In 1936, George Gallup correctly forecast Franklin Roosevelt’s presidential landslide at a time when one of the country’s well-known straw polls was predicting a comfortable win for Republican Alf Landon. Gallup was one of the first to adopt a statistical development—probability sampling—that soon became the gold standard of American public opinion research. In probability sampling, each member of the sample is selected with known probability (simple random sampling, in which each member of the sample has an equal probability of selection, is the simplest example). According to statistical theory, a probability sample will be representative of the larger population to a calculable degree and thus free of the selection biases that plagued straw polls. In short order, American public opinion researchers adopted probability sampling as the means to the end of representative samples.

Of course, real world samples always have been only approximations of probability samples. No comprehensive lists of people or households in the U.S. exist, so to some extent the larger population is unknown. And not everyone agrees to be interviewed when reached, even in the early days of face-to-face interviewing in respondents’ homes. As telephone polling replaced face-to-face interviewing in the 1970s, random digit dialing enabled samplers to attempt to contact representative samples in telephone-equipped households. During the past decade, however, the challenges of telephone surveying have mounted. Social changes, such as increases in two-worker families, may make it harder to reach members of a sample. Technological changes such as answering machines and caller ID enable potential respondents to avoid attempts to reach them, and the abandonment of land lines in favor of cell phones may further exacerbate this difficulty. Furthermore, negative
reactions to telemarketing may spill over to telephone surveying and lead to automatic hang-ups.

Such developments cumulate into a simple fact about contemporary public opinion surveys that is rarely discussed: the polls reported in the lead stories of the nation’s leading media outlets are usually based on response rates of less than 30 percent. Because much of the target sample is never reached or is not interviewed even if reached, there is a potential for substantial non-response bias. To date, much evidence documents that general public survey samples under-represent people who are low in educational attainment, low in household income, minorities, and men. Statistical weighting is done to correct for these sample composition biases, but there is no way to know for sure whether the respondents never interviewed in these under-represented categories can be effectively replaced by the members of their demographic categories who were interviewed.

A number of studies have explored whether drops in response rates from the 60% or 70% level (typical of a few decades ago) to the 20% level (more common today) have yielded increases in non-response bias in terms of demographic characteristics (observed prior to statistical weighting) and answers to substantive questions. But to date, no such studies have yet found evidence of any increases of notable size (e.g., Curtin et al. 2000; Holbrook et al. 2003; Keeter et al. 2000; Merkle and Edelman 2002; The Pew Research Center for the People and the Press 2004). Furthermore, there are some striking instances in which notably lower response rate surveys were more accurate than much higher response rate competitors (Visser et al. 1996). These sorts of findings suggest that we need not panic about low or dropping response rates. But most such findings are also not yet well understood, and many scholars in the field have the sense that we are just waiting “for the other shoes to drop,” when sizable relations of response rates to substantive results finally appear (see Groves and Couper 1998).

To be sure, though, the +/- x percent margins of error that accompany many widely-publicized survey results are misleading. This is true partly because these margins of error describe only sampling error, whereas we know many other sorts of error are present in survey data, including errors caused by interviewers and respondents when reporting and
recording answers to questions. But in addition, at least in theory, these margins of error underestimate sampling error itself. Such estimates are accurate only if the respondents ultimately interviewed are a random draw from the original sample, and the less than perfect response rates that typify public opinion polls certainly come with non-response bias in terms of demographics. It is not an exaggeration to say that conventional public opinion surveying today begins with probability samples, then loses successive portions of the sample but hopes that in the end, the losses will cancel out or be corrected by statistical weighting using demographics, so the sample that remains will be a reasonable approximation of the population. This may well be true, and no tests of this assumption to date have yet turned up powerful evidence that it is inaccurate with regard to substantive responses. But the margins of error so often reported are certainly lower bound estimates of the likely error in survey results.

