TOWARD AN OPEN-SOURCE METHODOLOGY
WHAT WE CAN LEARN FROM THE BLOGOSPHERE

MARK M. BLUMENTHAL

Abstract During the 2004 election campaign, millions of political enthusiasts downloaded poll data on the Internet, while “Weblogs” provided a new forum for commentary on survey methodology. At the same time, traditional public opinion surveys came under pressure from declining cooperation, contact, and coverage rates, and many automated and Internet surveys began to proliferate. This article provides some examples of “blog” commentary on automated and Internet polls and then explores the lessons to be learned from the spirit of innovation and openness of the Internet in evaluating new survey methods such as automated polls and those conducted over the Internet.

Andy Warhol (updated): “. . . In the future, everyone will be a polling expert for fifteen minutes.”

A user comment posted on the Weblog on DailyKos.com (2004b)

For many in the survey research profession, the future as described by the DailyKos commenter seems to have already arrived. During the 2004 election campaign, millions of political enthusiasts downloaded poll data on the Internet, while the emergence of “Weblogs” fostered an unusual amount of commentary and discussion on polls, much of it partisan and combative. The birth of this medium occurs in an environment where technological and social change put continuing pressure on the conduct of surveys. Yet for all the partisan rancor directed at pollsters in 2004, the spirit of innovation and openness of the Internet may have lessons to teach us about how to evaluate emerging new survey methods such as automated polls and those conducted over the Internet.

This essay has three sections. The first reviews some of the discussion of polling and survey methodology on the Internet during the 2004 campaign, with special emphasis on developments in automated and Internet-based surveys. The second will consider the challenge to the survey research profession
posed by these new methodologies and suggest a response. The final section will offer advice to survey consumers.

**The Internet and Polling in 2004**

In 2004, more Americans than ever turned to the Internet for news and information about politics. As of June 2004, the Pew Research Center reported that 63 percent of Americans were using the Internet and 31 percent went online to look for “news and information about the campaign” in June 2004, up from 16 percent four years prior (Rainie 2005). At the same time, a small but growing subset of those going online turned to Internet-based diaries known as Weblogs. The Pew Center estimated that by November 2004, 9 percent of Internet users said they regularly or sometimes read political blogs during the campaign (Pew/Internet and American Life Project 2005). In February 2005, Gallup reported that 15 percent of Internet users and 12 percent of Americans read political blogs at least a few times a month (Saad 2005).

During the 2004 campaign, talk of polls and survey results seemed every bit as ubiquitous on the Internet as in other forms of news media, as references to poll results appeared on major political blogs on an almost daily basis. While we have no precise estimates of Internet traffic resulting from the public’s interest in polling and survey data, we do have some powerful anecdotes. For example, a little-known website named Electoral-Vote.com posted constantly updated predictions of the Electoral College vote based on the latest publicly available polling data for each state. In October 2004, the site reported roughly 15 million unique visits, nearly as many as the 16.6 million reported the same month by DailyKos, the most popular of the liberal blogs.¹

Another example comes from the explosion of traffic on Election Day to websites and blogs that posted leaked midday exit poll numbers. The Drudge Report reportedly received over 36 million visitors on November 2, “the biggest day in the site’s history” (DrudgeBlog.com 2004; Traugott, Highton, and Brady 2005). Many popular sites—such as DailyKos, TalkingPointsMemo, and Instapundit—crashed or were overwhelmed by the unusually heavy traffic (Dolinar 2004; Dreazen 2004).

