

Survey Practice This Month

Friday, December 19, 2008, 7:50:49 AM | Editor

The December issue starts with a question that was raised last month. David Moore posted an article that described the [mystery of the convergence of polls](#) in the week prior to the presidential election. We asked a number of survey and polling researchers to comment on the issue and many responded. In addition, David summarizes the responses and raises additional concerns. We also have a psephological analysis of the exit poll results from Robert Worcester.

Three articles this month focus on methods to improve surveys. Panel surveys have many challenges in tracking participants. The article by Linda Kimmel and Jon Miller contains useful advice on methods to track panel respondents. Denise Cobb and her colleagues describe their research with race matching for telephone interviews. And, the article by Rita Thissen and others demonstrates the use of technology to improve survey quality.

Coming next month - Survey Practice will publish another article on the viability of the list-assisted 100-series RDD samples.

As always, we welcome your comments on Survey Practice.

The Editors

John Kennedy

Diane O'Rourke

David Moore

Andy Peytchev

Comments: 0

 [Comments](#)

The Experts Reply on the Poll Convergence Mystery

Friday, December 19, 2008, 7:46:49 AM | Editor

Last month, Survey Practice posted an article by David Moore that noted the "[convergence mystery](#)" - the fact that while pre-election polls show wide disparities during the month of October, they tended to converge in their final pre-election predictions. Moore's data, presented in the article, was based on weekly averages. The data are in last month's article.

We asked numerous experts their views of why the polls showed this convergence, which included invitations to virtually all the major media pollsters. None of the latter chose to offer an explanation.

Mark Blumenthal posted [an article](#) on [pollster.com](#) in response to the article. David Moore contributed more information about the [convergence mystery](#) on pollster.com.

Our thanks to the experts who provided their responses to Survey Practice.

Paul J. Lavrakas
Independent Consultant

There are some reasons that the convergence that Moore notes occurred at the end of the 2008 presidential election campaign would be expected to have happened, and none of them suggest any collusion or conspiracy among the pollsters.

1. Screening and weighting pre-election polls is both an art and a science. Responsible pollsters should always be making explicit decisions about whether the approaches they used in the past (including the last poll they conducted) to screen their samples and weight their data are the best approaches for their most recent poll. There is nothing suspicious about pollsters changing their screening techniques or weighting algorithms based on what they think about the appropriateness of their prior approaches. Taking into account what other polls on the same topic are reporting is one useful and appropriate piece of information in deciding how to “tweak” the screening and weighting used in one’s most recent poll.

2. As shown in the table in Moore article, the variance among the polls represented in the table is decreasing each seven-day period through the month of October, with the exception of the 10/21-10/27 time period. Without this one anomaly, the final variance is wholly in line with the convergence that was occurring consistently during the month prior to the election. As voters become more certain of how they will vote, the polls should reflect this with less volatility across different polls and different polling methods.

3. The economic crisis seems likely to have increased the accuracy of pollsters if it made likely voters more certain of how they would vote. (Data exist that can test whether voters were more certain about how they would vote in the 2008 election versus voter certainty in prior elections, especially right before the election, but I do not have access to those data.) The economic crisis also may have driven an increase in early voting, especially for those who chose to vote for Obama, as early voting can have a cathartic effect for anyone concerned about the urgent need for changing the country’s top leadership — that is, one may immediately “feel better” about the disastrous economy once one votes for the future change (i.e., the new president) that one hopes will be able to resolve the crisis.

Michael Traugott
University of Michigan

David Moore has raised some interesting questions in his posting about the convergence of pre-election estimates of the outcome the 2008 election as Election Day approached. And he and Mark Blumenthal have expanded upon them in a subsequent exchange on Pollster.com. The original observation is not new, although the previous explanations offered for the phenomenon may not be any more compelling than the “methodological fixes” that are now proposed. But at least the original explanations have a sounder theoretical basis and merit some attention as plausible alternative hypotheses about why the estimates converge.

Moore begins with an observation about the convergence of pre-election estimates in the last week of the campaign, suggesting that Barack Obama would win by a smaller margin than the estimates from the previous weeks indicated. But he also comments on the fact that the variance in the estimates is reduced in the final week compared to the preceding weeks and asks why.

The prior explanations for convergence center on the forces that have historically (post-World War II) produced an apparent equilibrium in electoral outcomes in the American two-party system. They begin with a concept of the normal vote developed by Phil Converse (1966), one of the co-authors of *The American Voter*. Based upon the standing division between Democrats and Republicans in the 1960's in the United States, he estimated that - taking into account the higher turnout levels of the Republicans and the greater propensity of the Democrats to defect - the outcome of the prototypical campaign (involving a prototypical pairing of candidates and issue agendas) would be 54% Democratic and 46% Republican. In addition to modeling the dynamics of the system at the aggregate level, the normal vote provides the context for explaining how the outcome of any specific election differs from this baseline expectation by looking at the short term deviations from this expected outcome due to the specific pairing of candidates and the issues that defined the campaign, measured as the voters' reaction to both. While the relative weights for defection and turnout have been adjusted slightly in the intervening period, the basic theoretical underpinnings of the model remain valuable to this day.

As an extension of this model, Gelman and King (1993) wrote a well-reasoned and provocative essay that asked why we pay any attention to polling at all during the campaign when an historical review of post-World War II election returns, in conjunction with the normal vote model, suggested what the outcome would be minus the distraction of variations in the polls along the way. They asked the question "How can political scientists reliably predict the outcome of presidential elections months in advance of election day in the face of active campaigns?" They expanded on the explanation for the "typical" outcome, indicating that the variability of the polls during the campaign was a result of citizens starting the campaign generally poorly informed, and then the stability and convergence of preferences comes about because of the campaigns' effect in informing voters through a variety of mechanisms. Their implication is that as citizens learn more about the candidates and what they stand for, their opinions crystallize and become more stable, hence the reduction in variability in the preference distribution. While there are other factors that contribute to the variations in estimates, including house effects (systematic methodological differences between polling firms), sampling error, and question wording differences, this crystallization of support for a candidate is a central explanatory factor.

There have been some "landslide" elections in the post-war period, including Johnson-Goldwater and Reagan-Mondale. But most of the outcomes have fallen very close to the expected outcome estimated through the normal vote. When there have been significant third party candidacies, the outcomes between the leading candidates have often been closer. Of course, the 2000 election demonstrated how close outcomes can be under specific sets of circumstances, and the 2008 senatorial election in Minnesota illustrates this from the most recent campaign. But these outcomes reflect the impact of the short term forces determined by who the candidates are and how the issue agenda is defined.

In this light, the outcome of the 2008 presidential election seems pretty "normal," despite the discussion early in the general election campaign of whether Barack Obama was "underperforming" and then later in the campaign, after the debates, whether he was pulling out to a runaway lead. Even though his final lead was of the expected magnitude according to the normal vote model, it masked a significant shift in a growth in support for the Democrats as a party and for their standard bearer compared to John Kerry's effort in 2004.

Mark Blumenthal adds his own twist to the convergence phenomenon by suggesting that the explanation lies in the adjustments or "fixing" that pollsters who produced earlier outlier estimates make by examining their procedures and modifying them. This may at one level accurately describe how the pollsters respond to having produced an estimate that is quite different than the majority of the others, but at another level it could

be taken to imply devious or nefarious behavior that would suggest to some a fudging of the data the next time an estimate is produced. This is also a potential alternative explanation, but one that would be difficult to confirm empirically for obvious reasons; and I don't share this view of commercial pollsters. A better prospect for explaining convergence would be to update the Gelman and King results from 15 years ago, with more emphasis on media effects as suggested by Moore and Blumenthal when they note that the variance jumps in the short period after debates. This is a function of direct exposure to the event plus the follow up media coverage of them, especially given their important role in the 2008 campaign.

Converse, Philip E. 1966. "The Concept of a Normal Vote" in Campbell, Converse, and Miller, **Elections and the Political Order**, pp. 9-39 (John Wiley and Sons: New York).

Gelman, Andrew and Gary King. 1993. "Why Are American Presidential Election Campaign Polls so Variable When Votes Are so Predictable?" *British Journal of Political Science*, vol. 4, pp. 409-451.

Micheline Blum
City University of New York

The final three days of national polling before the 2008 Presidential election produced results close to the final election outcome, and which appear to have converged from earlier disparate results. While some earlier volatility in the vote and variation in the polls might be expected over the final weeks, David Moore points out the appearance of convergence in only the last three days and questions the reasons for the accuracy of the later polls and the divergence of the earlier polls.

Some of the reasons for the convergence would be welcome news, while other explanations would not reflect as well on the results. The most reasonable explanations for the "convergence mystery," beyond the solidifying of the vote choice itself in only the final three days, would seem to be:

1. Allocation of the undecided on the final pre-election polls only, with more of the undecided allotted to the candidate previously "underestimated" by the particular organization and in the direction of the average of the polls.
2. Participation of a few influential "outlier" polls and organizations in the weeks leading up to the election but not in the final 3 days.
3. Refinement of, or changes to, the likely voter models, including index questions, weighting formulas and turnout estimates by "outlier" polls.

As I was not privy to the information necessary to examine the last of these, I have chosen to concentrate on reasons 1 and 2 only.

