election Service continued to collect the vote as a separate consortium until 1993 when both VRS and NES were combined to form Voter News Service or VNS. That is what we have today. In the early 1990s the cost per network ran at about two million a year, more or less. This was for VRS and NES. This

was not really very much compared to what they had been spending. It was about one-third.

In the first years of VRS, there was no competition to call elections first. All projections were made by VRS. But, aside from ABC, nobody who mattered cared – certainly nobody on Wall Street, where the economic health of the networks became the only thing that mattered. ABC started the competition anew by making its own projections in 1994. The other networks were embarrassed when it took VNS about an hour or so to catch up to ABC. ABC had announced the George W. Bush win in Texas, and the losses of Mario Cuomo in New York, and Ollie North in Virginia, among others, while NBC, CBS, and CNN were unable to report on those races. Ever since, all the networks have had their own capability for calling elections.

For the broadcast networks today, the costs are low compared to what they had been spending before they pooled the projections and exit polls. Primary election broadcasts have dwindled to the vanishing point. So, too, have off-year election nights. What the broadcast networks did in 1998 and will likely do in 2002 is announce a few winners on the hour and then offer an hour of election coverage from 10 to 11 p.m. Cable networks are another matter. There, reporting news is the core business of their company, and the competition for bragging rights about their election broadcasts is very important to them. The competition among the cable networks is like it was at the broadcast networks in the sixties. Moreover, while only the presidential election broadcast is given lots of attention by the broadcast networks, for cable the off-year election is a big deal. If you want to watch lots of election coverage in 2002 cable will be more likely to have it.

4. Florida 2000

No discussion about Election Night would be complete without discussing the errors in Florida in the last presidential election. Over the years there have been very few bad projections. From 1967 through 1988, which was the last year we worked only for CBS, we had five mistakes out of the 1,500 or so state elections we made projections for in that period. During the 1990s the VNS pool had only one error in about 700 projections. However, we have since learned that 99.7 percent is not good enough when that rare event is highly visible.

For those of us who make Election Night projections, making a mistake has always been the one thing we most strive to avoid. The risk of making a mistake twice in one night in the same state is something we actively guarded against on the few occasions when a mistake was made. That nightmare, however, was not as bad as what actually happened in Florida last year. (You can't make this stuff up.) Florida was the key to the outcome of the presidential election, and there were two bad projections in the state. The first was for Albert Gore early in the night. That was made by VNS and each of the networks almost an hour after most of the polls closed. For the second blunder the networks called George W. Bush the winner when almost all the precincts in the state had reported. That projection stained the solid journalism in the reporting of elections for the last 38 years. Voter News Service and the Associated Press were widely blamed along with the networks even though neither of them called Bush the winner. VNS, however, did provide the election returns used by the networks to make the wrong calls

and the Associated Press, for most of election night, reported the same incorrect vote counts.

5. The Gore Call

Here is what happened in Florida. At seven o'clock when 95% of the polls closed there was a projection that the Democrat, Bill Nelson, won the Senate contest. This was based on the exit poll. There was no projection for President. A call for President based on just the exit poll looked too risky, so everyone waited for real vote returns in the sample precincts. The real votes confirmed the Gore lead and Gore was called the winner 50 minutes after the polls closed.

There was no rush to this conclusion. Before the call was made we checked the exit poll against real vote returns to see if there was any possible overstatement in the exit poll for one of the candidates. The model produces a variety of estimates, each with different statistical assumptions. They all showed Gore ahead by margins of five to ten percentage points. They all showed *t*-values between 2.0 and 4.8. The best estimator, the one with the lowest sampling error, said that Gore led by 7.3 points with a *t*-value of 3.8. From the early precincts with actual votes it appeared that the exit poll slightly overstated *Bush!* We saw the same Bush overstatement in Kentucky, which was the only other state where we had data at that time for this calculation and this gave us even more confidence in the projection.

There are four estimators for manipulating the precinct data. Two are simple unbiased estimators and two are ratio estimators. Each type of estimator also is made using two types of post stratification, a past party vote stratification and a geographic area stratification. The estimator with the smallest sampling error on the difference is considered the ''best estimator.''