Not surprisingly, the intense interest in polling data in the “blogosphere” (the collective term encompassing all blogs) has helped spawn a new type of poll taker, the “entrepreneurial pollster,” best exemplified by Scott Rasmussen. As the *Washington Post* described, Rasmussen conducts polls not for political

---

¹. These traffic statistics come from counter software maintained by the website. In the parlance of the Internet, a “unique visit” counts each time a visitor browses a particular website, viewing one or more pages on that site. A single “visitor” could log multiple “visits,” while a single visit might involve one or more “page views” (or “hits”). See www.sitemeter.com for details. These statistics were obtained from Sitemeter reports for ElectoralVote.com (www.sitemeter.com/default.asp?action=stats&site=s10ElectoralVote&report=36) and DailyKos.com (www.sitemeter.com/default.asp?action=stats&site=sm8dailykos&report=36) accessed on July 24, 2005.
candidates or clients in the news media but “for himself, with publicity and profits as the motive” (Harris 2004, p. A05). Rasmussen conducted daily national and statewide samplings using an automated, recorded voice, a methodology known as interactive voice response (IVR).

He released overall results free on his website, RasmussenReports.com, and offered more detailed cross-tabular reports to paid subscribers. By mid-October 2004, Rasmussen estimated the daily traffic to his site at 1.3 million hits per day (Harris 2004). Another company, SurveyUSA, which conducted IVR surveys for 50 local television news stations during 2004 and made poll results available on its website (surveyusa.com), averaged over 800,000 hits per day during the final week of the campaign.2

Although virtually all of the public pollsters at the national level continued to use live interviewers in 2004, the IVR surveys of Rasmussen and SurveyUSA were a growing presence among state and local surveys. Of 344 public statewide polls released in the final month of the campaign, 59 were conducted by SurveyUSA or Rasmussen (Bloom and Pearson 2005).

Some organizations also experimented with self-administered surveys conducted over the Internet, with mixed success. *The Economist* magazine, in conjunction with the British polling company YouGov, conducted 17 national surveys drawn from an “opt-in” panel of respondents who had volunteered to participate (Economist.com 2004). The *Wall Street Journal* Online edition partnered with Zogby International to conduct polls in 16 “battleground” states using a similar opt-in panel methodology. Zogby and the *Journal* stopped reporting results from Internet surveys in the nine most competitive states after October 19, instead reporting results from telephone surveys cosponsored by Zogby and the Reuters news service. In seven states, they continued to report results from online surveys through November 1 (*Wall Street Journal* 2004). Harris Interactive, the company that pioneered the use of its opt-in panel to conduct political surveys in the 2000 campaign, did not release political surveys during 2004. Finally, the company Knowledge Networks, which maintains an “Internet-enabled panel” recruited with “random digit dialing” (RDD) probability sampling, conducted surveys of debate watchers for CBS News and for Democracy Corps, an organization associated with the Democratic Party.

Bloggers obsessed over polls, and, not surprisingly, some took note of these relatively obscure developments. However, most simply focused on reporting and dissecting the latest polling results, typically demonstrating little awareness of sampling error, much less the finer points of survey practice. Of course, many of these fierce partisans used the Internet and blogs to mount harsh and largely unwarranted partisan attacks on pollsters themselves, ostensibly involving methodological controversies.

---

2. E-mail to the author from Jay Leve, SurveyUSA.
The debate over weighting by party identification provides some pointed examples of what the Washington Post’s Richard Morin (2004) called “smash-mouth attacks” directed at pollsters from all directions during the campaign. From the Republican side, Matthew Dowd, a senior strategist for the Bush–Cheney campaign, sent an e-mail to ABC’s online tip sheet, The Note, attacking a June poll by the Los Angeles Times. Dowd called it “outrageous” and “a mess” because he believed it had sampled too many Democrats and too few Republicans (Halperin et al. 2004b). “Patently fraudulent,” agreed blogger John Hinderaker (2004) of Powerline.com, as did a blog named Deinonychus Antirrhopus (2004), concluding, “The L.A. Times polling group is either run by idiots or are up to some dishonest shenanigans.”