1. Allocation of the undecided vote.

In the last 3 days before the 2008 election 20 national polls conducted by 19 organizations appeared on Pollster.com and were cited by David Moore in Survey Practice. Of these, 7 polls allocated the undecided, (1 poll did not report, but did not allocate the undecided). Allocation allowed the pollsters to apportion more of the undecided to the candidate they had underestimated on previous polls, in comparison with other polling

organizations. Six of the seven reported the vote before allocation on their websites. The undecided vote on these six polls ranged from 2-7 points, with an average of 5 points. Three allotted more to McCain, one gave more to Obama, and two allotted the vote evenly. The 3 polls giving more to McCain showed a larger pre-allocation margin than did the 1 poll which allocated more to Obama. This allocation toward the mean clearly accounted for some of the "convergence" and the reduction in the variance. In the previous 4 days, there were 13 polls done by 11 organizations, and only 1 poll allotted or did not report the undecided. Had some of these polls allocated the undecided in the direction of the average, the variance would have been lower and they would have appeared to converge earlier.

2. The influence of the "outlier" polls.

Much of the variance seen in the polls done before the last three days is due to very few "outlier" polls, and about half of this group either did not release polls in the last three days or were among those which allocated the undecided in their final polls.

The range in the Obama-McCain margin in the final three days was from 5 to 11 points. In the previous four days, 3 of the 13 polls released (by 11 organizations) had margins beyond the 5-11 range and could be considered "outliers." All three had margins of less than 5. No poll had a margin larger than 11 in that four day period. If we remove those 3 "outliers" from the 13 reporting during those four days, the variance is reduced to only 3.3pts.

Similarly, if we look at the previous week (10/21-27), when we could expect more true variance due to voter uncertainty and volatility, only 7 of 29 polls, conducted by 6 of 24 polling organizations, were "outliers." Three of the six outliers were the same three seen above.

Of the six organizations with outlier polls, three reported margins larger than 11 points, and all three consistently showed larger margins in their polls. In the final three days, however, one of these organizations allocated the undecided, giving more to McCain, one did not release a poll, and one had a margin of exactly 11 points on its final poll. Of the three organizations releasing four polls with margins smaller than 5 points, one organization (with 2 "outlier" polls) allocated the undecided in the final three days, one did not release a poll in the final three days, and one "converged." So, of the 6 organizations, 2 organizations (accounting for 3 "outlier" polls) allocated undecided in the final three days in the direction of the previously underestimated candidate, 2 did not release polls in last 3 days, and only 2 "converged." If we remove the 7 "outlier" polls from the 29 released in the week of 10/21-27, the variance is reduced to only 2.9 points.

Moving beyond the last two weeks, we get to the two most extreme outliers seen in the final month of polling, both cited by David in his analysis. These were the only 2 organizations (out of 26 organizations) to report margins of less than 3 points or more than 13 points, and those results were both more than two weeks before the election. Again the pattern seen above is repeated, with one of the two most extreme outliers not releasing a poll in the last three days and the other allocating more of the undecided in the direction of the average of the polls.

Basically then, both of the explanations examined, the allocation of the undecided by seven organizations in the final three days and the absence or favorable allocation of a few "outlier" organizations, appear to be major contributors to the "convergence" seen. Apportioning the undecided in the favorable direction and the absence of previous outliers virtually guarantees less variance and the appearance of "convergence." So,

perhaps, rather than convergence, what we saw was that much of the earlier variance was due to a few outliers and that the final three days benefited from their absence or favorable apportionment of their undecided vote. This may help solve the mystery, but it is both good and bad news in evaluating the 2008 pre-election polls. The bad news is that some of the final poll results and the "convergence" seen and lauded must admit to being assisted by favorable allocation of the undecided vote. The good news is that there were few outlier polls, and, in fact, most of the national polls were very accurate predictors for more than just the final three days.

Cliff Zukin
Rutgers University

The question of why pre-election poll estimates of vote intention converge at the end of a campaign is an interesting one. I suspect the answer may have less to do with our methods than the phenomenon we are measuring. Absent some major occurrence in the campaign, one would expect citizens to become more certain about their behaviors the closer it gets to election day-both in whether and for whom to vote. If so, we might expect less variation in poll estimates (convergence). Screens to identify likely voters probably produce more consistent results from poll to poll, among other judgments that pollsters make at the end of the campaign.

Don Dresser
Sierra Club

Part of the answer is probably just sample size. The obvious part of this comes from some pollsters who deliberately choose to take a bigger sample for the final poll. Harris, ABC, and Pew in particular all ran polls of around 2400 respondents in the last couple of days, but had been doing polls closer to 1200 earlier in the cycle. I suspect there may be an additional piece (which is still basically sample size) that says everyone makes sure they have a poll in the last few days - but pollsters who consistently run large polls may be somewhat less frequent to put out additional results, while those who consistently run smaller polls can afford to do them more often, so the smaller polls may make up a larger fraction of the earlier polling.

I don't think there is enough in sample size to account for all of the convergence you note - but I suspect it tells part of the story.

We welcome your comments on this issue.

Comments: 0

 [Comments](#)

[Explaining the Convergence Mystery: A Tautology Problem?](#)

Friday, December 19, 2008, 7:42:36 AM | Editor

David Moore
University of New Hampshire

The two major types of explanations offered by our experts for the convergence of polls right before the election are based on 1) changes in the pollsters' methodology and 2) changes in the certainty of the vote

choice.

The first explanation suggests that in the final weeks of the campaign, many pollsters adjust their likely voter models (mentioned by Lavrakas and Blum) or they increase their sample sizes (Dresser).

Lavrakas argues that the adjustment of the voter models, even when done explicitly to make their outcomes more consistent with other polls, should be seen as a positive action rather than as a “suspicious” activity. However viewed, such last minute changes could account for some of the convergence.

Dresser mentions the tendency of many pollsters to substantially increase their sample sizes for their final pre-election polls, to insure as small a margin of sampling error as they can reasonably afford. He specifically mentions Pew, Harris and ABC/*Washington Post*.

Blum suggests that the outlier polls during the campaign were either a) less likely to poll in the final three days, or b) more likely to allocate their undecided voters, thus bringing them closer to the mean.

Zukin speculates that the convergence has less to do with pollsters’ methods and more to do with the “phenomenon we are measuring.” Specifically, as voters become more certain about their choices, polls will tend to converge toward each other. That sentiment is also found in the explanations by Lavrakas and Blum. This is the same reasoning that Pew’s Andrew Kohut made on [NPR](#) on Nov. 2, 2008, when he told the host, Andrea Seabrook, that “the closer we get to the election, the more crystallized public opinion is” and thus “it’s pretty typical that the polls - rigorous polls - all come together in the final weeks.”

Mike Traugott’s analysis is different from all the rest, because he focuses on the historical trends and their relation to a “normal” vote. My sense is that this approach, while important for understanding the 2008 election in the context of American elections generally, does not get at the heart of the polling issue being debated here. The question is not so much why Obama’s victory margin was about seven percentage points, but why the polls in October showed wildly varying results during October and then converged to a small variance right before election day. As Mark Blumenthal showed [in his commentary about the “convergence mystery,”](#) the same pattern of convergence was found in the state polls as well as in the national polls - relatively high variance during October, but dramatically lower variance in the final election estimates (even though Obama’s average lead among those polls varied by less than two percentage points).

Why Are Poll Results Related to Solidity of Vote?

Virtually all of the factors offered by our experts seem plausible partial explanations for the convergence. The *implications* of these arguments need to be analyzed further, especially as they relate to the confidence we can put in pre-election polls during the campaign, as well as the confidence we can place in polls that deal with public policy issues when the public does not seem to have crystallized opinions, when sample sizes are much smaller, and where there is no “screening” for likely voters or other indicators of citizen engagement in the issues.

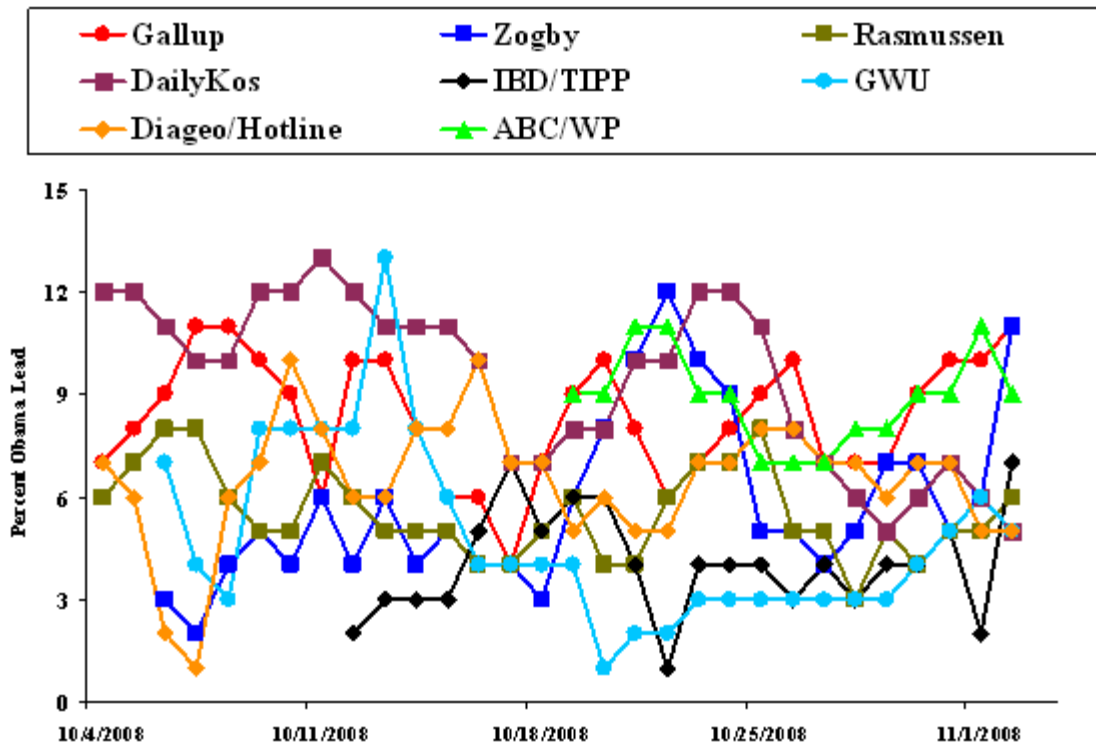
For the time being, however, it may be useful to ask the question - how can we accept the notion that the more “uncertain” public opinion is, the greater the disparity we are likely to find in the polls?