The best estimator at the time of the Gore call (which was at 7:50 p.m., EST) in Florida was the ratio geographic estimator using the 1998 Governor's race as the past race. After the election we had the luxury of time to examine the various estimates more carefully. This is what we learned:

Error breakdown of the Gore lead at 7:50 P.M.
Geographic ratio estimator using different past races in the denominator:

	98 Gov.*	96 Pres.*	98 Sen.*
QC Modification error	0.4	0.6	0.1
Survey error	2.6	2.6	2.6
Estimation/sampling error	2.6	0.4	0.3
Absentee vote est. error	1.7	1.3	1.3
Total error	7.3	4.9	4.3

*Past comparison race used by the ratio estimator.

The first column gives the breakdown for the ratio geographic estimator using the 1998 Governor as the past race, which was the best estimator used on election night to make the call. The 7.3 percentage point error has the following components. The quality control procedure contributed 0.4 percentage points to the total error. The sampling of voters

within precincts contributed 2.6 percentage points. This includes within precinct sampling error and any effects of nonresponse to the exit poll. Then there is a contribution to the total error from the sampling of precincts and the estimation process. That, too, contributed 2.6 percentage points to the total error. And lastly there was the effect of the missing absentee vote in the model.

We knew in advance that our model was underestimating the effect of the absentee vote because it was based on the 1998 Governor's race where absentees were only 7% of the vote that year. We expected the absentees to be closer to the 10% level of the 1996 presidential race. As the model did not take the 10% level for absentees into account, the decision-makers had to approximate the increased absentee effect outside the system. Even so, no one expected the 12% absentees that we actually encountered in 2000.

There are several past elections stored in the Election Night database. The correlations are computed between the Republican vote percentage in the current election and all past elections. The past contest with the best correlation is used in the ratio estimates. The other two columns in the table show the geographic ratio estimates for different past races that were in the database but were not used on election night. The comparison is illuminating. As luck would have it Bush correlated best with his brother's vote in the governor's race two years earlier, but only by a small amount over the correlation with Dole in the presidential race. Had the presidential race been selected for the ratio estimator, the estimated Gore lead would have been 4.9 percentage points. The correlation with his brother's race was only three-hundredths larger, but the resulting overall error was 2.4 percentage points larger. The 1998 Senate race had the lowest correlation, and it showed the smallest lead for Gore.

One clear lesson from this analysis is to make ratio estimates using other past races on election night, and not just the best-correlated race. At the time our models were originally developed, simplifying assumptions had to be made just because of the limits of what could be presented and comprehended in the time available for making decisions. With the modern user interfaces in VNS's new computer system, the decision maker will be able to model the value of this new information and anticipate other problems without being overwhelmed by the data.

The table below illustrates the different critical values that we would have had if we used other past races for the ratio estimates. To complete the analysis, we subtracted what would have been a reasonable pre-election guess of the effect of the absentee vote from the final election result and then divided the estimated lead by its standard error.

Error on the Gore lead from the Election Day vote as of 7:50 P.M.
Geographic ratio estimator using different past races in the denominator.

	98 Gov.*	96 Pres.*	98 Sen.*
Total error	7.3	4.9	4.3
Less absentee adjustment	1.0	0.5	1.0
Adjusted	6.3	4.4	3.3
Standard error on Gore lead	1.9	1.8	2.3
Critical value	3.3	2.4	1.4

*Past comparison race used by the ratio estimator.

In this case, the best-correlated race was not the best race to use. Our criteria would not have led us to call this race for Gore if any other past race had been used as the denominator for the ratio estimates. The minimum criterion used for a call status is a *t*-value with a probability of 1-in-200. No one using these models expects them to be totally accurate. The post-election analysis showed that we understated the sampling error and therefore we were understating the probability of an error. Even with this problem the race was very clearly an outlier.

Sometimes good samples produce bad estimates. Clearly, this race was an outlier. The fact that the Gore projection was wrong does not mean it was made irresponsibly or that the models were no good. Every once in a great while a carefully designed and tested sample yields an estimate that is wide off the mark. Of the thousands of races we have participated in this is only the second time we have seen this much solid evidence for a projection that turned out to be wrong. In all the years we have been making election projections we have only had one other error due to an outlier. That was for the presidential race in Illinois in 1988. That mistaken call was for Dukakis. Unlike Florida in 2000, the Dukakis call had no consequence for the outcome of the presidency.