Democratic partisans went into a similarly vituperous attack mode in September when polls looked gloomy for John Kerry. The filmmaker Michael Moore (2004) used his personal blog site to warn that the polls were “wrong” and “all over the map like diarrhea.” A few days later, the Internet-based liberal group MoveOn.org (2004) placed a full-page newspaper advertisement in several major newspapers that directly attacked the Gallup organization. Headlined “Gallup-ing to the Right,” the ad claimed the pollster “refuses to fix a longstanding problem with their likely voter methodology” and implied that Gallup slanted its polls to suit the whims of George Gallup Jr., an evangelical Christian no longer involved in Gallup’s political polling operation. “Take a look,” recommended Donkey Rising blogger Ruy Teixeira (2004), “I think it’s a striking and effective ad.” MyDD.com’s Chris Bowers posted just three words on the subject: “Thank you MoveOn.org” (2004).

These examples merely skim the surface of the name-calling and questioning of motives that typify partisan attacks on pollsters in the blogosphere. However, if the rhetoric that surrounded the party ID debate illustrates the dark side of blog commentary on polling, the more limited commentary on automated and Internet polling may prove more informative and, ultimately, more instructive to those of us in the survey research field.

As early as December 2003, D. J. Drummond, a popular conservative blogger known as Polipundit, noted that automated pollster SurveyUSA had been “scientifically proven to be more accurate” than other pollsters in 2002 and 2003, “call[ing] results that other polls missed.” As proof, he (2003) cited a passage from a paper presented at the 2003 American Association for Public

3. The next day, Los Angeles Times Polling Director Susan Pinkus responded to The Note (Halperin et al. 2004a) with an e-mail noting that their result was consistent with recent surveys by the Gallup and CBS News/New York Times polls. She also released a year’s worth of results for party identification from the L.A. Times poll showing that while the June survey was a bit more Democratic than average, the various party results were “pretty similar to one another (all within the margins of error).”

4. The Gallup “likely voter model” has been the object of more serious critiques even within the pages of this journal (see Erikson, Panagopoulos, and Wlezien 2004). However, the likely voter methodology condemned by MoveOn.org produced a final survey in 2004 that had George W. Bush running ahead of John Kerry by two percentage points, 49 percent to 47 percent. Bush won the national popular vote by a 2.4–percentage point margin, 50.7 percent to 48.3.
Opinion Research (AAPOR) conference (Bloom 2003), referring to it as a “University of Oregon study.” Bloom’s paper, as quoted on Polipundit, had not been nearly as emphatic. It noted only that SurveyUSA had “performed at roughly the same level as other nonpartisan polling organizations in 2002,” though it did “somewhat better” on “most measures” (2003, p. 1).

In March 2004, one of the original political bloggers, Mickey Kaus of Slate.com, took notice of the Rasmussen daily IVR tracking survey. He posted occasional speculation about minor (typically insignificant) day-to-day variation. On March 21, he noted that he had “been getting a lot of email guff” about his reliance on the survey but asked, “Hasn’t it been fairly accurately reflecting the actual trends in the campaign” (2004c)? A month later, he elaborated further:

I’m not saying Rasmussen’s automated survey should be relied on to accurately predict the results of the election (were it held today). But his poll does seem useful for spotting trends and trendlets. He takes the same poll every day, after all. If a candidate goes steadily up (or down) over the course of a few days, that candidate is probably going up (or down), no? (2004b)5

Finally, in May, Kaus received an e-mail from a reader that “cured me of my Rasmussen tracking poll addiction.” The reader modeled the hypothetical day-to-day variation that would result from chance alone assuming a trendless, dead-even race between Kerry and Bush. The simulated results looked much like the variation in the Rasmussen results that Kaus had interpreted as meaningful trends. The Rasmussen survey showed the race as very close in this period, with neither candidate showing a lead beyond the reported margin of error. Kaus conceded that, as a result of the reader’s argument, it would now take “a lead of 4–5 points for two straight weeks [to] convince me that one candidate was actually ahead” (2004d).