Since posting the original article on the convergence mystery, based on *weekly* poll averages, I posted a further

analysis on pollster.com which shows the results of eight *daily* tracking polls for the last four weeks of the campaign. That graph is reproduced below:

OCT 4 – NOV 2, 2008 OBAMA'S LEAD OVER MCCAIN

Compared by 8 Daily Tracking Polls

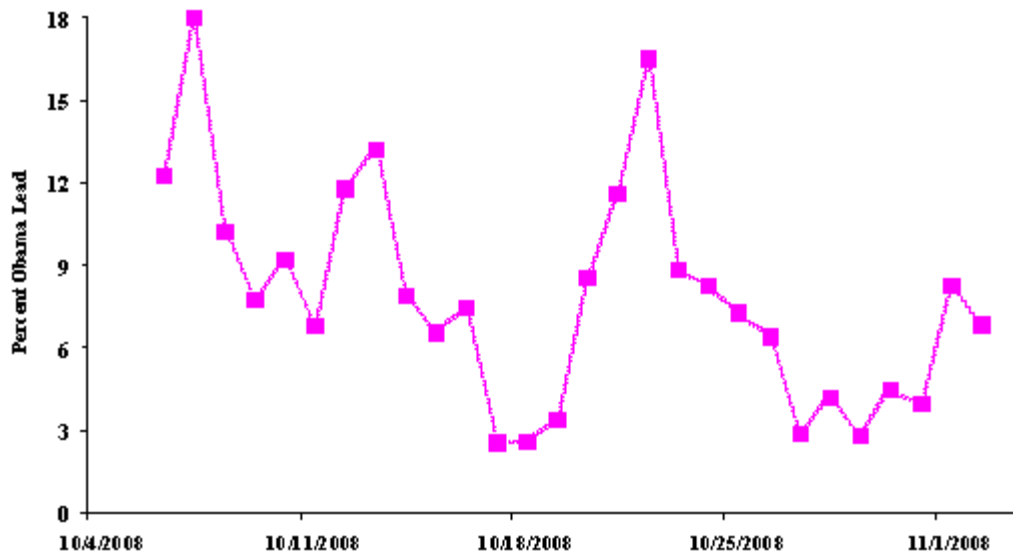


The next graph shows the variance in the polls over this same time period. Note that there is a persistent decline over the month of October, interrupted by three spikes - each one about five days after a vice presidential or presidential debate. Given that the polls are mostly three-day averages, that means the spikes all became evident about two days after the debate - which suggests poll variance probably followed commentary about the debate, rather than the debate itself.

OCT 4 – NOV 2, 2008 VARIANCE IN OBAMA'S LEAD OVER MCCAIN

Compared by 8 *Daily* Tracking Polls

Debates on Oct. 2 (VP), 7, 15



These results tend to confirm the experts' consensus that as election day approaches, (when, presumably, voter certainty becomes greater), the more likely the polls are to converge.

But these results also suggest that the debates had the effect of causing short-term *increased* variance in the polls. Was that because voter uncertainty increased after the debate (and the commentary surrounding the debate)? Is there any evidence for this?

Most important, what is the theory that suggests why polls become more unreliable (inconsistent) just because of voter uncertainty?

The margin of sampling error we calculate has nothing in its formula that relates to the character of what is being measured, only the sample size. Typically, pollsters will add that results can also be affected by nonresponse, question wording, and timing of the interviews - but in the case of the daily tracking polls, none of these additional factors seems to offer a plausible explanation for why results vary among the polls.

All of the daily tracking polls were, of course, conducted "daily," thus eliminating timing as an explanation for their differences. The vote choice questions are almost all identical, unlike, say, questions on a bailout or other policy matters, when question wording is an accepted explanation for variations in poll results. And nonresponse was not suggested by any of the experts as a factor in the polls' variability.

A Tautological or First-Level Explanation for Convergence?

Thus, the issue comes back to crystallized opinions. The experts and the data all seem to agree that the more crystallized are voter choices, the more "accurate" (consistent) are the individual polls (compared with each other). But this seems less an explanation than a different way of expressing an empirical observation. It's at best, I would think, a first-level explanation. We know that polls converge, and we have reason to believe that more people have made up their minds the closer polls are to election day. But the explanation also seems dangerously close to a tautology - the same as saying that polls converge toward election time, because we've observed that polls converge toward election time.

The question is: Why are polls - conducted at the same time, using virtually the same wording - supposedly more accurate (reliable) when opinion is more crystallized?

And if this relationship between poll variance and uncrystallized opinion is significant, do we need to design a new margin of "crystallization" error, to warn poll users of variation due to voter/citizen uncertainty about the issue being measured?

Your comments are welcome.

Comments: 3

 [Comments](#)

A Psephological Analysis of the US Presidential Election from Britain

Friday, December 19, 2008, 7:41:05 AM | Editor

Robert Worcester
MORI and University of Kent

America has elected a black, liberal intellectual. The British psephological tool, "swing," can be used to illuminate the what and the who of the 2008 American Presidential Election. "Swing" measures the votes **state-by-state** (Table 1) which compares the Bush-Kerry election of 2004 to McCain-Obama in 2008 and the **demographic and attitudinal outcome** (Table 2) in 2008 compared to 2004 (using exit poll data). In sum, swing measures the number of voters who have 'swung' from one candidate to another between two outcomes.

The formula is simple: if a = 2008 result and b = 2004 result, swing is $(b+a)/2$, e.g., in 2008 Senator McCain lost to Senator Obama by 46% to 53%, a 7 point difference. In 2004, President Bush beat Senator Kerry by 51% to 48%, a 3 point difference; $(7+3)/2 = 5$, in the aggregate, 5 people in 100 switched in 2008 from the Republican candidate to the Democrat.

[What Table 1 \(Swing by State\) tells us. \(Click for Table 1\)](#)

Across the USA's 50 states (plus the District of Columbia), totals added up to 127,142,278 votes. There were 67,066,915 (52.7%) for Senator Barack Obama; 58,421,377 (45.9%) for Senator John McCain, and 1,653,986

(1.3%) for other candidates. This gave the election to President-elect Obama by 8,645,538 votes (6.8%). This equates to a "swing" of 5%, five voters moving from the Republican candidate, George W. Bush, in 2004, to the Democrat, Barack Obama, in 2008.

The state-by-state vote gives us the best data to compare outcome with hypotheses for any number of measures; those who know each state's record and composition, attitudes and values will be able to use these findings to draw their own, and undoubtedly more sophisticated conclusions, but several stand out.

1. The biggest swing in any state was 18%, in Hawaii. There will be several reasons for this. The 'local boy' factor, voters proud to help elect someone born in their state, was one. There were other factors at work, as politics generally and psephology specifically are complex analysis tasks.

Obama attracted racial minorities in larger numbers than ever before. In many speeches and radio and TV interviews, I supported the idea of a slight Bradley effect (no more than 2% overall) because of the "Reverse Bradley effect" which seems to have impacted Hawaii and other states where a large racial minority exists. Also, where a true Bradley effect does exist, it is largely in already safe Republican ('red') states, mainly in the South.

2. The regional pattern sustains this hypothesis. In the Southern states the swing was less than half (2.3%) of the national average (5.0%). Voters in the Southwest were the next least 'enthusiastic' in their support for Obama (3.5% swing). The greatest swing to Obama was in the mountain states (7.4). Most others were close to the average, the mostly 'blue' (Democratic) states having a 4.5% swing, the central states 5.4%, although Indiana, a central state, had, next to Hawaii, the greatest swing. The 11% magnitude of the swing in Indiana moved Indiana into the Democratic column for the first time since 1964 when the Lyndon Johnson landslide included Indiana.

3. There were only three states which "counter-swung" to McCain - Tennessee (-0.5%), Louisiana (-2.0%) and Arkansas (-5.5%). Arkansas may be a combination of disappointment that Hilary Clinton did not win the nomination and a Southern state tradition. The other 47 states all swung to Obama.

4. I have always carried a 'compressed spring' hypothesis that it is easier to move politically from 50% to 60% than from 85% to 90%. This seemed at first inspection to be sustained by the normally strongest Democratic states having low (further) swings to Obama, such as Massachusetts (0.5%) and District of Columbia (2.5%), but a quick statistical analysis found otherwise.

