by Steve Ross
Sunday, January 12, 2003 12:00 AM

With races for the United States Senate and the Presidency running neck-and-neck, New Yorkers reside in a special hell this year -- a polling hell. If the election were being held today, all we REALLY know is that some politicians would be winners, and others would be losers.

The only advantage of the races being so close this year is that New Yorkers will be spared the normal spectacle of a candidate way behind in the polls suddenly, to the media's amazement, ALMOST closing the gap as Election Day approaches. This is not some odd inexplicable (and regularly repeated) miracle. It is a calculated strategy on the part of the candidate ahead in the polls: Better to run out the clock, letting the opponent get close, than to change the stump speech, introduce a new idea, and thus perhaps risk a gaffe that could cost the election.

But with Bush or Lazio ahead in one day's poll, and Gore or Clinton ahead in the next day's, we are getting a horse race that is less exciting than confusing - as well as pointless, and inaccurate.

I'm a big fan of polling. But I'm not a big fan of most reporting about polls. News organizations use polls to call a winner. We should instead be using polls to understand the pols, and how they are framing their campaigns. Journalists should be more upfront about the limitations of our polling methods -- and there are many. And we should spend more money to do the job of polling correctly.

MIXING APPLES AND ORANGES

Polls showed Al Gore way ahead in the race for the Presidency right after the Democratic National Convention in August, but then falling behind by the second week of October. Was this a measurement of an actual decline in support for the Democratic nominee? I think not. The early polls were polls of registered voters. The later polls were polls of likely voters (those who have voted in the last few elections). Democrats are somewhat less likely to vote than are Republicans. This makes the Democrats show up better in polls of registered voters, while Republicans show more strength in polls of likely voters. This crucial fact is occasionally explained in print stories about polls. It is never explained on radio or local TV.

If polling likely voters is more accurate than polling registered voters, and if movement up or down from week to week can only be measured accurately if the polling method is consistent from poll to poll, why don't we always just conduct polls of likely voters? One answer is simply

that it is more expensive and more time-consuming to do so. It literally doubles the time and cost to find a likely voter rather than make do simply with a registered one. But there is an even less honorable reason, which I will explain in a moment.

WHY POLLS ARE INCREASINGLY INACCURATE

Even if just one polling method were used consistently, the truth is, typical polls are much less accurate than readers (and, I suspect, editors) have been led to believe. There are many reasons for this.

First, voters are getting harder and harder to sample. Pollsters nowadays make five to ten calls to get one valid response. In the 1988 presidential race, pollsters made two or three calls to get a valid response, and complained about that.

Second, people move more frequently, changing their voting address on average once every five years or so. This makes it more difficult for pollsters to confirm likely voters by asking for whom they voted in previous elections.

Third, new patterns of quasi-marriage and racial intermarriage that have emerged over the past decade have made it more difficult to "profile" a voting district. Pollsters must make an estimate of all voters' age, family status, gender and race to draw a representative sample. If 50 white Christian males from a certain district are polled, do they stand for 5,000 likely voters in the district's population or 8,000? If 5,000, then their answers might be magnified in the poll 60 percent more than if they stand for 8,000. Is a family counted as black if only the husband is black? Pollsters usually do. Is a family counted as Hispanic if only the husband is Hispanic? Some pollsters do, at least for some election polls, and some don't. Does this square with YOUR views on family life?

This campaign season presents extra difficulties for pollsters because it has been a full ten years since the last federal census. The data collected in the count this April will not be released until well after the elections.

They simply are not ready. That means pollsters have been dependent on 1990 base data, updated by marketing data collected by commercial companies. Such data tends to be pretty good in wealthy districts, but not in poor districts. Marketers are simply not as interested in poor

districts, and thus don't put as much effort into understanding them. The typical American has moved at least once since the last census in 1990, and the poor have, on average, moved even more often.

Finally, voters' opinions are often quite fluid and changeable. Something in the news that day could affect their opinions, which could change the next day. They also have become more poll-savvy. They shape their responses to what they think pollsters want to hear. Seven years ago, polls showed David Dinkins winning over Rudy Giuliani. It turned out that many people who simply would not vote for a black candidate told pollsters they would. The same thing could happen this year in the Presidential race, with a Jew on the ticket.

INHERENT INACCURACY: THE "MARGIN OF ERROR"

This season's races are being called "statistical dead heats" and "too close to call," even though on any day a poll will show one candidate with at least a small lead. My own informal discussions with members of the public suggest they don't really understand how there can be a tie if the numbers show one candidate ahead in a poll, nor what the anchors mean when they throw in something about a "margin of error of plus or minus 3.5 percent." If they have any opinions at all about this, they ascribe the weasel wording to imperfections in polling techniques -- the difficulties in getting the most representative people to interview, due to the issues discussed above.

But even if the sample of actual voters polled were an exact demographic representation of the electorate as a whole (which is never the case), the results would still not mirror the real world exactly. This is a simple fact of statistics. Think about what happens when we "poll" a penny about "voting" heads or tails. If a perfectly balanced coin is tossed many thousands of times, the "poll" will end up even: Half the time the penny will wind up heads, and half the time tails. But if we toss it up only ten times, it is more likely to have uneven results, perhaps seven heads to three tails. We call this a "winning streak." Thousands of tosses have many such streaks of heads or tails, which cancel each other out.