A week earlier, Kaus had noted the apparent successes of the SurveyUSA “robo-poll” in a number of Democratic primaries in 2004 and summarized an e-mail I had sent him with a few reasons why IVR might be doing well:

1) They can poll more people more cheaply, creating a larger pool from which to pick only “likely” voters; 2) On some questions, people are less likely to lie to please a machine. In particular, they’ll often tell a live human *of course* they’re going to vote (when they’re not) because they think voting is what’s expected of them; 3) Robo-polling duplicates the impersonal and anonymous nature of voting itself. (2004a)6

---

5. Kaus’s theory about the potential reliability of an IVR study to “spot trends” has some support in the work on “parallel publics” by Page and Shapiro (1992). They examined roughly three thousand subgroup trends on 169 repeated policy questions on public polls and found that “most of the time . . . most groupings of Americans change their policy preferences—when they do so at all—in the same direction and to about the same extent” (1992, p. 294).

6. Kaus identified me as “kausfile’s Mystery Pollster,” the name I would ultimately use for my own Weblog, MysteryPollster.com, launched in September.
For those not familiar with IVR polling, let us consider some of the potential pitfalls that other pollsters have raised with this methodology. The political IVR surveys also use recorded messages rather than live interviewers to solicit respondents, and as such, detractors argue that their response rates must be significantly lower than surveys that use live interviewers (thus contributing to a potentially higher response bias [see Quigley 2003; Sabin 2004]). Without a live interviewer, random selection of a respondent within the household is problematic, and as such, IVR pollsters typically interview whomever answers the phone, using weighting schemes to correct demographic bias. Detractors argue that as a result of either nonrandom within-household selection or response bias, IVR pollsters weight much more severely than other pollsters, thus increasing the standard error. Without a live interviewer, the IVR pollster has no way to check who is answering for the household. A 10-year-old boy may claim to be a likely voter. Without a live interviewer, respondents cannot ask for clarification or repetition. Even those who conduct IVR polling typically concede that without a live interviewer, open-ended questions are too time consuming and impractical (see Quigley 2003; Sabin 2004).

Nonetheless, after the November elections, several websites and blogs took note of the apparent success of SurveyUSA and Rasmussen in polling statewide races in the general elections. The most widely cited was a postelection review by Slate.com:

Before the election, we publicly doubted and privately derided Rasmussen and SurveyUSA, which used recorded voices to read their poll questions. We rolled our eyes when they touted the virtues of uniformity and when they complained that live interviewers “may not know how to read or speak the English language,” could “chew gum,” or might “just make up the answers to questions.” It sounded to us like a rationalization for cutting costs.

Look who’s laughing now. Rasmussen and SurveyUSA beat most of their human competitors in the battleground states, often by large margins. . . .

We won’t settle the relative merits of the two approaches in this article or this election. But when the two major automated pollsters score either second and first—or third and tied for first, depending on how you count it—in round-robin match-ups with the three major human pollsters, it’s time to broaden the experiment in automated polling and compare results to see what’s working and why. Clearly, the automated pollsters are onto something, and the human pollsters who have fallen behind will have to figure out how to beat it—or join it. (Kenner and Saletan 2004)

Other bloggers took note. Polipundit ran his own scorecard, ranking SurveyUSA and Rasmussen as the two most accurate pollsters (Drummond 2004).

By comparison, the methodology of the opt-in Internet-based surveys conducted by YouGov and Zogby received relatively little attention, although bloggers often posted and dissected their results, as they did with other polls. Liberal bloggers were intrigued by the final YouGov poll showing Kerry running
three points ahead of Bush, a better performance for the Democratic nominee than in any of the other public polls.