5. After removing the explicable states of Hawaii, Arkansas and the District of Columbia, the ten most Democratic states had an average swing of 5.4%, the next ten 5.2%, the third ten 5.2%, the fourth 3.9% (which contained Indiana), and the last 5.2%, again. Repeating the exercise by ranking by Republican strength the analysis proved to vary the national 5.0% by as little a spread as between 4.8% and 5.2% across the five segments.

[So what then of Table 2 - the demographic and attitudinal analysis? \(Click for Table 2\)](#)

6. There was little difference in the gender swing. While in 2004 the 3% Bush victory turned into a 7% Obama win in 2008 overall, a "shift" of 10%, among men Bush led by 11 percentage points, 55% to 44%, while women voted for Kerry 51% to 48%, a 14% total, thus a 7% gender swing.

In 2008, Obama won both among men and women, but men only by 1% and women by 13% (which must give the Palin apologists pause for thought). Yet when turned into swing between 2004 and 2008, there was little difference, men 6% and women 5%. Very little discernable Palin effect.

7. Not so with the age analysis. As predicted, without the 18-29 age cohort (18% of voters this time vs. 17% in 2004) and the young's enthusiasm for Obama, the election would have been much closer. The 18-29s swung 12.5% to Obama between 2004 and 2008, with no fewer than two thirds, 66%, voting for Obama vs. 32% who voted for McCain, giving the Democratic candidate a 34% lead in 2008. This was a huge jump up from just 9% more of younger people voting Democratic in 2004.

There was also an above average swing among the next older cohort, 30-44, 6.5%. The next cohort (45-64s) was down to 1%, and a nil swing among the quarter (18% in 2008, 24% in 2004) of voters in the 65+ age group, who gave McCain almost exactly the same percentage of the vote that they gave Bush four years earlier.

8. Nor race, for if only the 77% of voters who are white had voted, Senator McCain would be the President-elect. Of the quarter (23%) of voters who are not, 19 out of 20 blacks voted for Obama and just 4% for McCain. Among Latinos, two-thirds voted for Obama.

Much was made of the voting patterns of Hispanics/Latinos by the media early on. The received wisdom was that they would not vote for a black man in great numbers. In 2004 this was true to some extent, but only to the extent that they were less supportive of Kerry (53%) than were blacks (88%).)

Still, Latinos gave Kerry a 9% lead then, but a massive 36% in 2008. Their 13.5% swing to Obama was among the biggest of any grouping. They now represent 9% of all voters. Asian-Americans are still but 2% of the electorate, and they too swung to Obama, by 7.5%, half again the 5% average across all groups.

9. One of the biggest swings came from a tiny but important segment of the electorate. Just three percent of the electorate in 2004 admitted to the exit poll interviewers that their household income exceeded \$200,000, and then they voted by 63% to 35%, nearly by two to one, for President Bush over Senator Kerry.

Despite the Obama promise that he would introduce tax cuts for 85% of the American taxpayers (and that clearly meant a tax rise for the wealthy), they nonetheless swung by a massive 17% to give Obama a 52% to 46% win in their household income segment (in 2008, their ranks had increased by 100%, from the 3% they amounted to in 2004 to 6% in 2008). This finding destroys the media's false premise that people tell pollsters they will vote Labour/Democrat, but if threatened by tax rises they don't in the voting booth. It was clearly not the case in Britain in the 1997 Blair landslide or in USA in 2008.

10. The other important income group for Obama was the least well off, which includes a disproportionate percentage of black voters, who have household incomes of under \$15,000, and who gave Obama a 10.5% swing. Only among the next income cohort, the \$15,000 to \$30,000 income group, was a below 5.0% swing measured, 4.0%.

11. Those 4% who are the least well schooled, with no high school, again disproportionately black, were strong supporters of Obama, giving the Democrat a 13.5% swing and a majority of 28% this time, compared with just 1% in 2004. Counter-intuitively, those with a college education, even those with post-graduate education,

swung somewhat less to Obama than the average.

Since the days of Adlai Stevenson there has been the hypothesis that less-well educated voters wouldn't vote for an intellectual. Those with no high school with their 13.5% swing, against the less than 4% for those with at least one college degree, showed that other things trumped anti-intellectualism.

12. Party ID is an important factor in American elections. According to the always useful Pew polling, towards the end of the long campaign 37% said they thought of themselves as Democrats while 31% were Republican supporters. In the popular vote, it is much easier to get to 50% +1 from 37% than 31%. That said, the nine in ten of Democrats told the Edison Media Research interviewers that they'd voted for Obama, and nine in ten Republicans supported McCain.

Before the election began, the Democrats had a record ten point lead in party ID. The 28% of those who wear the 'I vote for the person, not the party' badge, the independents (unknown in British political parlance) were running practically even in 2004, with just a one point lead for Kerry; this time it was a 3.5% swing to Obama, 52% to 44%. Obama did well with them, but not nearly as well as the pundits and analysts had anticipated. They represented 26% of voters in 2004, 29% in 2008.

13. America voted for a self-confessed 'liberal.' It wasn't long ago that 'liberal' was a dirty word in American folklore; anyone who was tarred with the 'liberal' brush could not be elected. Obama, a self-confessed liberal just was. The ideology profile of the 2004 American voter was 21% liberal, 45% moderate, and 26% conservative. In 2008 it was 22% / 44% / 34%, not much different.

In 2004, 85% of liberals voted for Kerry, in 2008, 89% for Obama. There the 'compressed spring' theory works, for Obama's swing was 3.5%. Among moderates, who split more evenly last time than this, the swing was 5.5%. Even among Conservatives there was an above average swing of 6.0%.

14. First-time voters in 2004 gave Kerry a 7 point lead; in 2008, they voted 69% to 30%, delivering a massive 16 point swing to Obama. They were 11% of voters in both 2004 and 2008. Another myth turned over.

15. One of the most intriguing findings from the exit poll this year is that while in 2004 those wavering voters who decided finally how they were going to vote in the last three days of the campaign gave a 13 point lead to Kerry, in 2008, they leaned to McCain, giving him a five point lead, 52% to 47%.

16. Protestants (54% of voters both then and now) swung to Obama by an average 5.0%, Catholics (27%) 7.0%, and Jewish (2%) by 4%. The respective leads were McCain by 9%, Obama by 9% and among Jews a massive 78% to 21%, nearly four to one.

17. McCain's choice of Sarah Palin as his vice presidential running mate probably captured more headlines than any other single event in the election, and certainly more discussion. There are no comparative figures from 2004, but the result is certainly interesting if a bit contradictory.

There are two pair of findings: of the 7% of voters who considered Palin's appointment as the most important issue of the election to them, there was a five point Obama lead and of the rest who said it was important (not 'most' important), a third of voters, it was a five point lead to McCain. On the 'minor factor' group, one in five, a massive 2:1 lead for Obama, but an equally huge 2:1 lead for McCain among those who said it was not a

factor, and they were a third of voters.

18. One most interesting finding relates to the 12% of the voters who are members of trade unions, especially because it had been reported that Barack Obama has given the unions a promise both to abolish the Right to Work Act and the secret ballot for officers of trade unions. Did they reward him? Not on your nelly. There was a nil swing from Kerry among union members, while those who do not belong to trade unions swung half again beyond the average, to 7.5%.

Comments: 0

 Comments

The Longitudinal Study of American Youth: Notes on the First 20 Years of Tracking and Data Collection.

Friday, December 19, 2008, 7:39:50 AM | Editor

Linda G. Kimmel, Michigan State University

Jon D. Miller, Michigan State University

There is broad agreement among analysts that longitudinal studies provide a more accurate picture of human change than any combination of cross-sectional studies. Most social science theories predict or imply change in a person or a group of persons over time. Although cohort analysis and other techniques can provide some insights into human change, measurement of the same respondents over time provides the best indicator of human change. In educational research, longitudinal studies have been the gold standard for nearly 50 years.

To provide a more intensive examination of student achievement in science and mathematics (and their relationship to career choices), the National Science Foundation funded the Longitudinal Study of American Youth (LSAY) in 1985 and has continued that support over the last 20 years (NSF grants MDR-8550085, REC96-27669, RED-9909569, REC-0337487, DUE-0525357). This article reports on tracking procedures that were able to account for 94 percent of the original 5,945 students and obtain current addresses for 90 percent of the original sample (see Table 1).

Longitudinal Study of American Youth

The original LSAY participants were recruited as 7th or 10th graders in 1987 from a probability sample of public school districts in the United States. During the first seven years, each student was given mathematics and science achievement tests each fall and was asked to complete attitudinal questionnaires each fall and spring. Parents were interviewed by telephone, and information was collected from mathematics and science teachers and school principals. In 2005, the NSF provided funding for two new cycles of data collection in 2007 and 2008 to allow the construction of an empirical model of educational and career development. Once funded, the first task was to re-locate the original participants, now approximately 32 and 35 years of age.

LSAY Tracking Methods

From the beginning of data collection in 1987, the LSAY staff collected family contact information, which

facilitated short-term tracking when students moved during the in-school years of the study and served as the foundation for subsequent tracking activities.

In April, 2006, LSAY staff began a new tracking activity using: (1) online database tracking, (2) newsletter mailing, (3) calls to parents and other relatives, (4) use of alternative online searching, and (5) questionnaire mailing. As a result of these tracking activities, we located contact information for approximately 90% of the original participants (see Table 1).

