David M. Lawrence
MASC 611 Research Proposal

THE PRESS AND THE POLLS: ACCURACY IN THE REPORTING OF OPINION POLL RESULTS IN THE RUN-UP TO THE 2012 REPUBLICAN PRIMARIES

One of the staples of the never-ending political reporting season is the public-opinion poll story: Who's up? Who's down? Who cares? The need to feed the gaping maw of the 24-hour news beast almost guarantees significant coverage of the most insignificant results. The need to watch the watchdogs has likewise spawned a cottage academic industry—analyzing the accuracy of press reporting of poll results.

Polls have a venerable history as a source for election reportage, dating back to the 1820s; the first syndicated polls, such as the Gallup, *Fortune* (Roper) and Crossley polls, appeared in the mid-1930s. (Crespi, 1980). All three correctly predicted Franklin Delano Roosevelt's presidential election victory of Alf Landon in 1936, cemented the use of polls in generating raw material for news coverage (Crespi, 1980). The importance of polls in generating fodder for news coverage has grown greatly since the Depression Era, and it is no mystery why, for they produce "visible manifestations of political developments that can be readily reported," (Atkin & Gaudino, 1984). Atkin and Gaudino continued:

> Such information is generally concrete, objective, and a product of a well-defined and organized system of record. Opinion surveys usually deal with politically relevant subjects such as candidate evaluations or feelings about public issues, and are therefore consistent with the high news value accorded political subjects. Polls provide explicit representations of public opinion, especially during vast stretches of time between elections.

Crespi, in documenting the coevolution between journalism and public-opinion polling, found several flaws in the polling process that are a result of that symbiosis. The primary flaw, Crespi (1980) wrote, is that "There is a preoccupation with reporting numbers, the 'objective' poll results, with a corresponding lack of interest in their underlying meaning or patterning." This

disinterest leads to one of the most common errors made in reporting poll results: treating the final numbers as "facts" whether or not such claims of "fact" are statistically justified.

The likelihood of such errors increases given the changing nature of election polling in particular. The presidential polls of the 1930s focused on predicting the eventual winner (Crespi, 1980). In time, pre-election polling inspired "horse-race" coverage that focused on magnitude of and changes in leads rather than on the substance and qualifications of candidates (Atkin & Guadino, 1984). This kind of coverage affects the races themselves. For example, polling rather than primary results catapulted Rep. John Anderson (R-Ill.) into contender status as an Independent candidate in the 1980 presidential election (Gollin, 1980). Ronald Reagan earned the Republican nomination and, ultimately, the White House. Anderson, through the media coverage inspired by his polling results, got 7 percent of the votes[1].

Sloppy horse-race reporting can create a potentially dangerous false narrative by creating the appearance of "facts" that are not true. These false "facts" arise primarily through sins of omission in terms of inadequate discussion of survey statistical errors such as the margin of error (Paletz et al., 1980; Rhee, 1996). Polls are based on a sample of a population. Even when the sample is selected perfectly (i.e., at random); the survey instrument (questionnaire) is perfect; the survey was implemented perfectly; and the data recorded, analyzed, and the results reported perfectly there is a chance that the sample value will differ from the actual population value it is supposed to measure.

The margin of error—usually given as a plus/minus value—indicates the range of values the actual population value should fall into a given percentage of time. For example, a sample measurement of 25 with a 95 percent margin of error of ±3 means the actual population value should fall between 22 and 28 95 percent of the time. The margin of error is to pollsters what the

[1] Including mine in my first vote in a presidential election.

confidence interval (Sokal & Rohlf, 1981) is to a biologist. For the difference between two values to be statistically significant, the margin of error (confidence intervals) around each value must not overlap (Sokal & Rohlf, 1981).

Journalists have a history of committing sins of omission in reporting the margin of error. Paletz et al. (1980), in a study of poll reportage in *The New York Times* and on CBS and NBC evening news coverage from 1973, 1974, and 1975 found that the Times was more thorough than the networks. It reported sampling error (margin of error) a whopping 7 percent of the time. The network results were too paltry to mention. Continued failure to report margin of error helped fan confusion over whether the incumbent, George W. Bush, or the challenger, John Kerry were leading in the 2004 election. In the closing weeks of the campaign, some polls suggested Kerry was in the lead, but statistically the differences between the two candidates were within the margin of error (Patterson, 2005).

Given the fact that these well-recognized problems persist into the twenty-first century (Larson, 2003), and that we have what appears to be developing as a hotly contested 2012 campaign, it seems worthwhile to examine whether our journalistic brethren are properly doing their jobs. We propose a content analysis of coverage of the race for the Republican nomination to investigate the following two research questions:

RQ1) Do news organizations report accurately on political polls?

RQ2) Are there differences among news organizations in the accuracy of their reporting on political polls.

Methods

In order to address these questions, I propose to conduct a content analysis (Neuendorf, 2002) of polling coverage from the run-up to the 2012 Republican presidential primaries. We will look

at news reports about 60 polls from major polling organizations in a six-month period from May 2011 to November 2011 and evaluate the content of those reports for accuracy using criteria that will be described below.

