

A STUDENT-RUN EXIT POLL SUCCESSFULLY PREDICTS 2008 ELECTION RESULTS

ANNALAISSA JOHNSON, ANDREW MCCARTY, JASON PARSLEY, AND ROBERT REINHARD

ABSTRACT. In 2008, students in a first-year seminar conducted an exit poll to predict election results in their county. This article, written by three of these students and the seminar's professor, details their efforts and describes their results. They correctly predicted four of the five contests they surveyed; the fifth served as a useful lesson on how bias can affect polling. The article concludes with a discussion of the minimal resources and background required to attempt such an effort.

1. INTRODUCTION

If you watched coverage of the 2008 election, you might recall that the major news organizations sometimes predicted a winner in a state immediately after the polls closed, before any precincts had reported their results. For other races, the media might declare a winner after only a few precincts reported results; sometimes, the candidate who was trailing was declared the winner. Yet each major media prediction was correct that fall; how did they prognosticate so well? Were the media just speculating historically (e.g., *a Republican has won this state for the last 12 years, so it is bound to happen again . . .*), or did they know something the rest of us did not?

The media organizations did, in fact, have access to information that the public did not – data from surveys of people who had just voted. Such surveys are known as *exit polls*, and along with actual reported election results, they constitute the major data source for predicting winners on election night. Of course, these polls are not always correct; both Florida in 2000 and Ohio in 2004 showed major discrepancies between presidential exit poll data versus actual results. In the wake of such controversy, only one organization, Edison Media Research/Mitofsky International, now conducts exit polls for the media. Their precise statistical method is proprietary. They release the data only to their customers, typically major media companies, who (eventually) pass some of the data along to the public.

In fall 2008, a seminar at Wake Forest University conducted its own small-scale exit poll. Using a simple random survey, we measured the outcomes of five different races in our home county (Forsyth Co., North Carolina) and made our own predictions, available online within an hour of polls closing. By surveying a few hundred voters, we correctly predicted the winners of four of these races (president, senator, governor, local bonds), all within a few percentage points. This was a wonderful learning opportunity for students, many of whom had no knowledge of statistics prior to this course.

In this article, written by three of the seminar students and their professor, we describe how we conducted our exit poll and detail our results. Like all exit polls, we collected demographic data

from the voters, and we use this to correlate voting trends by age, race, gender, etc. We include a discussion the flaws inherent in our survey and detail how this might affect the accuracy of our results, especially for our one incorrect prediction. The final section describes the educational impact of this endeavor and encourages other groups to try this approach. For more information, please refer to our website: <http://www.wfu.edu/~parslerj/exitpoll/>.

This article specifically does not discuss or analyze the controversy behind exit polls of the past. There is a tremendous amount of literature on this subject already, and often it leads to quarrels along party lines, which we happily avoid. We do, however, begin with some background regarding exit polls.

2. EXIT POLL BACKGROUND

Exit polls are a recent idea in our political history; the 2008 presidential election was only the tenth with exit polling. Warren Mitofsky, the executive director of CBS News, conducted the first prominent exit poll in 1967 for Kentucky's gubernatorial election. Later, he expanded exit polls to a national scale for the 1972 presidential election (and correctly predicted Nixon would win). Following this, exit polls became widespread and popular; not only could they predict who would win, but the demographic data they collected helped answer why a candidate won.

The first significant controversy involving exit polls occurred just a few years later. In 1980, exit polls predicted a landslide victory for Reagan over Carter; NBC used this data to declare a winner several hours before the polls even closed on the west coast. The timing led to political uproar – did NBC's prediction discourage voter turnout? Consequently, the major news networks agreed not to release exit poll results or to declare winners for a state until after all of its polls have closed.

Although exit polls have provided valuable insights into elections, they have also produced some faulty results. By 2000, the major news networks and the AP had pooled resources to use one polling group (Voters News Service) for exit polls. That year VNS, and some networks, predicted that Gore had won Florida; hours later they switched their prediction to Bush; finally they settled on "too close to call". After the demise of VNS two years later, the networks created the National Election Pool (NEP) run by Edison/Mitofsky. Not free from mistakes, NEP predicted Kerry winning Ohio in 2004, while Bush won by 2.0%; for an explanation of this discrepancy, see for instance [3]. The networks continue today to use NEP as their only source of exit polls.