Online Tracking. The development of online databases has made tracking far easier than in the early years of the LSAY. During the fall of 2005, we examined several online databases and selected **Intelius** (www.intelius.com), using a premium search feature that verified addresses and phone numbers against utility bills. Customer service representatives will negotiate bulk rates - about 50 cents per search.

Table 1. Results of LSAY Tracking, 2006-2008.

Status of Case	Number	Percent
Original Sample	5,945	100.0
Have participant address, phone, and e-mail	2,591	43.6
Have participant address and phone	1,415	23.8
Have participant address and e-mail	149	2.5
Have participant address	1,137	19.1
Have participant phone	8	.1
Sub-total: Some participant location information	5,300	89.2
Have parent address and phone	7	.1
Have parent address	6	.1
Sub-total: student or parent location information	5,313	89.4
Participant is in military in foreign country	18	.3
Participant is in jail - no address	67	1.1
Participant is deceased	108	1.8
Participant is incapacitated/unable to participate	38	.6
Participant is a foreign student who has returned home	37	.6
Sub-total: student, family, or other disposition	5,581	93.9
No verified information for participant	364	6.1

We used Intelius to search for all participants in Spring, 2006. We located 37 percent of the participants through Intelius, as well as addresses for parents of another 35 percent of the cases (see Table 2). Intelius is inexpensive and easy to use, but it was not useful for finding individuals with common names that had moved from their original city or who changed their names.

In a second cycle of online tracking, we acquired access to **Accurint** (www.accurint.com), a database operated by Lexus-Nexus. Accurint includes access to drivers' licenses in most states and these records include the exact

birthday of each person, which the LSAY collected in the initial cycle in 1987. Because Accurint links records by social security number (blinded from our view), we were able to do nationwide searches including women who had changed their last name. Using Accurint, we located addresses for 75 percent of the participants (see Table 2).

Fall 2006 Participant Newsletter. A fall of 2006 newsletter served as our first contact with the participants since 1994. The newsletter included information about the ways the LSAY data have been used and a few results from earlier data cycles.

United States Postal Service National Change of Address (NCOA). To reduce the number of undeliverable pieces of mail, the U.S. Post Office makes available through licensed firms a computerized data base of changes of address filed throughout the country. Prior to mailing the LSAY newsletter, we sent 5,690 participant addresses to Anchor Computer, a licensed NCOA vendor, (www.anchorcomputer.com), and we received forwarding addresses for 681 of the addresses. The cost of this service was about \$75 and saved us far more than that in postage and labor costs.

Table 2. Results of LSAY Tracking, 2006-2008.

Status of Case	Intelius	Accurint	Current Status
	June, 2006	August, 2006	September, 2008
Located participant	47.2	75.4	89.2
Located parent	34.6	2.5	0.2
Located other family member	1.5	0.1	0.0
No verified information for participant	16.3	0.3	6.1
Participant deceased	0.2	1.1	1.8
Other disposition for participant	0.2	0.2	2.7
Incorrect student address (letter RTS)	–	20.3	–
N =	5,945	5,945	5,945

Parent and other contact calls and letters. For respondents who could not be located through databases, we made follow-up calls to their parents and other relatives. Thirty-six percent of the relatives provided us with participant contact information and another 16 percent said they would forward information to the participant.

Alternative Search Methods. For cases not located otherwise, we used the online Social Security Death Index (www.ssdiregistry.com). If we could not locate the person through the SSDI, we turned to prison records. Federal prisons (www.bop.gov/iloc2/LocateInmate.jsp) and prisons in 37 states have online databases. If a participant has a distinctive name, we search on Google.

Current Tracking Activities. Tracking is a continuous activity in any longitudinal study. LSAY participants move, get married or divorced, change names, and change jobs. We maintain a Web site (www.lsay.org) and encourage our participants to use it to report changes of address and other related information. Although we have e-mail addresses for 75 percent of our respondents, we continue to screen our mailing addresses through NCOA and use Accurint as needed.

The 2007 Questionnaire Cycle

In 2007, we initiated a new cycle of data collection to update our educational and occupational record for each individual. Early in 2007, each LSAY participant for whom we had an address received a letter asking them to participate in the new survey by going to an online site and completing a questionnaire or to return an enclosed postcard indicating that they would prefer a printed questionnaire or a telephone interview. Only two percent of participants asked for a telephone interview, 69 percent used the online questionnaire, and 29 percent completed a printed questionnaire. Each respondent was offered \$20 for their cooperation in the study. After consulting with other studies of this kind, we increased the payment to \$30.

Because the 2007 questionnaire was a report on educational and occupational activities since high school, we continued to write and call respondents throughout 2007. Follow-up e-mail messages were sent when an e-mail address was available. When e-mail was not available, follow-up mailings to non-respondents included a letter asking for cooperation and a printed questionnaire. When telephone numbers were available, follow-up telephone calls were made. By the spring of 2008, approximately 68 percent of the eligible sample completed the new questionnaire (see Table 3).

Discussion

The LSAY tracking activities and the 2007 cycle of data collection demonstrate that it is feasible to locate respondents in longitudinal studies even after a period of more than a decade without contact. We think that this should be encouraging to investigators thinking about launching a new longitudinal study or about the possibility of re-contacting respondents from previous studies. At no time in American history has it been more feasible to track samples of adults over a period of years.

For any longitudinal study, the fundamental requirement is to build an extensive tracking record in the early cycles and to maintain those tracking records through periodic contact. For new longitudinal studies, building a solid tracking information base should be a core objective. Building an initial tracking database requires staff, equipment, and persistence, but there is no substitute for the resulting tracking file.

It is important to encourage respondents to see their participation as contributing to something useful and important. The LSAY emphasized throughout the in-school years that the study was designed to help improve education and that it was only through their cooperation that educational leaders would know what was working and what was broken. When the LSAY data collection was resumed, we adopted a letterhead that reminded participants that they are a part of Generation X and that the LSAY is the only study that chronicles the history of their generation.

Despite the extraordinary tracking and data collection record, there are some unsolved problems. The recent Pew study of prison populations in the U.S. provides estimates on the percentage of young adults (by age groups) who are currently incarcerated (Pew, 2008). The estimated number

Table 3: LSAY Eligible Sample and 2007 Response Rate.

	Number	Percent
Original Sample	5,945	100.0
Ineligible		

Participant in military in foreign country	14	.2
Participant in jail - no address	59	1.0
Participant deceased	108	1.8
Participant incapacitated/unable to participate	38	.6
Foreign student returned to home country	37	.6
Only one or two years of original data	340	5.7
Insufficient tracking information	289	4.9
Eligible Sample	5,060	85.1
Active Sample		
	5,060	100.0
Completions	3,421	67.7
Refusals	156	3.1
Verifying location and seeking cooperation	1,483	29.2

of young adults from our sample who we have identified as being in prison is significantly lower than the Pew estimates. We also note that the number of deceased LSAY participants is significantly lower than the number that would be estimated from reports from the Centers for Disease Control and Prevention (www.cdc.gov/nchs/deaths.htm).

If we were starting the process over ... It is useful to look back at the last 20 years and think about what advice we would give an investigator who is planning to launch a new longitudinal study or is thinking about reviving an earlier study.

First, we would collect the full name - including middle initials - for all LSAY participants and their family members. We would also make certain that we had each participant's legal name rather than a nickname or initials.

Second, we would keep an electronic record of all previous addresses for students, parents, and relatives. In the late 1980's, computer storage was expensive and only the most recent address and phone number were retained in the electronic database, but today memory is cheap and all tracking records should be retained.

Third, we would have used Accurant from the beginning of the study and we would have used an NCOA vendor for all of our mailings. Neither of these services was available in 1987, but we recommend them for any current longitudinal study.

On balance, we think that the LSAY tracking and data collection experiences should be taken as encouraging. And we hope that younger investigators will think about 20-year studies when they

Comments: 0

 Comments

Race-Matching: Interviewers' Reactions to the Race-matching Process.

Friday, December 19, 2008, 7:38:42 AM | Editor

P. Denise Cobb, Southern Illinois University Edwardsville
Cindy Boland-Perez, University of Arkansas at Little Rock
Patricia LeBaron, Research Triangle Institute

Although there has been widespread attention among public opinion scholars to problems of data quality when dealing with racially sensitive items, there is little empirical evidence of how interviewers' understand their work on survey projects that deal with racially sensitive questions and use race-matching protocols. For this exploratory study, we asked thirteen telephone interviewers involved in a racial attitudes survey to help us understand their experiences with that project and particularly with the race-matching process.

Public opinion scholars have established that the social environment matters, affording the characteristics of interviewers and perceptions of respondents weighting equivalent to the content itself (Anderson et al, 1988; Schuman and Converse, 1971; Hatchett and Schuman, 1975; Davis, 1997b). Repeated studies have generated considerable knowledge about how interviewers affect data quality, but little is known about interviewers' understandings of the process and its implications(1).

Project Background and Method

In 2003 the University of Arkansas at Little Rock, Institute of Government, Survey Research Center (SRC) initiated a telephone survey to collect, analyze, and disseminate community data relating to black/white relations. Most of the questions were based upon an instrument used by The Gallup Organization for a series on black/white relations in the U.S. (c1997-2007 The Gallup Organization (2)).