Selection of polls

The 60 polls will be selected from the Web site PollingReport.com. The site offers an abstract of many public opinion polls by major categories, including elections: in this case White House 2012. The White House 2012 group is further broken down into Democratic and Republican nominations, which—at least in the Republican nomination case—date back to October 2008. In addition to offering a synopsis of poll results, the site lists the following: dates the polls were conducted, sample or subsample size, margin of error, and usually a quote from the published report or press release.

Selection of sources

News reports of the polls will be obtained from either the Lexis/Nexis or Factiva databases. Researchers will be divided into four teams that focus on a different type of news source: newspapers, television networks (broadcast transcripts if available, if not then reports on the networks' Web sites), wire services, and online portals (i.e., blogs and online-only publications). These categories represent traditional news media (newspapers and television), major providers of content to news outlets (wire services), and the so-called new media (online portals) that are becoming more and more prominent as news providers. Three sources in each category will be evaluated.

The newspapers selected are *The New York Times*, *The Wall Street Journal*, and *Washington Post*. The first two aspire to be and are still considered national newspapers. The *Washington Post*, while no longer claiming national newspaper status, is still regarded, as are the other two,

as one of the top publications in the nation. (And given the address of the White House, the effort to find its next occupant still counts as a local story for the *Post*.) Other national newspapers, such as *The Los Angeles Times* and *USA Today* were considered, but lack of access to archives or insufficient material in preliminary searches eliminated them from consideration.

The three television networks chosen are CNN, FOX News, and MSNBC. They are the most prominent television news providers in the nation today as people have switched from traditional broadcast news programs, such as CBS Evening News, to 24-hour news channels. The three networks likewise represent something of the political spectrum of the country, with Fox allegedly biased toward the right, MSNBC allegedly biased toward the left, and CNN struggling somewhere in the center. In addition to the availability of program transcripts, all three networks have a significant news presence on the Web.

The three wire services selected are the Associated Press, Thomson/Reuters, and Agence France-Presse. These three represent the three largest news wire services in the world. While based in France, Agence France-Presse has a significant English-language operation[2].

Given the proliferation of online news sources, the choice of online portals is far from easy. Nevertheless, the three chosen—*Huffington Post*, *Politico*, and *Salon*—are justified given their prominence as oft-quoted or oft-shared sources. *Huffington Post* is problematic in that it also aggregates content from other sites, but preference will be given to original content from *Huffington Post* contributors.

In all cases for all sources, preference will be give to original—not aggregated or republished—content, despite the fact that the coding instrument will enable us to document whether or not the report is staff- or contributor-written or comes is republished or obtained from some other source.

[2] In full disclosure, I have freelanced for Agence France-Presse in the past.

Coding of content

The unit of analysis is the individual news story. With the exception of stories selected for content analysis (see below), each of the four team (newspaper, television, etc.) members will focus on stories from a particular source and code the stories according to the protocol given in Table 1. Coding will include basic identifying information: coder, case, media category, news source, poll, article month, and type of content (e.g., staff report, wire service, etc.).

The rest of the variables will provide the data needed to address the two research questions. Two questions are specific to the margin of error: Is it reported accurately? and Where is it reported in the text? The margin of error is reported accurately if the reported value matches the margin of error for the sample or subsample of interest (i.e., likely Republican voters). The margin of error is not reported accurately if the reported value does not match the value for the sample or subsample of interest. (This usually occurs when the reported value applies to the entire survey sample, but the results of interest are from a subset of that sample.) The coding instrument includes two other choices, one for cases where the accuracy of the reported value cannot be determined, and another for when it is not reported at all. There are seven choices for the location of the reported value: headline, sub-head, lead (defined as the first paragraph), body of the text, end (defined as the last paragraph), other (such as when the margin of error is provided in a graphic), and not reported.

Two questions are specific to the sample size. As for the margin of error, the questions are: Is it reported accurately? and Where is it reported? The decision rationale and coding choices for sample size are identical to those for margin of error.

The final question goes to the heart of our study: Is the overall reporting on the horserace among the top candidates inaccurate or misleading in regards to the margin of error (radius)? The choices are simple: yes, no, or cannot tell.

Table 1. The proposed coding instrument.

V1. Coder 1-3 (see Blackboard) V2. Case 1-18 V3. Media category 1-newspaper 2-TV 3-wire service 4-online portal V4. News outlet 1-WSJ 2-NYT 3-Washington Post 4-CNN 5-Fox News 6-MSNBC 7-AP 8-Reuters 9-AFP 10-HuffPost 11-Politico 12-Salon V5. Poll # (also name article file with #) 1-60 V6. Article Month 1-12 V7. Story origin 1-Staff report (non-blog) 2-Staff report (blog) 3-wire service 4-other 5-cannot tell	V8. Margin of error reported accurately? 1-Yes 2-No 3-Cannot tell 4-Not reported V9. Where is margin of error reported in text? 1-Headline 2-Sub-header 3-Lead (first paragraph) 4-Body of text 5-End (last paragraph) 6-Other 7-No reported V10. Sample size reported accurately? 1-Yes 2-No 3-Cannot tell 4-Not reported V11. Where is sample size reported in text? 1-Headline 2-Sub-header 3-Lead (first paragraph) 4-Body of text 5-End (last paragraph) 6-Other 7-Not reported V12. Is the overall reporting on the horserace among the top candidates inaccurate or misleading in regards to the margin of error (radius)? (If the % difference between the top two candidates multiplied is equal or smaller than double the margin of error, no candidate should be reported to be in the lead. A trend based on previous polling can be reported and can be coded as accurate.) 1-yes 2-no 3-cannot tell