Elections provide a unique opportunity for an empirical evaluation of the statistical methods used. They are one of the only applications where the population is counted soon after the estimates are made. If other applications were subjected to the same scrutiny there is no telling what improvements might be made in statistical methods.

6. The Bush Call

The announcement starting at 2:15 in the morning that Bush won Florida and the Presidency was the more troubling of the mistakes. While we do not want to minimize the confusion caused by the Gore call, that mistake was not as significant: the Bush call decided the presidency. The Bush mistake was not the result of sampling error; it was the result of operational error and a poor model of the outstanding vote. With any reasonable improvements this kind of mistake can be avoided in the future.

Here is what happened. The Bush projection was based on an analysis of the county vote tallies supplied by local election officials. (It was not based on exit polling or on a projection model.) VNS reporters, stationed in each county, forwarded the vote counts periodically to the VNS processing center. Because they were all looking at the same vote returns all the networks called this race within a few minutes of each other. Fox made the first projection at 2:15 a.m. VNS and the AP did not call Bush the winner. They deserve credit for their restraint. However, VNS's handling of the vote tally played a major role in what everyone else knew, and the AP did not put out a caution about the closeness of the vote count on their election wire until an hour after the first network projection was made. The vote tally in Florida, prior to the network mistakes, seems to have been processed just like it was in all the other states, and not like a vote tally that would decide the outcome of the presidency in the closest election in years.

The Bush-wins-Florida call was made when the vote count had Bush leading by 51,000 votes with 97% of the precincts counted. An analysis showed that Gore would need about 63% of the *estimated* 179,000 votes outstanding to catch Bush. The only counties where Gore was getting that high a percentage of the vote were Broward where only

2% of the vote was estimated to be outstanding and Palm Beach where 15% was estimated outstanding. Gore also would pick up votes in Miami-Dade County, but collectively this would not be enough to catch Bush. Using these estimated votes, all three counties combined would only make up about 20,000 votes of the Gore deficit. Other counties with vote to be counted were supporting Bush. During the next 30 minutes the Bush win looked even more certain and the networks remained confident. Still there was no projection from VNS or the AP.

In our post-election evaluation, we learned that these numbers were based on a large error in Volusia County – where Daytona is located – and poor estimates of the size of the outstanding votes in others. Below is a table of the reports from Volusia County.

Time	Presidential vote for Volusia County, Florida Election Night 2000					
	Precincts		Vote			
	IN	OF	GORE	BUSH	BCHN	NADR
8:40 PM	145	172	73,363	57,308	361	2,148
9:54 PM	165	172	82,619	63,265	396	2,436
2:08 AM	171	172	72,152	73,146	396	2,436
2:49 AM	172	172	97,063	82,214	396	2,436

At 2:08 a.m., shortly before the network projections, an incorrect vote was entered into Volusia County by VNS, which caused an increase of 20,000 votes in the Bush lead over Gore. The error was caused by a problem with the tabulation software at the county election office. The AP had reported the same incorrect vote in its tally since about 12:30 a.m., and corrected it at 2:16 a.m., about the same time as the first Bush call. VNS did not correct it until 2:48. There also was a 4,000 vote-count error overstating Bush in Brevard County.

The Bush lead in the state at 2:15 a.m. was not 51,000 votes, as VNS reported and the network analysts believed. It was closer to 27,000 votes. Couple that with the expected gain of 20,000 votes for Gore in the three counties just mentioned and the Bush lead was only 7,000 votes.

There also was a problem with the estimated number of votes outstanding. The 179,000 votes that were thought to be outstanding turned out to be 359,000. About half that miscalculation was from Palm Beach County, where Gore was doing well.

It is clear that several things contributed to this error. First, there is almost always some error in the vote tally. With all the votes in a state counted and verified we have found errors in the past of up to half a percentage point. The errors in Volusia and Brevard counties alone were almost that much. Second, the procedure used to estimate the number of votes outstanding is too simplistic. A better model is needed than we have been using.