Consistent with calls to mitigate the race-of-interviewer effects, the SRC developed an interviewer/respondent race-matching protocol (Krysan, 2002; Cotter et al, 1982). Each interviewer was trained specifically for questionnaire content and the race-matching process.

We draw on qualitative semi-structured interviews lasting approximately 30 minutes. These thirteen interviewers represented thirty-three percent of the total workforce during the Year 2 survey. Nine respondents were black, nine were female; seven of the interviewers were employed for both the Year 1 and Year 2 surveys. Their ages range from 16 to 70. They completed 635 of the 1694, thirty-seven percent of all surveys in the Year 2 survey.

Findings

Senior staff at SRC conducted extensive training with interviewers to explain project content and the purpose of race-matching (RM). Some interviewers accepted the idea, others were ambivalent, and some were resistant. Acceptance of the practice indicated that interviewers understood RM as an attempt to increase respondent's veracity.

While the interviewers may have heard the messages from training, they didn't accept the premise necessarily. Interviewers were often confident in their skills and believed they should be able to interview anyone - a helpful trait to have among interviewers but one that led to confusion in this situation. Larry's comments illustrate this tension.

"I was told that the purpose was that maybe some black people or some white people would feel more comfortable talking to someone of the same race, but I really didn't see what was the difference."

For some, RM emphasized race and accentuated difference. Making race salient to their work was uncomfortable.

"I didn't agree with it at first . . .because in my eyes everybody is the same." (Shirley)

Regardless of interviewers' receptivity, training messages of how RM would promote comfort resonated. Almost all used the word "comfort" or some derivation.

"You feel more comfortable speaking with people of your race." (Leslie)

While interviewers recognized RM might increase respondents' comfort, they did not always connect it to themselves. Some acknowledged that RM increased their comfort.

"...It would be so uncomfortable for me personally to ask someone of a different race." (Susan)

Our findings also suggest that interviewers may be affected by interviewing someone of a different race when the content is racially sensitive. Some interviewers noted they would have found it uncomfortable to interview respondents of a different race.

"If I was interviewing a white respondent, and they said something bad about my race, I would have been offended." (Eleanor)

Some of these differences of opinion could happen when interviewers are race matched. For at least one interviewer, RM seemed paradoxical to the goals of ameliorating racial divisions.

"...I didn't think it was a big deal that only black people should talk to black people and that only white people should talk to white people if they were willing to conduct a study, it really shouldn't matter if you were black or white..." (Larry)

Regardless, most interviewers eventually embraced the idea because they believed RM increased honesty.

"I think that is the only way you were going to get positive, well, not positive but honest results regardless of what was said or how someone felt..." (Christine)

Nevertheless, interviewers' perceptions of respondent honesty are complex. Some interviewers suggested that respondents were searching for socially desirable responses.

"They were trying to analyze the question and figure out what they think that we expect them to answer, and they were giving us those answers." (Jean)

The commentary between questions may be more revealing than the answers themselves.

"...Always the white women would tell me '(whispers) we would move but we wouldn't tell anybody that

was why' and I'm like that's not an answer...Can you give me what I've asked you? And then they'd give me, 'no, we wouldn't move'..." (Shirley)

This finding reinforces the point that what goes on between questions is critical to data quality. Interviewers must be attentive to the total interaction. In so doing, they may feel frustrated by respondents' contradictions.

Interviewers' comments also suggest that respondents were frank -even if it was off-putting.

"I mean, I'm young, and there were a lot of older people, and they'd try to get me to agree with them. That's what I really didn't like about the answers, because you can express your opinion, but don't say, "You know what I'm saying?" ... "You're one of us." I didn't like that. I may be black, but just because you feel that way doesn't mean I feel that way." (Christine)

Team leaders must help interviewers devise strategies to deal with respondents' needs for affirmation.

Some interviewers suggested that when respondents recognized the match, they were increasingly honest.

"...When I would talk to black women, you could tell they were being distant, even through the phone, just by some of the answers they were giving, but then I think I would say a certain thing or a certain word a certain way, and, oh, she's black, she's a sister, and then they'd be, oh girl, let me tell you...and start talking to me like I'm their friend, and they'd be a little more honest, yes."

Of course, respondents may not have recognized the match at the same points.

There were some unintended consequences for workers. Although there is contradictory evidence, the process may have altered the work environment. Some interviewers thought RM facilitated interaction among groups.

"...Since we had to occasionally stop and ask someone else to do the survey for us, that kind of made us interact more...like, no one hated each other going into it, no one hated each other coming out of it."
(Beth)

The team leaders shared this belief. Some interviewers disagreed and reported new levels of sensitivity.

"I believe some people did get upset about some things that were said. Like, we were sitting black person, white person, and I was, oh, God, don't let this person say anything and then they're going to hear it, and think I agree, and it's not like that." (Christine)

Interviewers expressed that they sometimes felt separated from each other because of RM.

"It was very divided...It was almost like we had the black people and the white people sitting on the other side of the room. Everything was, like, there's this line; you can't cross it. With the call records...With who I could talk to...And I knew that when I got on the phone with a black person, if I could tell by their voice that they were black, okay, well, hold on while I find someone of your race to answer these questions. That really separated caller from respondent, and it separated the respondents from each other, and it separated the workers from each other. It was very, very divided." (Marie)

In spite of deliberate efforts to clarify the value and purpose of RM, interviewers felt some resistance to the process. While there were clearly differences in opinion, what emerged was a sense that respondents' assumptions about the similarity of their positions frustrated interviewers. Their honesty and dishonesty posed challenges. Finally, a significant unanticipated consequence of race-matching was that the process highlighted racial difference among interviewers and may have created divisions.

Accordingly we offer the following tentative recommendations:

Inform respondents of the race match at the beginning of the survey in a standardized fashion.

Recognize interviewing is a social process; improving data quality depends on understanding respondents and interviewers.

Create debriefing opportunities for interviewers to share their experiences and concerns.

Offer mechanisms for interviewers to communicate knowledge they gain during the interview that may speak to the validity of respondents' answers.

Integrate interactive workshops into training to develop greater appreciation of diversity and promote respect among interviewing staff.

Understanding how interviewers perceive their work and how they understand the data collection process is critical to improving data quality.

References

Anderson, Barbara A., Brian D. Silver, and Paul R. Abramson. "The Effects of the Race of the Interviewer on Measures of Electoral Participation by Blacks in SRC National Election Studies." *Public Opinion Quarterly* 53 (1988a): 53-83.

Campbell, Bruce. "Race-of-Interviewer Effects among Southern Adolescents." *Public Opinion Quarterly* 45 (1981): 231-34.

Converse, Jean M., and Howard Schuman. *Conversations at Random: Survey Researchers as Interviewers see it*. New York: Wiley, 1974.

Cotter, Patrick R., Jeffery Cohen, and Philip B. Coulter. "Race-of-Interviewer Effects in Telephone Interviews." *Public Opinion Quarterly* 46 (1982): 278-84.

Davis, Darren W. "The Direction of Race of Interviewer Effects among African Americans: Donning the Black Mask." *American Journal of Political Science* 41 (1997a): 309-22.

Hatchett, Shirley, and Howard Schuman. "White Respondents and Race-of-Interviewer Effects." *Public Opinion Quarterly* 39 (1975-76): 523-28.

Hyman, Herbert H. "Review of *Conversations at Random: Survey Research as Interviewers See It*." *Public*

Opinion Quarterly. 39 (1975): 277-278.

Krysan, Maria. "Data Update to Racial Attitudes in America." (2002): An update and website to complement Racial Attitudes in America: Trends and Interpretations, Revised Edition, by Howard Schuman, Charlotte Steeh, Lawrence Bobo and Maria Krysan, Harvard University Press, 1997.

<http://tigger.uic.edu/~krysan/racialattitudes.htm>.

Rhodes, P. J. 1994. "Race-of-Interviewer Effects: A Brief Comment." Sociology. 28 (1994): 547-558.

Schuman, Howard, and Jean M. Converse. "The Effects of Black and White Interviewers on Black Responses in 1968." Public Opinion Quarterly 35 (1971): 44-68.

1. There are some analyses of interviewers' experiences, many regarding qualitative interviewing and some regarding quantitative surveys (see Rhodes, 1994; Converse and Schuman, 1974 for examples).
2. We gratefully acknowledge The Gallup Organization for allowing UALR to utilize questions from their Minority Relations trend surveys.

Comments: 0

 Comments

The Evolution of Audio Recording in Field Surveys

Friday, December 19, 2008, 7:37:39 AM | Editor

M. Rita Thissen, RTI International
Sridevi Sattaluri, RTI International
Emily McFarlane, RTI International
Paul P. Biemer, RTI International

By taking advantage of new technology, survey managers can boost the effectiveness, efficiency and quality of data collection. One such technology is computer audio-recorded interviewing (CARI), a way to ensure the quality of data through unobtrusive digital recording. This article reviews the evolution of CARI technology, with an emphasis on its feasibility for routine use with field surveys.

Traditional methods of monitoring field staff include live observation or re-contacting the respondent to confirm the interview's authenticity and inquire about the professionalism of the interviewer. With CARI the process is much easier. Sound files can be created electronically without the need for external equipment and can be transmitted along with the usual response data files. Because the recording process is "invisible" once consent has been given, it can provide a faithful representation of the reality of in-person data collection.