Despite the simplicity of choices, this question has the potential to be the hardest one to answer. In order to simplify the process, a conservative statistical approach will be implemented.

If the margin of error about two candidates' poll numbers overlap, then one cannot be judged to have a lead over the other. For example, if Newt Gingrich sits at the top of a poll with a 25 percent result, and Herman Cain sits at the bottom with a 16 percent result, but the margin of error for the relevant subsample is plus or minus 5 percentage points, then Gingrich statistically cannot be said to lead the field, for his "true" number may range anywhere from 30 to 20 while Cain's may range from 21 to 11. In this case, it is statistically possible for Cain to actually be ahead. Therefore, the report must clearly address the statistical uncertainty to be judged accurate.

Intercoder reliability

Given the number of coders on each team, intercoder reliability must be assessed. A number of measures are available to choose from (Lombard, Snyder-Duch, & Bracken, 2002). Some of the simpler methods, such as Holsti's method and Scott's *pi*, are designed measure the agreement between two coders. Others, such as Cohen's *kappa* or Krippendorf's *alpha*, are designed for more than two coders. The more coders involved, the more complicated the calculations. Cohen's kappa is appropriate for the number of coders and the type of measurements we need to assess reliability of. An online calculator is available (http://justusrandolph.net/kappa), therefore I recommend it be used. Given the rather explicit decision rationale provided in the coding instrument, likewise recommend a target reliability value of 90 percent. Intercoder reliability tests should be applied to the two margin of error questions, the two sample size questions, and the overall accuracy question.

Analysis of data

Summary statistics (i.e., sum and percentage) will be calculated for each response to the margin of error, sample size, and overall accuracy questions. The percentages can be subjected to non-parametric tests (such as *chi*-square) for differences in frequencies. The number of

comparisons must be taken into account, however, to reduce the likelihood of finding statistically significance differences by chance. Determination of appropriate null hypotheses for the questions may be problematic. One would hope that every media outlet will be more likely to report polling data accurately, nevertheless, in a case like the "Is the margin of error reported accurately" question, the traditional null hypothesis is that there will be no statistically significant difference among the "Yes," "No," "Cannot Tell," and "Not Reported" options. As for tests of differences within and among media types, we have no basis to predict which newspaper, for example, will report polling results more accurately. The null hypothesis in these cases make more sense: There should be no statistically significant differences within or among media types.

Whatever results are obtained from this study, it will be worthwhile to compare the current results to similar research in the past—there appears to be a sufficient number of studies from the past few decades to support such an analysis. If the accuracy of our reporting has improved over the years, great! If the accuracy of our reporting is the same, or (gasp!) worse, we as a profession need to focus on how to do better.

REFERENCES

Atkin, C. K., & Gaudino, J. (1984). The Impact of Polling on the Mass Media. *The Annals of the American Academy of Political and Social Science, 472*(1), 119-128. doi: 10.1177/0002716284472001011

Crespi, I. (1980). Polls as Journalism. *The Public Opinion Quarterly, 44*(4), 462-476.

Gollin, A. E. (1980). Exploring the Liaison Between Polling and the Press. *The Public Opinion Quarterly, 44*(4), 445-461.

Larson, S. G. (2003). Misunderstanding Margin of Error. *The International Journal of Press/Politics, 8*(1), 66-80. doi: 10.1177/1081180x02238785

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research, 28*(4), 587-604. doi: 10.1111/j.1468-2958.2002.tb00826.x

Neuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, Calif.: Sage Publications.

Paletz, D. L., Short, J. Y., Baker, H., Campbell, B. C., Cooper, R. J., & Oeslander, R. M. (1980). Polls in the Media: Content, Credibility, and Consequences. *Public Opinion Quarterly, 44*(4), 495-513. doi: 10.1086/268619

Patterson, T. E. (2005). Of Polls, Mountains. *Public Opinion Quarterly, 69*(5), 716-724. doi: 10.1093/poq/nfi065

Rhee, J. W. (1996). How polls drive campaign coverage: The Gallup/CNN/USA today tracking poll and USA today's coverage of the 1992 presidential campaign. *Political Communication, 13*(2), 213-229.

Sokal, R. R., & Rohlf, F. J. (1981). *Biometry: The Principles and Practices of Statistics in Biological Research* (2nd ed.). San Francisco: W.H. Freeman.