Other groups conducting exit polls rarely garner much attention. One exception is the student-led Utah Colleges group, which has conducted exit polls statewide since 1982 [2].

3. SURVEYING METHOD

The seminar students assumed full responsibility for all aspects of conducting this exit poll, beginning with deciding which contests to cover and writing a clear, concise survey through the actual interviewing of voters and ending with tallying the results and placing them online. The nature and size of the seminar constrained our survey design. Exit polls were merely one of three topics for our one-semester seminar, and no statistics background was required of the first-year students.

Thus, we chose a simple statistical design for our poll, which was able to survey at fewer locations than ideal.

Our exit poll predicted how our home county would vote in 5 contests:

- President
- U.S. Senate
- NC Governor
- County Commissioners
- Local Bonds

In addition to these five contests, our survey (available via our website) collected 8 types of demographic data: gender, age, race, party affiliation, education level, income level, 2004 presidential vote, and time-to-decide: when a voter settled on her presidential candidate.

At each precinct, we used systematic sampling: we surveyed every fifth voter to leave the poll. Some voters declined to participate; we estimate roughly 90% of polled voters did respond. We chose immediate replacement of non-responders, meaning that if one voter refused to take our survey, we polled the next voter leaving the precinct.

One issue inherent in every exit poll is clustering: a county's election board subdivides its voters into precincts. To eliminate all bias, a poll would ideally sample randomly within all precincts, or a poll might select a significant number of precincts at random, and sample randomly within each one. However, the large quantity of precincts (over 170,000 in the country, 102 in Forsyth Co.) makes this infeasible.

With merely 18 students in the seminar, we had manpower to poll at two precincts. Using only two precincts introduces potential bias (and is responsible for our one incorrect prediction), but it was necessary based on our logistics. With more students, we could use a more sophisticated design to ensure better results. This led to our main statistical compromise: we did not select our precincts completely at random; we wanted to avoid precincts which were heavily skewed by party affiliation (Forsyth County has several of these) or scarcely populated. The precincts we surveyed were at a nearby high school and at a recreation center located further away from campus. Both precincts contained slightly fewer Democrats and slightly more Republicans than the county average. This compromise meant that our results would certainly apply on a precinct level but introduced more error at a county-level.

Each student (and the professor) interviewed some voters during the day. Via class discussion, we had agreed on our procedure and an outline of what would be said; each interviewer was to identify herself as being from a seminar at Wake Forest and then ask the voter to complete our anonymous survey.

4. OUR PREDICTIONS VS. THE RESULTS

Our exit poll interviewed 250 voters¹. We estimate our response rate to be 90%. Students began tallying the results at 3pm on Election Day; polls in North Carolina closed at 7pm, which means that anyone in line at that time is allowed to vote. We first posted our results online around 7:45pm.

A summary of our results appears in Tables 1-4 in this section; due to votes for third-party candidates, the totals do not sum to 100%. We correctly predicted the results of all contests, except for one county commissioner. Actual results are reproduced from the county website [1]. For the president, senate, and governor races, our results were all within a couple percentage points of the actual results. If we naïvely estimate the standard error as $\frac{1}{2\sqrt{n}}$, then our survey has a 95% confidence margin of error of 6.20%. Because of our compromise in selecting two precincts, the actual countywide margin of error is greater.

TABLE 1. Election for **President**

Candidate	Our Predictions	Forsyth Co. Results
Barack Obama (D)	52.00% ✓	54.83% ✓
John McCain (R)	45.20%	44.35%

TABLE 2. Election for **U.S. Senate**

Candidate	Our Predictions	Forsyth Co. Results
Kay Hagan (D)	53.68% ✓	56.86% ✓
Elizabeth Dole (R)	42.42%	40.37%

TABLE 3. Election for **Governor**

Candidate	Our Predictions	Forsyth Co. Results
Bev Perdue (D)	50.44% ✓	54.94% ✓
Pat McCrory (R)	43.86%	42.17%

We consistently underestimated each Democratic candidate's votes (by 2.83%, 3.18% and 4.50%) and overestimated the Republican candidates' votes (by 0.85%, 2.05%, and 1.69%).