Audio recording supports quality control for field surveys (Biemer et al, 2000) and telephone surveys (Basson, 2005), including those invoking web instruments (Suresh, 2005). The technology provides a potent tool for

detering and detecting falsification, providing performance feedback and enabling study of questionnaire item effectiveness.

Audio Recording Equipment, Past and Present

From the marketing of the Dictaphone in 1907 (Nuance Communications 2005) to the availability of miniature recorders embedded in portable electronic devices today (Dwyer et al, 1998), people have been using audio recording to capture voices for later review. While the early recorders helped journalistic interviews, they were not usable for large-scale research surveys. The introduction of cassette tapes improved convenience (Stockdale, 2002) but introduced logistic problems and disrupted the flow of interviewing.

As computers began offering built-in sound cards, digital audio recording became feasible for field surveys. In 1999, CARI was first deployed on a national field survey (Biemer et al, 2000), the result of innovative technical work by RTI developers R. Suresh, A. Bethke and P. Cooley. Use of CARI has spread since then as the utility of the approach has been confirmed.

Many laptops now have built-in microphones, sound cards and more disk space. Recording with internal microphones requires no additional equipment, and offers no distraction. Feedback from respondents and interviewers indicates that most people forget about recording when the microphone is hidden (Biemer et al, 2000). Depending on microphone placement and laptop settings, audio fidelity from internal microphones is adequate to capture voices within about 8 feet of the laptop at a quality level that allows a listener to distinguish multiple voices and discern the spoken content.

Benefits of Audio Recording for Quality Assurance and Questionnaire Evaluation

Among the advantages offered by CARI, perhaps the most compelling are to confirm the authenticity of data for a reduced cost compared to traditional verification methods and to provide detailed "observation" of interviewer performance (See Figure 1). CARI can act as a deterrent to curbstoning and a means of detecting poor interviewing technique. The presence of CARI might reduce cheating if interviewers are aware of being recorded.

Figure 1 Performance issues found in review of 5600 cases

Count	% of Cases	Problem Definition
13	0.2	Authenticity Questionable
217	3.9	Reading - Minor Deviation
72	1.3	Reading - Major Deviation
73	1.3	Recording Errors
44	0.8	Unprofessional Behavior
86	1.5	Inappropriate Probing
79	1.4	Feedback not Neutral
1	0.01	Incorrect Incentive Provided

Using CARI, a survey may reduce effort and costs. CARI monitoring may replace verification calls or re-interviews. It reduces respondent burden and allows confirmation of interviews from households which lack

telephones or are hard to contact. However, it remains desirable to follow up a sample of the cases since some respondents may refuse to allow audio recording after consenting to the interview, and interviewers may take advantage of that option to prevent detection of poor interviewing habits or curbstoning.

Another benefit of CARI is to provide a method for identifying questionnaire problems and data collection difficulties in interviewer-respondent interactions. CARI offers a unique opportunity to listen to the interview exactly as it took place, without the interference of personal observation. Using CARI allows questionnaire specialists to evaluate the success of the survey items in eliciting the desired information and the success of the interviewer in faithfully capturing responses (Mitchell et al, 2008).

Audio File Formats

Many audio file formats have been developed over the years, including wave, MP3, RealMedia, AIFF, CD Audio and others. The sound recording algorithm affects the audio file size, quality, playback software, platform requirements, cost and licensing. The size of a particular recorded file depends on the parameters selected in its creation (see Figure 2). File compression techniques can reduce the space needed to store the audio files.

Figure 2 File sizes and quality for uncompressed wav files

Band-width	Sampling	Chan-nels	Sound Quality	MB PerMin
8 bit	11.25 KHz	1	Low	0.66
16 bit	11.25 KHz	1	Medium	1.31
8 bit	22.5 KHz	1	Medium	1.79
16 bit	22.5 KHz	1	High	1.19
16 bit	44.1 KHz	1	Very High	5.25
16 bit	44.1 KHz	2	Very High	12.3

Integrating Audio Recording with Survey Software and Information Systems

CARI has been implemented on survey instruments in a variety of languages including Blaise (Statistics Netherlands) (Thissen and Rodriguez, 2004), CASES (University of California, Berkeley) (Biemer et al 2000) and ASP.NET (Microsoft) (Suresh, 2005). Many telephone systems used by call centers offer the capability for recording. One of the challenges of incorporating audio recording is to make the process unnoticeable to the interviewer. The recording process must not slow the system or provide any visual or audible clue as to when it starts and stops.

Once a survey instrument has been enabled with CARI technology, survey information systems (Thissen, 2004) must also be expanded to handle the audio data files. From a case management and data security perspective, CARI files are simply response data stored in a different format. The files can be transferred to the central servers using dialup transmission, broadband, or removable media like flash drives. The choice of transmission option may depend on the size of files being transmitted.

For CARI to be used during an interview, most states and countries require that participants give express consent for the interview to be recorded. Because audio files could potentially have personally identifying

information, and given the heightened consciousness of confidentiality and security concerns, audio files are best treated as sensitive data. Encryption can be used to ensure protection of the files in transit and storage.

After audio files are received at a central location, the monitoring process may be as simple as opening up the files and making notes. However, manual case management is impractical for all but the smallest of surveys, and it is best to build an interface for reviewing the files and storing evaluations.

Operational Results

In this section, we present a brief discussion of RTI's experiences with CARI technology.

At RTI, files are recorded with Windows Sound Recorder from Blaise or CASES, resulting in file sizes of about one MB per recorded minute. Use of the LAME compression algorithm (The LAME Project) yields an average compression ratio of approximately 11:1 without loss of audio quality, reducing storage to about 100KB. Recording directly to a compressed format makes a more compact file but requires more processing power, producing time lag and visible indication of recording, thus limiting its usefulness.

Audio files collected by RTI were considered adequate in quality if voices could be heard plainly and understood. Problems included background noise, static, faintness of voices, key tapping, hum and other recording problems which interfered with detection of vocal content, but generally the quality was acceptable. See Figure 3.

Figure 3 Example of CARI sound file quality distribution

Sound Quality	Number of Interviews
1 - Poor	4
2 - Passable	5
3 - Adequate	21
* - Acceptable	48
4 - Good	49
5 - Excellent	37

* Raters were allowed to leave the score blank for Acceptable.

We estimated the minimum number of CARI audio files required for making consistent monitoring evaluations. 165 interviews were coded by three independent reviewers for a pair-wise comparison of rater convergence, in which each rater listened to recordings from two groups of 55 interviews. From this, we concluded that three audio files each of 30-second duration would be adequate for verification purposes.

A theoretical cost-analysis model was created to compare the expected costs of operating traditional verification processes with CARI systems at the "steady state" in which systems had been implemented already. Analysis suggests that the cost of verification is 10% to 40% lower for CARI than for the traditional approach, depending on levels of traditional and CARI review. Results of production surveys confirm that prediction due to reduced labor.

CARI provides added benefits of reduction in respondent burden, the opportunity to review cases which are too old for re-contact and the option of having multiple reviewers for a questionable case or performance problem.

Visions of the Future

Looking forward, we see expanded use of CARI in field surveys for monitoring survey quality and as an integral part of data collection. Advances in digital signal processing may allow automation of some monitoring activities such as detecting files with just one voice, applying speech analytics or leveraging "pausology" (O'Connell and Kowal, 1983). CARI can be used to collect open-ended responses (Mitchell et al, 2008). At some future time, recordings may be transcribed automatically to text for behavior coding. Research is underway on speech-to-text conversion tools in uncontrolled surroundings (Ming, et al, 2006), which may broaden its applicability to include home environments. These research areas will enhance CARI's usefulness for years to come.

Acknowledgements

The authors would like to acknowledge the work of our colleagues at RTI in the studies mentioned here. A version of this paper was presented at the 62nd Conference of the American Association for Public Opinion Research, Anaheim, CA, 2007.

References

Biemer, P.P., Hergert, D., Morton, J. and Willis, W.(2000), "The Feasibility of Monitoring Field Interview Performance Using Computer Audio Recorded Interviewing (CARI)", *Proceedings of the American Statistical Association's Section on Survey Research Methods*, pp. 1068-1073.

Dwyer, J.J., Godin, D.K., Colon, R.S., Sr., Rothschild, S. Pawlowski, J.J., and Vaughan, J.C. (1998), "Voice file management in portable digital audio recorder", United States Patent 6671567.

The LAME Project, LAME Compression Software, <http://lame.sourceforge.net/index.php>

Mitchell, S. B., Fahrney, K. M., Strobl, M. M., Thissen, M. R., Nguyen, M. T., Bibb, B. S., & Stephenson, W. I. (2008). Using computer audio-recorded interviewing to assess interviewer coding error. In 63rd AAPOR Conference (No. 127664). New Orleans, LA.

Ming, J., Hazen, T.J. and Glass, J.R. (2006), "Speaker Verification Over Handheld Devices with Realistic Noisy Speech Data," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp I-637 to I-640.

Nuance Communications, Inc. (2005), "About Dictaphone", <http://www.dictaphone.com/aboutus/history.asp> .

O'Connell, D.C. and Kowal, S. (1983), "Pausology". In *Computers in Language Research*, Sedelow, W. A. Jr. and Sedelow, S.Y. (eds), Berlin-New York: Walter de Gruyter & Co., pp. 221-301.

Statistics Netherlands, Statistical Informatics Department, P.O. Box 4000, 2270 JM Voorburg, The Netherlands.