One post on the popular liberal blog DailyKos is illustrative of the blogosphere’s role in disseminating information to its readers. On Thursday, October 28, Washington Post Polling Director Richard Morin authored a long article detailing the “smash-mouth” attacks on pollsters during the campaign and reviewing the various difficulties now facing all pollsters, including the use of cell phones, caller ID, “soaring” costs, and cooperation rates “at or near record lows” (2004, p. C01). A DailyKos author identified as “Armando” excerpted a passage from Morin’s piece on falling response rates under the headline: “The Death of Polling: WaPo Pollster Admits They All Suck” (DailyKos.com 2004a). Yet despite the headline, in an exchange of comments about the article, Armando and fellow Kos writer “DemFromCT” saw a future for polling:

Armando: What Do You Think? I mean, it’s pretty damning. And it really makes mincemeat of the trackers.
DemFromCT: the trackers will live or die together. that’s for sure. The proof is [the election outcome on] 11/3/04.
Armando: Not Zogby. He’ll apply his “special sauce.” Watch.
DemFromCT: seriously I wonder if YouGov and internet polling is the future.
Armando: It has to be. There really is no other way. Weighting is the future.

Do not let the informality or the anonymous “handles” obscure the importance of the participants. “Armando” and “DemFromCT” are both regular contributors to the DailyKos front page, a platform from which they can reach an average half million visitors a day.7

This exchange is emblematic of much of the blogosphere’s commentary on polling methodology: Some is good; some, bad; some, ugly. It embodies many of the themes common to all of the passages cited above: a fascination and openness to new technologies, an instinct to test the quality of polling by validation against election results, and the enduring paradox of consumers who view polling as badly flawed and untrustworthy and yet remain obsessed with every twitch of the numbers.

The Dilemma for Survey Research and a Practitioner’s Perspective

The commentary of the blogosphere also highlights the various challenges now facing public opinion research. As has been documented exhaustively in this journal and elsewhere, we face rapid and continuing change in the technology of telephone communication and the way Americans use and answer their

7. The exchange quoted did not appear on the DailyKos front page but, rather, in a more obscure and far less trafficked “diary” within the site.
telephones that threatens to undermine the theoretical basis of telephone surveys. What impact these will have over the next 10 years is anyone’s guess, but we know that response rates have plummeted over the last two decades, cell phone–only households are starting to reduce coverage, and costs are rising sharply.

Meanwhile, as the blogosphere commentary demonstrates, the technology available to survey researchers and the demands of the marketplace for innovation are moving faster than our ability to understand its implications. That pressure is made greater by two meta-analyses presented at the 2005 AAPOR conference that generally confirm what Slate’s correspondents observed, that the IVR surveys performed at least as well as or better than other statewide surveys during 2004 in projecting election results (Bloom and Pearson 2005; Shipman and Leve 2005).

So what is the survey research profession to do? Here is one practitioner’s admittedly subjective opinion.

LET A THOUSAND FLOWERS BLOOM

Now more than ever our profession needs to embrace a spirit of innovation tempered by experimentation and evaluation within the context of the Total Survey Error framework (see Groves 1989). Yet all too often we respond to the introduction of a new technology with a reflexive fear. As Mick Couper put it after reviewing responses to advances such as computer-assisted telephone interviewing and computer-assisted personal interviewing, “Each such innovation is viewed as a portend of the end of surveys as we know them, as a threat to all we hold near and dear as survey researchers” (2002, p. 1). But, as Couper also notes in the same paper,

the reality is often neither as wonderful as the proponents of the technologies argue, nor as dire as the major detractors fear. Each new technology enhances and extends the range of possibilities and opportunities for survey research, but also often introduces new challenges and issues for further research. (2002, p. 2)

Consider our reaction to the use of IVR surveys to conduct election polling. Michael Traugott, one of our most prominent and respected academic colleagues, refers to political polls done using IVR as “CRAP—Computerized Response Automated Polling,” arguing that “there is no sound theoretical basis for the way in which these surveys are conducted” (in Quigley 2003; see also Sabin 2004).

But is the issue with respect to IVR election polling one of theory or a question of how to best measure and reduce survey error? Consider the issue of within-household selection. The surveys conducted by SurveyUSA, for example, administer an RDD sample with callback procedures similar to those used by pollsters that use live interviewers. Yet SurveyUSA does not attempt to select a random respondent within each household, choosing instead to interview the person who answers the phone and weighting the results to conform
to Census estimates for gender and age. Does this deviation from true probability sampling fatally undermine the “theoretical basis” for the survey?