It is important to recognize that when we can compare survey data to alternative, sometimes more direct measures of the same phenomena, the accuracy of surveys, even those with relatively low response rates, is sometimes remarkably impressive. For example, the Columbus Dispatch’s mail polls forecasting the outcomes of Ohio elections between 1980 and 1994 predicted candidates’ vote shares with an average error of just 1.6 percentage points (Visser et al. 1996), and the best polls forecasting the outcomes of very visible national elections have generally been just as accurate (Visser et al. 2000). Likewise, surveys measuring crime rates, rates of drunk driving, voter turnout rates, and the like yield over-time trends that match other indicators of these trends (e.g., police report data bases) very closely (e.g., Chang and Krosnick 2004). But at the same time, there are some well-known instances of poll inaccuracy. For example, the National Election Study surveys and the U.S. Census Bureau’s post-election surveys substantially and reliably overestimate voter turnout, election after election (see McDonald and Popkin 2001). Thus, the traditional survey method is very valuable and often spot on, but it is not perfect.

The development of the Internet has stimulated a rethinking of the practices of public opinion polling that have prevailed for two generations. The internet makes it possible to poll
large numbers of people at minimal expense and to use visual and audio instruments difficult or impossible to incorporate into traditional telephone surveys. But how can one get a probability sample of the population when not all American households are wired? The figure for those with internet access currently stands at about 60 percent in the United States. Moreover, even if the entire population were wired, there is no registry of Internet addresses from which to sample (for a discussion of the issues see Couper 2000).

Two general avenues of Internet surveying have developed. One is to emulate the methods of traditional probability sampling: draw the sample, then connect anyone not already wired. This is a complicated and expensive undertaking, but the founders of the firm that was the first to implement it received the 2001 Innovator’s Award from the American Association for Public Opinion Research. Furthermore, a survey conducted using this method was the first to detect that Arnold Schwarzenegger would win the 2003 California Gubernatorial Recall Election, quicker than any of the competing telephone polls (see Brady et al. 2004).

The other approach to Internet surveying entails a reexamination of first principles; namely, the principle that a probability sample is the only means to achieve a representative sample. The advocates of this approach argue that probability sampling may be a sufficient condition, but it is not a necessary condition. Increasingly, researchers have begun to use volunteer samples or samples of convenience: send out invitations or contact potential sample members and recruit from among those who respond. Although the latter practice would seem to be a regression, a return to the straw polls of the pre-1930s, such procedures cannot be dismissed so simply. A great deal more information about the population is now available—its demographic characteristics, interests, and activities—than was available two generations ago. Armed with such information, some analysts believe that recent statistical developments permit statistical correct for the discrepancies between such samples and populations.

Thus far, some results produced by this approach have been surprising to the skeptics. For example, in 2000, none of the leading final pre-election telephone polls were more accurate in forecasting the outcome of the presidential election than one such Internet poll,
and nearly all traditional competitors were notably less accurate (Traugott 2001). However, 
other evidence suggests that volunteer Internet survey samples can be much less 
representative of the nation than are Internet samples generated by probability sampling or 
probability samples interviewed by telephone (Chang and Krosnick 2003). Only as more 
such research is done can we understand how volunteer Internet polling unfolds and what 
determines its accuracy in particular contexts.

The *Economist*, which already uses Internet polling in Britain, is entering this new 
world in the United States. *YouGov*, a British Internet polling firm, will conduct Internet 
surveys of the American electorate from early July to mid-November. Via advertisements on 
Google and Overture, and emails to commercial opt-in mailing lists, about 20,000 voting-age 
Americans are being recruited for the study. Incentives to participate include payment for 
survey completion and chances to win an end-of-study prize. Invited respondents will have 
two days to respond to each poll.

We have provided advice on the design of the study and will conduct analysis of the 
data on an ongoing basis. Some of the data will be published each week in this magazine, and 
we will post more detailed data, some analysis, and occasional commentary on this website. 
Data will be made available to the larger public opinion research community on a regular 
basis.
REFERENCES


