tion day. For example, the vast majority of the newly registered will vote, albeit at a slightly lower rate than those who have been habitual voters in the past.

An arbitrary but reasonable guess is that about 85% of the likely voters voted while 60% of the less likely voted. (These numbers add up to 80% of registered voting.) If so, the true projected margin would be neither the +2 for likely voters alone nor the +6 for registered voters as a group, but instead +5.

$$\frac{.8 * .85 * (+2) + .2 * .60 * (+22)}{.8} = +5$$

This estimate is much closer to the registered voter margin of +6 than to the likely voter margin of +2. This +5 estimate could be achieved by properly weighting respondents by their likelihood of voting.

Did those less likely voters who did go to the polls in 1992 vote disproportionately Democratic? At least for those who were judged less likely because they were first-time voters, we can get some help on this from exit poll data. According to the VRS exit poll, self-described first-time voters comprised 11% of the electorate and supported Clinton by a strong +18 margin—which is generally consistent with the argument made here. For the overall numbers to add up to the national +5 Clinton margin, the 89% of the electorate who had voted previously must have given Clinton roughly a +3 margin. This is an unusually large gap in the vote by new vs. established voters.

The moral of the story is twofold. First, surveys usually do not need to correct for turnout beyond screening out non-registrants. Second, in the rare case where screening for turnout matters, it is better not to correct at all than to correct incorrectly.

Robert Erikson is Distingushed University Professor, department of political science, University of Houston

COUNTING LIKELY VOTERS: A REPLY TO ERIKSON

By G. Donald Ferree, Jr.

Robert Erikson raises provocative questions about the propriety of screening for "likely voters" and its impact on the projections of the vote. These questions are worthy of close attention in the survey community. But his conclusion—that any scheme other than assigning a probability of voting to each respondent and then weighting on the basis of that probability is methodologically indefensible—is seriously flawed.

Erikson assumes that because introducing a likely voter screen lowered the Clinton margin (as compared to all registered voters) at the start of the week prior to the election, it must follow that this screen systematically overstated Bush's vote. But, as Hugick *et al.* noted in their piece (see *Public Perspective*, January/February 1993, pp. 12-13), the Gallup surveys showed a widening gap later in the week—as did others. The final Gallup tally in fact *overstated* Clinton's margin. If Erikson is right, and adopting

a likelihood-weighting procedure for all respondents would have boosted Clinton's estimated vote, Gallup's use of this procedure would have left it with an even higher overestimation!

Some of Erikson's argument is unobjectionable. Compared to the possible electorate, the actual electorate consists

The final Gallup tally in fact overstated Clinton's margin. If Erikson is right, and adopting a likelihood-weighting procedure for all respondents would have boosted Clinton's estimated vote, Gallup's use of this procedure would have left it with an even higher overestimation!

of 100% of those who had a certainty (p=1.0) of voting, 90% of those who actually had a .9 chance of voting, 80% of those with a .8 chance of voting, and so on

down.[1] If there *were* a way of accurately assessing this actual probability prior to the election, no sane researcher could possibly object to weighting respondents by this factor.

But what if, as is undoubtedly the case, it's not possible to determine accurately individuals' likelihood of voting? Does *any* attempt, however inaccurate, do more good than harm, as Erikson would seem to assert? The answer depends on such factors as how similar turnout patterns are across elections, how estimated turnout relates to preference, and whether events occurring between the survey and election day change individuals' likelihood of voting.

Many concerns here parallel those that apply to any question of weighting. It is a commonplace that any actual sample may differ from the theoretical population from which it is drawn not only randomly (quantified in the so-called "margin of error"), but systemati-

cally. If something causes some persons to be more likely than others to come into the sample, those who are easier to find will be overrepresented, while those who are harder to find will be underrepresented. The consequences of this fact depend on how much and in what ways, "easy to find" persons differ from those who are "harder to find". The more the two groups differ, the greater the difference will be between the "biased" sample and what one would get from a truly random sample.