We know that real-world samples are always a far cry from perfect probability. Even the most rigorous methodologies encounter some degree of nonresponse or imperfect coverage. Quality is not absolute. The Total Survey Error framework allows trade-offs. Efforts to reduce one type of error may increase another, and the goal of quality must also consider the limitations of budget and time.

The issue of within-household selection is no exception. The “best practices” of our profession include a wide variety of different methods to select a household-level respondent, many of which involve some deviation from perfect probability sampling. Three of the most prominent and respected survey researchers (Gallup, Harris, and the Pew Research Center) employ a “non-probability” method of selection known as the Youngest Male/Oldest Female (YMOF [see Keeter and Fisher 1997–98; Voss, Gelman, and King 1995]). Scores of other survey researchers follow their lead. The justification for YMOF is a trade-off: a true probability method would increase nonresponse, especially within the context of a survey that must be completed within a relatively short time period. But, as Keeter and Fisher note in a comparison of respondent selection methods,

Any divergence from probability methods should be undertaken with caution, if at all. At the same time, . . . if our ultimate goal is the reduction of error, in all its forms, we may sometimes have to make difficult choices. (1997–98, p. 1)

If an IVR study interviews the first person to answer the phone in order to facilitate a technology that reduces measurement error, is that an unacceptable deviation from theory? Or is it an extension of the same principal, a trade-off among sources of potential error made within the constraints of budget and time?

We must also consider the role for validation in our effort to measure and quantify error within this framework. In other contexts, including consideration of methods to select respondents within households, we arrive at our “best practices” with experiments and studies that attempt to measure survey error. These studies are not about theory—they attempt to measure reliability and validity in some way.

Yet unlike our consumers on the Web, we often shun arguments for newer technologies based on crude efforts to compare preelection survey results against election returns. What makes for a good sample, we often say, is not what comes out (the results) but what goes in (probability sampling, rigorous field practices, and standardized questionnaires). That is sound advice, and no one can deny that in any single application, chance alone can determine which preelection survey comes closest to the final result. However, with a meta-analysis involving enough cases, we can at some point overcome the limitations of sampling error and conduct comparisons that add up to “validation.” If we can include and control for other variables that might affect survey accuracy,
such meta-analyses can become powerful tools for assessing the efficacy of newer methods like IVR.

Claims about response rates and their potential to increase or decrease response bias are also in need of more empirical evaluation. Skeptics argue that response rates for IVR surveys are lower than those of surveys completed with live interviewers. Proponents of IVR surveys claim that their response rates are comparable to “industry averages” and, separately, that the reduction in social discomfort may reduce nonresponse bias in the selection of likely voters. How much do we really know about these claims? If response rates are lower, are they low enough to undermine the “theoretical basis” for such surveys?

In 2002, for example, SurveyUSA (2003) released data indicating that their response and cooperation rates were lower than the average values reported in a blind study of five top media pollsters conducted by Jon Krosnick and his colleagues (2003) but still well within the range of rates reported by that study. These claims are difficult to evaluate, of course, without access to full disposition reports and given the limitations of our standards for computing cooperation and nonresponse. An inbound IVR study can identify whether a call reaches a working telephone and whether a human being answers the phone, but it cannot readily distinguish a business from a residence or households that fall outside the sampling frame. To better evaluate the claims and counterclaims regarding response rates, we need to reach consensus on how to apply the AAPOR standards to inbound IVR.

TOWARD AN OPEN-SOURCE METHODOLOGY

One of the buzzwords of the moment in the field of computer programming is the concept of “open-source” software and computing. The term refers to software created by a community of programmers rather than a single vendor. The best example is the popular Web browser known as Mozilla. The “source code” of the software was created by programmers from different organizations and is made freely available to others to modify and copy as they wish. Open-source software and operating systems have proven so popular that many for-profit corporations are adapting open-source products into their business models.