Third, and most important, it is necessary for analysts of this late-night vote count to have knowledge of any drop in the cumulative vote count in a county or corrections to the tally within a county. Because there are many drops and corrections in the vote tally during election night this is how VNS handled them: depending on the seriousness of a correction or a vote drop, it is reviewed first by the state manager, then by the area

manager, then by the overall manager who could then notify the analysts. No one notified the analysts at VNS or at the networks when this correction happened – a simultaneous drop of 10,000 votes for Gore and an increase of 10,000 votes for Bush in Volusia County. The AP does not regularly notify subscribers of significant vote drops and did not know what their vote drop in Volusia County was responsible for their state total getting much closer until the next day.

While the network analysts, along with the country, were focused on the closeness in Florida where slight changes in the vote were highly significant, the VNS people doing the tabulation were disconnected from the national drama and were focusing on keeping the data flowing as accurately as possible in all of the states. The VNS manager who was reviewing the vote drop in Volusia County was confident that the error really was a correction and that the shift in votes was not all that unusual so he did not pass the correction up to the overall manager who could have passed it on to the decision makers.

Had the vote tabulation operation alerted the network analysts to the entry of the first vote adjustment, the analysts might have been suitably suspicious of the vote count *before* Bush was called the winner. Similarly, if the network analysts had had access to the entire trail of inputs into each county, they might have discovered and questioned that vote drop themselves and given the closeness of the race, might have waited.

7. The Future

VNS is in the process of evaluating and re-programming all its systems. A statistical committee representing all the networks has been examining all the statistical methods. Tonight's speakers and discussant are part of that committee and other notable statisticians from the outside have reviewed the models and made suggestions. For the most part the system we put together over thirty years ago was sound. Even so, after all these years it needs updates and improvements.

When the computations were first designed absentee voting was not a problem for the models. Today it is. There are significant numbers of absentees in a number of states and a proper accounting for them means that the models will have to more clearly distinguish between the estimates of election day votes and the estimates of absentees. The absentee vote will need to be modeled so that analysts will be able to have bounds for its effect on a state's vote.

Another problem concerns the estimate of the sampling error. The post-election analysis of our statistical methods turned up a deficiency that must have existed all the years we have been using these methods. The sampling error calculation *understates* the magnitude of the sampling error. The understatement is greatest when a small share of the total sample has reported. As more sample precincts report their vote the understatement of the sampling error decreases. It vanishes when all sample precincts have reported. We have been using the Taylor approximation of the sampling error. A sampling error calculation using the Jackknife method has a similar understatement. The cause of the understatement appears to be our use of incomplete sample returns to make estimates. As this is unavoidable for this application, in the future we will increase the sampling errors by a factor inversely proportional to the number of precincts reporting that will

be determined from historical data. Had we been doing this in 2000, all other things being the same, we might not have made the Gore projection at 7:50.

In Florida we saw the effect of selecting one past race rather than another for the ratio estimate based on the best correlation. We have found through our post-election analysis that this selection procedure has the effect of reducing the variance of the estimate without a commensurate decrease in the actual error of the estimate. In the future when VNS makes this kind of selection, it will include a factor that increases the estimate of the error appropriately. VNS also will be using multivariate ratio estimators, which are more likely not to understate the error in this way. They may also prove to be the most stable.

The end of night analysis of the vote count for very close contests needs improvement. The estimation of the size of the outstanding vote is one problem. Also, analysts must give greater emphasis to the possible errors in the counting and tallying of the vote by election officials. VNS and the AP need to coordinate their vote validation. They also need to report to the analysts on serious problems with the vote tally.

Errors in the vote tally are likely to be repeated because of the difficulty of the task and the fact that we have no control at the local level where the vote is counted. However, VNS will have a more sophisticated quality control on the vote input. The history of each county's vote and the evaluation of that vote will be available to the analysts throughout the night.

We look forward to the elections in 2002. It will mark a new beginning. It will demonstrate that VNS and the various decision desks learned from their mistakes. It also will show that the mistakes of the last election were a fluke.

Received March 2002