These discrepancies can be understood by recalling that we surveyed in precincts comprised of slightly fewer Democrats and slightly more Republicans than the county average. In addition, North Carolina offers early voting in the weeks before an election. We were unable to survey these

¹It is mere happenstance that this is a round number.

voters – media interviews suggested that the early vote might trend Democratic. Also, African-Americans were underrepresented in our survey, which might further bias our projections.

The county commissioner race had 7 candidates vying for 3 seats; see Table 4. We correctly predicted the top two Republicans would win; however we showed that a Democrat, Nancy Young, would claim the third seat. Instead, she lost decisively, though she finished well ahead of the other three candidates. Were our results biased for Young? Probably. Following the election, we learned of a simple explanation for Young’s appeal to our responders: she lives near the high school where we polled, and in this precinct her ‘hometown’ status helped her to finish second, a mere two votes behind Plyler. This precinct’s results closely match our exit poll data. With a better, more complicated poll design, we might have avoided this error.

TABLE 4. Election for **County Commissioners** (vote for 3)

Candidate	Our Predictions	Forsyth Co. Results	HS Precinct
Plyler (R)	17.28% ✓	19.92% ✓	17.38%
Whisenhunt (R)	17.46% ✓	19.01% ✓	17.24%
Linville (R)	15.81%	17.98% ✓	14.91%
Young (D)	17.28% ✓	14.57%	17.35%
Ziglar (D)	15.07%	13.42%	15.76%
Gladman (D)	13.97%	12.25%	14.66%
Norman (L)	3.13%	2.83%	2.70%

Human error must be taken into consideration when considering any exit poll. All of our data was compiled by hand, then entered on spreadsheets. As quality control, we had two separate tallies on election night, followed by a third afterwards. Human error by the polled is also likely with such an operation; many factors (e.g., haste, unfamiliarity with candidates, ...) might cause a respondent to unintentionally provide mistaken information. Also, there exist respondents who purposely falsify any survey they complete. Despite these sources of error, our results provided accurate predictions at the county level.

5. DEMOGRAPHICS

In our county (Forsyth Co., NC), Obama defeated McCain by receiving 55% of the vote to Senator McCain’s 44%. According to our exit poll, President Obama won several key demographics which enabled him to win by such a substantial margin. Over 92% of voters who voted for John Kerry in the 2004 election continued their support of the Democratic Party by voting for Obama. This continued support was important for Obama as he was able to retain the base of the party. Conversely, McCain lost the support of 17% of the voters who cast their votes for George W. Bush in the 2004

election, which was one reason why Obama was able to turn many states ‘blue’, including North Carolina. Table 5 reports the presidential results broken down via five demographic groups.

TABLE 5. Demographics: Our Presidential Data

Demographic group		Obama	McCain
<i>Race</i>	White	41.05%	55.26%
	African-American	100.00%	0%
<i>Gender</i>	Female	57.03%	41.41%
	Male	49.55%	45.95%
<i>Race+Gender</i>	White Female	43.30%	54.64%
	White Male	33.93%	55.06%
<i>Age</i>	18-35	68.24%	30.59%
	36-59	45.80%	49.62%
	60+	39.29%	60.71%
<i>Income</i>	under \$40K	77.78%	20.37%
	\$40-99K	52.69%	45.16%
	\$100-149K	45.65%	50.00%
	over \$150K	25.00%	68.75%

We were surprised at the scope of the contrast along racial lines: our 41 African-American respondents unanimously voted for Obama, while among white voters he lost by over 14%.

The demographic data we collected reinforces many widely reported factors in the 2008 race. Obama enjoyed strong support from both younger voters and poorer voters, while John McCain was more popular with middle-aged voters and elderly voters. Was age or income more important? Our data suggests income: Obama’s strongest category of support (aside from race) was among voters earning less than \$40,000. Almost 80% of these voters chose Obama, and they formed roughly a quarter of our electorate. Meanwhile, McCain’s strongest support came from wealthy voters; our wealthiest group favored him by over 40 points. The correlation between income and party vote was stronger than we expected.