In the world of academic survey research, a form of “open-source” methodology reigns. Survey data and methodology are fully disclosed and shared. While we certainly lack a single unified survey methodology analogous to the Mozilla browser or Unix operating system, academic survey researchers have reached broad consensus around a set of “best practices” on many issues. Moreover, the “source code” of academic survey research is free and available to all practitioners to modify as they see fit.

In the media and commercial world, however, a more proprietary model still applies. The bigger survey research companies conduct their own research and development. Yes, many of the well-known national pollsters deposit raw data with the archives of the Roper Center for Public Opinion
Research, the Inter-University Consortium for Political and Social Research, and the Odum Institute for Research in Social Science (University of North Carolina). While many of these same pollsters will share aspects of their internal research and procedures at AAPOR conferences and for studies that appear in this journal, much methodology remains closely guarded or inaccessible, including details regarding the selection of “likely voters” and specific procedures for weighting data.8

It is worth noting that even among media pollsters, an “open-source” philosophy exists with respect to the dissemination and sharing of question text. Spurred in part by the standards propagated by the National Council on Public Polls (NCPP), most survey organizations routinely release the full text of the questions they ask. As a result, pollsters in the process of designing a new questionnaire need not contact colleagues to peruse the language of previous studies. They can go online to resources like the Polling Report or the I-Poll online databases of the Roper Center for Public Opinion Research to search and compare numerous questions from comparable past studies.9 As with “open-source” software, they can replicate the language used in previous studies or modify it as they see fit. By putting our question language out into the public domain, we collectively strengthen the quality of all opinion research.

Yet, given the competitive pressure, why should we extend this concept to other areas of methodology deemed more proprietary? The reason is the rapid rate of change, as well as our pressing need to adapt. As Mick Couper has observed:

The rate of technological change is likely to outpace our capability to do sufficient research on every new method prior to adoption. We need to strike a balance between a headlong rush to adopt each new technology at once, and waiting for all the research evidence to be accumulate before making the transition. . . .

Research and practice will of necessity go hand in hand. The era of large-scale methodological inquiries are probably behind us. On the other hand, every survey can be seen as an opportunity, not only to produce data of substantive relevance, but also to advance our knowledge of surveys. (2002, p. 19)

Moving toward such an “open-source” collaboration on methodology could take many forms, but it would generally call for more experiments by practitioners

---

8. When Voss, Gellman, and King gathered information from eight prominent national pollsters for a Public Opinion Quarterly article on pre-election survey methodology, they had to rely on what they describe as “highly diverse sources, often under rather chaotic conditions. In no case could we get all details about an organization’s methodology from one person or source. . . .” The rigor with which we were able to verify it varied from organization to organization” (1995, p. 128). In 2004, I attempted to gather information on likely voter models for MysteryPollster.com from 25 public pollsters. All but two responded in some way, and most were generally responsive, but the degree of disclosure varied widely. Some were willing to describe their likely voter selection procedures in great detail, while others relied on more general descriptions that withheld important details. For example, when I asked pollsters to specify the percentage of their adult samples that qualify as “likely voters,” roughly half refused.

9. Other relevant archives include the Inter-University Consortium for Political and Social Research (ICPSR) and the Odum Institute for Research in Social Science.
and more disclosure to facilitate more and better meta-analyses across surveys. Few practitioners regularly conduct controlled experiments—they lack either the budget, the training, or both. Most of us choose to “tinker” in an evolving way. But if one can imagine an environment in which practitioners were willing to collaborate on methodological studies, the possibilities are unlimited.

Further, simple disclosure of more methodological information on studies that are already being released into the public domain could facilitate a far richer and more useful brand of meta-analysis. For example, the existing studies that compare preelection surveys to vote results could be that much stronger if we could include variables for sample frame (RDD or list), whether the pollster weights, and by what (demographics? regional breaks? party ID or other attitudes? weighted and unweighted frequencies for common demographic variables?).