Stockdale, A. (2002), "Tools for digital audio recording in qualitative research", *Social Research Update*, pp 1-4.

Suresh, R. (2005). "Web-Based Computer Audio Recorded Interview (Web-CARI)." Presented at the *International Field Directors and Technology Conference 2005*, Atlanta, GA.

Thissen, M. R., and Rodriguez, G. (2004), "Recording Interview Sound Bites Through Blaise Instruments", *Proceedings of the International Blaise Users' Conference*, pp. 411-423.

University of California, Berkeley, Software Support Services, "Computer-Assisted Survey Execution System (CASES)," CSM Program, 358 Barrows Hall #3820, Berkeley, CA 94720.

Comments: 0

 [Comments](#)

TABLE 1: STATE-BY-STATE SWING ANALYSIS									
2004 USA Presidential Election Result					2008 USA Presidential Election Result				Obama
State	Bush	Kerry	Other	Kerry Lead	McCain	Obama	Other	Obama Lead	Swing
	%	%	%	%	%	%	%	%	%
Alabama	63	37	0	-26	61	39	0	-22	2.0
Alaska	61	36	3	-25	60	38	2	-22	1.5
Arizona	55	44	1	-11	54	45	1	-9	1.0
Arkansas	54	45	1	-9	59	39	2	-20	-5.5
California	44	54	2	10	37	61	2	24	7.0
Colorado	52	47	1	-5	45	54	1	9	7.0
Connecticut	44	54	2	10	38	61	1	23	6.5
Delaware	46	53	1	7	37	62	1	25	9.0
Florida	52	47	1	-5	49	51	0	2	3.5
Georgia	58	41	1	-17	52	47	1	-5	6.0
Hawaii	45	54	1	9	27	72	1	45	18.0
Idaho	68	30	2	-38	61	36	3	-25	6.5
Illinois	45	55	0	10	37	62	1	25	7.5
Indiana	60	39	1	-21	49	50	1	1	11.0
Iowa	50	49	1	-1	45	54	1	9	5.0
Kansas	62	37	1	-25	57	41	2	-16	4.5
Kentucky	60	40	0	-20	58	41	1	-17	1.5
Louisiana	57	42	1	-15	59	40	1	-19	-2.0
Maine	45	54	1	9	41	58	1	17	4.0
Maryland	43	56	1	13	37	62	1	25	6.0
Massachusetts	37	62	1	25	36	62	2	26	0.5
Michigan	48	51	1	3	41	57	2	16	6.5
Minnesota	48	51	1	3	44	54	2	10	3.5
Mississippi	60	40	0	-20	57	43	0	-14	3.0
Missouri	53	46	1	-7	49	49	2	0	3.5
Montana	59	39	2	-20	50	47	3	-3	8.5
Nebraska	66	33	1	-33	57	42	1	-15	9.0
Nevada	51	48	1	-3	43	55	2	12	7.5
New Hampshire	49	50	1	1	45	54	1	9	4.0
New Jersey	46	53	1	7	42	57	1	15	4.0
New Mexico	50	49	1	-1	42	57	1	15	8.0
New York	40	58	2	18	37	62	1	25	3.5
North Carolina	56	44	0	-12	49	50	1	1	6.5
North Dakota	63	36	1	-27	53	45	2	-8	9.5
Ohio	51	49	0	-2	47	51	2	4	3.0
Oklahoma	66	34	0	-32	66	34	0	-32	0.0
Oregon	47	51	2	4	41	57	2	16	6.0
Pennsylvania	48	51	1	3	44	55	1	11	4.0
Rhode Island	39	59	2	20	35	63	2	28	4.0
South Carolina	58	41	1	-17	54	45	1	-9	4.0
South Dakota	60	38	2	-22	53	45	2	-8	7.0
Tennessee	57	43	0	-14	57	42	1	-15	-0.5
Texas	61	38	1	-23	55	44	1	-11	6.0
Utah	72	26	2	-46	63	34	3	-29	8.5
Vermont	39	59	2	20	31	68	1	37	8.5
Virginia	54	46	0	-8	47	53	0	6	7.0
Washington	46	53	1	7	41	58	1	17	5.0
West Virginia	56	43	1	-13	56	43	1	-13	0.0
Wisconsin	49	50	1	1	43	56	1	13	6.0
Wyoming	69	29	2	-40	65	33	2	-32	4.0
Dist Columbia	9	90	1	81	7	93	0	86	2.5
USA	51	48	1	-3	46	53	1	7	5.0
Analysis by Sir Robert Worcester and Tomasz Mludzinski, Ipsos MORI, London: rmworcest@yaho.com				Voting in 2008	45.9%	52.7%	1.3%	6.8%	©
				127,142,278	58,421,377	67,066,915	1,653,986	8,645,538	28.11.08

TABLE 2: DEMOGRAPHIC SWING ANALYSIS

Demographics		2004 USA Presidential Election Result					2008 USA Presidential Election Result					Obama Swing
		Voters	Bush	Kerry	Other	Kerry Lead	Voters	McCain	Obama	Other	Obama Lead	
		%	%	%	%	%	%	%	%	%	%	%
All Voters		100	51	48	1	-3	100	46	53	1	7	5.0
Sex	Men	46	55	44	1	-11	47	48	49	3	1	6.0
	Women	54	48	51	1	3	53	43	56	1	13	5.0
Age	18-29	17	45	54	1	9	18	32	66	2	34	12.5
	30-44	29	53	46	1	-7	29	46	52	2	6	6.5
	45-64	30	51	48	1	-3	37	49	50	1	1	2.0
	65+	24	54	46	0	-8	16	53	45	2	-8	0.0
Race	White	77	58	41	1	-17	74	55	43	2	-12	2.5
	African - American	11	11	88	1	77	13	4	95	1	91	7.0
	Latino	8	44	53	3	9	9	31	67	2	36	13.5
	Asian	2	44	56	0	12	2	35	62	3	27	7.5
	Other	2	40	54	6	14	3	31	66	3	35	10.5
Income	Under \$15,000	8	36	63	1	27	6	25	73	2	48	10.5
	\$15-30,000	15	42	57	1	15	12	37	60	3	23	4.0
	\$30-50,000	22	49	50	1	1	19	43	55	2	12	5.5
	\$50-75,000	23	56	43	1	-13	21	49	48	3	-1	6.0
	\$75-100,000	14	55	45	0	-10	15	48	51	1	3	6.5
	\$100-150,000	11	57	42	1	-15	14	51	48	1	-3	6.0
	\$150-200,000	4	58	42	0	-16	6	50	48	2	-2	7.0
	\$200,000 or More	3	63	35	2	-28	6	46	52	2	6	17.0
Education	No High School	4	49	50	1	1	4	35	63	2	28	13.5
	H.S. Graduate	22	52	47	1	-5	20	46	52	2	6	5.5
	Some College	32	54	46	0	-8	31	47	51	2	4	6.0
	College Graduate	26	52	46	2	-6	28	48	50	2	2	4.0
	Postgraduate	16	44	55	1	11	17	40	58	2	18	3.5
Party ID	Democrat	37	11	89	0	78	39	10	89	1	79	0.5
	Republican	37	93	6	1	-87	32	90	9	1	-81	3.0
	Independent	26	48	49	3	1	29	44	52	4	8	3.5
Ideology	Liberal	21	13	85	2	72	22	10	89	1	79	3.5
	Moderate	45	45	54	1	9	44	39	60	1	21	6.0
	Conservative	34	84	15	1	-69	34	78	20	2	-58	5.5
First time voter	Yes	11	46	53	1	7	11	30	69	1	39	16.0
	No	89	51	48	1	-3	89	48	50	2	2	2.5
When did you decide who to vote for?	Today	5	45	52	3	7	4	45	50	5	5	-1.0
	Last Three Days	4	42	55	3	13	3	52	47	1	-5	-9.0
	Last Week	2	51	48	1	-3	3	50	48	2	-2	0.5
	In October	N/A	N/A	N/A	N/A		15	43	54	3	11	
	In September	N/A	N/A	N/A	N/A		14	45	54	1	9	
Before That	N/A	N/A	N/A	N/A		60	47	52	1	5		
Religion	Protestant	54	59	40	1	-19	54	54	45	1	-9	5.0
	Catholic	27	52	47	1	-5	27	45	54	1	9	7.0
	Jewish	3	25	74	1	49	2	21	78	1	57	4.0
	Other	7	23	74	3	51	6	22	73	5	51	0.0
	None	10	31	67	2	36	12	23	75	2	52	8.0
Union Members	Union member	14	38	61	1	23	12	37	60	3	23	0.0
	Non-union member	86	54	45	1	-9	88	46	52	2	6	7.5
McCain's choice of Palin	Most Impt Factor	N/A	N/A	N/A	N/A		7	47	52	1	5	
	Important Factor	N/A	N/A	N/A	N/A		33	52	47	1	-5	
	Minor Factor	N/A	N/A	N/A	N/A		20	66	33	1	-33	
	Not a Factor	N/A	N/A	N/A	N/A		33	33	65	2	32	
Analysis by Sir Robert Worcester and Tomasz Mludzinski, Ipsos MORI, London: rmworcester@yahoo.com		Source: Edison Exit Poll Data				Voting in 2008		45.9%	52.7%	1.3%	6.9%	©
		Obtained from CNN Website				127,142,278	58,421,377	67,066,915	1,653,986	8,645,538	28.11.08	