We also examined the demographics of the senatorial and gubernatorial races of the state of North Carolina. The same trends appeared in these two races as in the presidential race: all produced Democratic winners who received large support from young voters, lower-income voters, female voters, independent voters, as well as voters from minority groups (though the minority vote was no longer unanimous).

6. CONCLUSION: PLEASE TRY THIS YOURSELF

We encourage other courses and groups to conduct exit polls as a valuable learning opportunity. The seminar students had a natural enthusiasm for this election, both because as 18-year-olds it was their first chance to vote in an election and because of the historic context of Obama's campaign. Simply by being at a precinct for two hours, they observed democracy in action; they appreciated the local news coverage that our survey drew. They gained ownership of a topical project, which provided an excellent setting to learn about polls and statistics.

For several students, this seminar served as their first (and perhaps, only) introduction to statistics. Such a project provides immediate motivation to understand the ideas of confidence intervals and margins of error. While our survey method was too simple, it provided a launching point to discuss how it could be freed from bias, given more resources. Depending upon the statistical sophistication of the students, this discussion could naturally lead into an understanding of several complicated sampling techniques.

Even with a small number of students and a statistical survey design that was too simple, we were able to obtain good predictions about the election results; the one incorrect prediction provided a wonderful teaching moment regarding bias. The actual costs involved were minimal (copies were our biggest expense), but this effort requires a significant investment of time, both inside and outside the classroom.

ACKNOWLEDGEMENTS

This project would not have succeeded without the determined efforts of all 18 students in the seminar; see our website for a listing. Chuck Wessell clustered the Forsyth County election data to look for trends,

We dedicate this article to our fellow seminar student Mike Corrigan, who passed away in August 2010. Mike was a compassionate, articulate, and engaged student who played a significant role in this project. He is greatly missed.

REFERENCES

- [1] Forsyth county board of elections website. http://www.forsyth.cc/Elections/election_results.aspx, December 2010.
- [2] Scott D. Grimshaw, Howard B. Christensen, David B. Magleby, and Kelly D. Patterson. Twenty years of the Utah Colleges exit poll: learning by doing. *Chance*, 17(2):32–38, 2004.
- [3] Susan Kyle, Douglas A. Samuelson, Fritz Scheuren, Nicole Vicinanza, Scott Dingman, and Warren Mitofsky. Explaining discrepancies between official votes and exit polls in the 2004 presidential election. *Chance*, 20(2):36–43, 2007.

APPENDIX: PHOTOS

All photos copyright Jason Parsley, 2008.



FIGURE 1. Students Annalissa Johnson (center) and Hope Mirts (right) survey a voter



FIGURE 2. Students Colin Royal (center) and Ashley D'Auria (right) survey a voter early on election day



FIGURE 3. Students Ashley Gore and Robert Reinhard start tallying our surveys at 3pm on election day



FIGURE 4. Students Ashley Gore and Robert Reinhard start tallying our surveys at 3pm on election day

ABOUT THE AUTHORS:

Annalissa Johnson is a junior at Wake Forest University double majoring in mathematical economics and Latin. She began research on power indices with Jason Parsley after the exit poll sparked her interest in the mathematics of voting. Annalissa's passion for math and logic complements her passion for the classics; she recently completed a three-month stay in Italy where she conducted research on Augustus Caesar's house on the Palatine Hill.

Andrew McCarty is a junior at Wake Forest University majoring in mathematical economics and has continued to work on voting related projects. Andrew is a member of the Wake Forest debate team; he and his partner are currently ranked among the top 10 pairs nationally in debate.

Jason Parsley has twice taught a first-year seminar on the mathematics of voting; the second version included this exit poll project. This seminar also led him to research power indices, along with the first author. He is the recipient of the 2010 MAA-Southeastern Section Award for Distinguished Teaching by a Beginning College or University Mathematics Faculty Member. He recently

completed his ninth cross-country road trip and is well over halfway towards his goal of visiting every national park in the US.

Robert Reinhard is a junior at Wake Forest University majoring in business. Among his other virtues, Robert volunteered for the earliest shift (before 6:30 am) at one polling location. On election night, he was featured on the local news explaining our project (see Figure 3).

Department of Mathematics
Wake Forest University
Winston-Salem, NC 27109-7388