Is this a pipe dream? Perhaps. But the growing popularity of sites that provide access to polling data on the Internet creates an opportunity to set a new standard of disclosure for that medium. The NCPP standards have had tremendous reach far beyond the NCPP membership. Nonmembers and private pollsters routinely include field dates, number of interviews, information about the “margin of error,” and a description of who was interviewed largely because consumers demand it. They demand it because the most prominent public polls provide it. The original NCPP standards make sense in a world of very limited column inches and airtime. But the Internet has few such limits. Public polling organizations now routinely provide long detailed reports on their websites with no such constraints. The opportunity for a new standard is ripe.

ADVICE TO CONSUMERS

In this environment, with technological change outpacing our ability to reach firm conclusions about the efficacy of some technologies, what advice can we offer those who procure and consume survey research? Here, from a practitioner’s perspective, are some cautious suggestions.

Probability Sampling—Trust but Validate. As Keeter and Fisher put it, “Much of what we trust about survey research rests upon the foundation of probability sampling” (1997–98, p. 1). But as Fiorina and Krosnick remind us, “Real world samples have been only approximations of probability samples” (2004, p. 1). We should continue to put our trust in probability sampling, regardless of the mode of data collection. Nonetheless, no survey is infallible. The more we can do to validate our samples against reliable benchmarks, the greater the confidence we can all place in our work.

Opt-in Internet Polls—Experiment and Validate but with Caution. At what point do the compromises to probability sampling become too great? At what point, if ever, might we place greater trust in surveys drawn from opt-in panels? The only way we will know is by continued experimentation, disclosure, and attempts to evaluate the results through the Total Survey Error framework. Opt-in panels are gaining popularity, whether we approve or not. We should
encourage those who procure and consume such research to do so with great caution and to demand full disclosure of methods and results. If nonprobability sampling can ever routinely deliver results empirically proven more valid or reliable, we will need to understand what produces such a result.

**IVR Surveys Deserve Serious Consideration.** The concerns raised by traditional survey researchers regarding the use of inbound IVR to conduct pre-election polls remain unresolved. Healthy skepticism is appropriate, as always, but a reflexive rejection of IVR as “theoretically unsound” seems unwarranted. The sample frames and callback procedures can be identical to those used in conventional surveys. Yes, IVR studies appear to push the envelope with respect to in-house selection and demographic weighting, but they are extending similar compromises already made by conventional surveys. What evidence do we have that respondents obtained through an inbound IVR study are more biased than those obtained on the phone with live interviewers? The vote validation studies we have available show that the IVR studies are consistent with, and possibly superior to, surveys done with live interviewers. Though the jury is still out, they show great promise in some applications and deserve more careful study and consideration.

**There Is No Perfect Poll.** We have always seen uses for nonprobability methods of measuring public opinion. We use focus groups to get a nonprojective, “qualitative” view of those who participate; we use surveys based on probability samples to make “quantitative” estimates about larger populations. We can also see reasons to compromise some aspects of a pure probability design (such as within-household selection) in order to achieve specific goals (such as surveys with short field periods). Automated and Internet surveys have some place within that framework. Again, the jury is still out, but we may find that automated surveys are most appropriate for very short surveys that forecast elections, monitor reactions to breaking news, or track opinion on a few discrete variables, while live interviewers remain more appropriate for longer, more complex studies that require open-ended probing, clarification, or more than a minute or two of respondents’ time.

**“Meta-analyze” the Results.** The logic of the probability sample tells us that we can never be 100 percent certain of the result. The challenges to coverage and response rates should heighten that caution. All polls may not “suck,” but all are susceptible to nonsampling error. Where they can, consumers should examine results from a range of surveys, conducted using different methodologies, as a check against that possibility.

**Demand More Disclosure.** Ultimately, the market is the most powerful force for creating greater disclosure by pollsters, public and private. If customers demand disclosure, pollsters will provide it. Reporters and ordinary consumers need to start asking that surveys disclose more than just survey dates and sample sizes. A pollster hiding something may well have something to hide.
References