Suppose, however, that one takes demographic characteristics in the sample and then weights to known (or estimated) population characteristics. For example, because *as a group* women differ from men in ways that may affect how easily they can be found in a survey (e.g., working outside the home), samples often get more women than they should. Weighting by gender would "correct" the gender marginals. But it would also mean that women who are more like men (as a group) would be unfairly *down*weighted; those men who are more like women in these same characteristics would be overweighted. This could be worse for the accuracy of the final marginals than not weighting at all.

Likelihood of voting is typically calculated by looking at demographic characteristics, attitudinal variables such as interest in an election, and past voting behavior. After an election, one can—in a properly constructed model—"postdict" actual turnout based on these characteristics. Further, one can also determine to what extent turnout patterns were similar or dissimilar to previous elections. Before an election, one must just assume that past patterns will hold. If they do, weights for individual potential voters merely introduce one more source of random variation into the data.[2]

If patterns change, however, the problem is far worse. Erikson (quite properly) faults Gallup for not taking into account the possible presence this year of many "new" voters. But weighting based on past behavior won't catch such people

either. Moreover, weighting based on predicted likelihood of voting is vulnerable to factors that can change between the survey and election. For instance, a "new issue" could arise that mobilizes some voters, but not others. Expected outcome may render an individual's vote more or less efficacious or crucial. The possibilities are almost endless even before one gets to factors (such as campaign organization, weather, etc.) which cannot be measured on a survey in any event.

Erikson is simply incorrect when he asserts that it is 'common knowledge among survey statisticians' that dichotomizing into likely and nonlikely voters overcorrects for any vote differential. Everything depends on the pattern of preference and how it relates to (estimated) likelihood to vote.

There are also reasons for assuming that the actual estimate may itself be related to preference. Suppose, for instance that Clinton and Perot voters were motivated by the desire to cast a negative ballot against the incumbent. This could well show up in statements about certainty to vote, importance of the election, and so on. It is simply not possible to know in advance what the various relationships between likelihood to vote, and preference are.

It's clear, then, that assigning an individual a precise probability of voting is fraught with imprecision. But is doing this imperfectly *necessarily* better or worse than relying on the more straightforward dichotomy? Erikson is simply incorrect when he asserts that it is "common knowledge among survey statisticians" that dichotomizing into likely and non-likely voters overcorrects for any vote differential (see p. 22). Everything depends on the pattern of preference and how it relates to (estimated) likelihood to vote.

If one dichotomizes a sample into "likely voters" (all fully counted) and "non-likely" voters (not counted at all), any person who has a likelihood below

1.0 but above the cutoff is *over*counted (more so as actual likelihood goes down); any person falling below the cutoff but with a true likelihood above 0.0 is *under*counted (less so as actual likelihood goes down). Even if preference for Bush, say, was known to go down linearly with likelihood to vote—a highly dubious proposition in any event—the impact depends on how many such people there are at each level of likelihood. If people cluster around 1.0 and .6, say, where that is the cutoff, weighting all "likely" voters 1.0 would actually overestimate the expected vote for Clinton— exactly the opposite of what Erikson contends.

If alternate ways of weighting, or otherwise seeking to identify the actual electorate from among the theoretically eligible electorate, produce sharply different projections of the vote, perhaps the best that can be done is to point this out as a caveat in interpreting results. The quest for the elusive grail of the single best way of identifying "likely voters" is doomed to failure.

Endnotes
[1] The cited probabilities are chosen as examples only. Presumably, true likelihood of voting is a continuous variable which could take on *any* value between zero and one, inclusive.
[2] Since there is imprecision in the weight actually assigned to any individual, the results will differ more, sample to sample, than if data were unweighted. Moreover, the overall "marginals" are highly influenced by the groups which are upweighted, but which were originally smaller, with more group-based "margin of error."

_____

G. Donald Ferree, Jr. is associate director, the Roper Center for Public Opinion Research, University of Connecicut