Choosing a person to be polled is like tossing a penny. One person may say Lazio (Heads!). A second person, demographically identical, might choose Clinton (Tails!). To make sure the "streaks" cancel out, the pollster must talk to many, many voters -- more, indeed, than they usually talk to.

That is why polls usually carry a disclaimer -- a "margin of error" -- which is based loosely on the

work of a mathematician named Jacques Bernoulli. Bernoulli died more than 80 years before the Constitution was ratified, so he never met a United States Senator. In fact, the city where he lived was close to the border between what is now Italy and Switzerland, and changed hands during his lifetime, so he tended to stay out of politics. But he understood the process. [Math-averse readers may want to skip the next several paragraphs] He said that if we "poll" the coin 1,000 times, we would rarely get a precise 500-500 tie between heads and tails. But 19 times out of 20, he discovered, we will get a result between 465 and 535.

Statisticians call that "19 times out of 20" an expression of the "confidence level." If you divide 19 by 20, you get 0.95, or 95 percent. The difference between 465 and 500 (or 500 and 535) represents the margin of error--35, in our example. The 35 divided by 1,000 (the size of the sample) is 0.035, or 3.5 percent. That 95 percent level is the most commonly used in the statisticians' version of the real world.

Thus, a poll of 1,000 people would say that 95 percent of the time the reported results fall within 3.5 percent of the results that could be expected if the entire electorate were polled. That statement is roughly in line with one approved more than a decade ago by the American Statistical Association
.

The smaller the number of people polled, obviously, the larger the margin of error. But that is true as well for a "subsample" in a large poll. Given our hypothetical newspaper poll of 1,000 people, for example, if we want to talk about men's opinions (based on the 500 men interviewed for the poll), the margin of error has increased to plus or minus five percent. If we talk about black men's opinions (based on the 100 black men interviewed), the error limit reaches more than 10 percent. And if we talk about Hispanics' views (based on 50 Hispanics interviewed) the error limit cannot even be calculated using Bernoulli's math.

The problem is not only that the public is generally uninformed about these inherent inaccuracies in polling statistics. It is also that the media rarely provide the basic information about specific polls -- such as sample sizes, dates the poll was run, error limits - in order for even informed viewers to assess them. This is certainly true when they are reporting on polls by other organizations -- and most people read or hear about one news organization's poll only when another news organization reports it.

WHY THE PRESS USES POLLS ANYWAY

by Steve Ross
Sunday, January 12, 2003 12:00 AM

It is not just the public who understands little about polls. Most reporters deliver the results of a steady stream of polls -- simplistically reducing the results to "who's ahead, who's behind" -- without themselves having seen (much less analyzed) any of the raw data or any detailed explanations of their methodology. Why do they do this? By using polls for "horse race" reporting, news reporters don't really have to think, and the news organizations that employ them don't have to pay them extra for the time to do so. Why actually read and understand the candidates' position papers when you can report on poll results, especially results of others' polls you haven't put any money or effort into?

But saving time, expense and thought are not the only motivations for bad polling methods and bad reporting of poll results. One definition of "newsworthy" is "surprising." A badly done poll may produce surprising results, and thus be more widely repeated!

Paradoxically, our shallow use of polls often results in the least surprising, and most dubious, interpretation of the reality of candidates and their campaigns. We as journalists wind up reinforcing the standard images of the candidates (often created, or helped along, by their opponents). It is hard for me to believe that George W. Bush is as dumb as he looks, that Al Gore is so wooden, that Hillary Clinton is so dishonest or that Rick Lazio is so mean.

WHY THE PRESS SHOULD PAY ATTENTION TO POLLS

Despite their inherent problems, news organizations and academic institutions should be polling, to give them better insight into the candidates' campaigns. Consider this defensive polling. Candidates themselves poll constantly - first to produce numbers they can show to early donors, often even before they announce a candidacy. Candidates then poll to see potential voters' reactions to specific issues, and thus refine their positions on those issues. And candidates track their progress during the campaign itself with weekly and sometimes even daily polls.

Polling is thus integral to modern campaigning. Before Election Day itself, candidates' carefully honed images, guided by polling, are shaping the outcome.

The media should be focusing on the reality behind those images, images that tend to become more cartoon-like and one-dimensional as Election Day approaches. The closeness in the polls this year, for example, can help explain why the candidates are still experimenting with their messages, only a few weeks shy of the election.

by Steve Ross
Sunday, January 12, 2003 12:00 AM

There are many polling techniques that have been developed over the years, mainly for use in survey work on sexual practices, that can overcome the inherent errors in polling. Yes, they would cost more, about double what it normally costs now (which is up to $50,000 for a poll of 1,000 voters). It seems a small price to get it right.

Steve Ross teaches at Columbia University's Graduate School of Journalism, and conducts several polls and surveys each year, including the largest annual survey of how journalists use cyberspace. He has authored 18 books, including one on statistics.