

AAPOR Statement: Understanding a “credibility interval” and how it differs from the “margin of sampling error” in a public opinion poll

Summary:

AAPOR urges caution in the interpretation of a new quantity that is appearing with some nonprobability opt-in, online polling results – the *credibility interval*. The credibility interval is not the margin of sampling error (MOE) that the public has come to understand as the statistical uncertainty of probability based scientific polls. Instead, credibility intervals are being used when reporting on some nonprobability polls (typically opt-in online polling), or for model-based inferences such as arise in small-area estimation. The credibility interval reflects the statistical uncertainty generated by a statistical model that relies on Bayesian statistical theory. It requires the pollster to make statistical modeling choices that translate the observed participant observations (e.g., from an online opt-in poll) into results reflecting the targeted group to which the poll was intended (e.g., likely voters, general public). The credibility interval adopts a very different statistical approach from conventional probability sample polls, and this results in an interpretation that differs from that of the MOE. Using this framework, statistical insights (i.e., inferences, estimates and assessments of error) are derived from models that explain the outcome of interest (e.g., the percentage favoring Romney) and link the sample to the specific target population of interest (e.g., likely voters). In contrast, statistical insights from probability samples do not depend on any explicit model linking the sample to the target population (although implicit models are needed to adjust for nonresponse and noncoverage). While the adoption of the credibility interval is an appropriate use of statistical model-based methods, the underlying biases associated with nonprobability online polls remain a concern. As a result, the public should not rely on the credibility interval in the same way that it can with the margin of sampling error. Moreover, the Association continues to recommend the use of probability based polling to measure the opinions of the general public. Our full statement appears below.

AAPOR Statement on Credibility Intervals

As the post-convention election season swings into full throttle, the media and public are being inundated with national and local polling results. Nonprobability opt-in online polls are included in the

mix. AAPOR has already issued a caution on the scientific quality of nonprobability online polls. See <http://poq.oxfordjournals.org/content/early/2010/10/19/poq.nfq048.full.pdf?ijkey=0w3WetMtGltMuXs&keytype=ref> for a full discussion. But a new twist on the presentation of nonprobability-based polling results has emerged that deserves commentary – publication of **credibility intervals** in place of conventional margins of sampling error (MOE). AAPOR is issuing this statement to inform the public and the media about this newly emerging concept in polling.

What is a credibility interval? Credibility intervals reflect the uncertainty of statistical results generated through the use of *Bayesian statistical methods*. Both MOEs and credibility intervals are designed to relay the statistical uncertainty of polling results. But there is a big difference: credibility intervals are explicitly dependent on underlying assumptions tied to the statistical model chosen for the study, whereas classical margins of sampling error depend only on sampling design (as well as implicit assumptions underlying the weighting adjustments). These differences affect the way the intervals are interpreted.

The credibility interval is a range by which the polling estimate can vary (e.g., +/- 3 percentage points), where the range is determined by a probability chosen by the pollster – aka the credibility level -- typically set to 90 or 95 percent, similar to the *confidence* levels used in a margin of sampling error. To illustrate, suppose that a poll found a 47 percent favorable rating for Romney with a 95% *credibility interval* of 3 percentage points. The interpretation of the credibility interval is as follows: *There is a 95 percent chance that the true percentage of people supporting Romney is between 44 and 50 percent (i.e., 47% plus or minus 3%).* The credibility interval depends on the statistical model that the researcher chose for the study. If the underlying assumed model fails to hold, so too does the validity of the credibility interval.

Where do credibility intervals come from? Credibility intervals come from models within a broad area of mathematical statistics called *Bayesian statistics*. Under this approach, statistical models are chosen by a professional expert to fit the situation (e.g., a political poll of adults in the U.S. during the presidential election season) and objectives of their study (e.g., estimating presidential candidates' favorable ratings). The actual estimate along with its corresponding credibility interval are both affected by the choice of model. If the model is chosen well, then the polling results are reasonably free of systematic error and the uncertainty of the statistical estimate is properly reflected in the credibility interval. But if the assumptions underlying the statistical model fail to hold, then the study's results could be misleading and seemingly more accurate than they actually are. Bayesian theory and methods are well established and accepted in the statistical community. The big challenge -- especially for polls using opt-in online nonprobability methods -- has always been in *knowing to what extent the underlying assumptions are satisfied for any give poll.*

Bayesian analysis is sometimes referred to as *subjective probability*. The *subjective* characterization stems from the starting point -- specifying one's prior knowledge about what you are measuring (via the expert's specification of a mathematical model), and then updating it via the collection of new information (e.g., via a nonprobability poll). The Bayesian approach itself can be used with probability or nonprobability sampling.

Most applications of Bayesian methods in probability-based sample surveys adopt minimal expressions of ‘prior knowledge.’ At the extreme, it can be shown that conventional estimates from probability samples can be obtained using Bayesian methods coupled with a prior knowledge model reflecting total ignorance. However, in practice a Bayesian analysis will not assume total ignorance, as the reason for using such methods in the first place is to make use of additional assumptions and knowledge to obtain estimates in settings such as nonprobability sampling where classical statistical methods fail. A challenge arises when these methods are used with nonprobability polls such as opt-in online panels. Experts’ expressions of ‘prior knowledge’ can vary from one non-probability sample to another when estimating the same quantity (e.g., percentage of likely voters in favor of Obama) across vendors or polls.

How do credibility intervals differ from the usual margin of sampling error? Margins of sampling error (MOE) reflect the statistical uncertainty of an estimate generated from a probability sample and based on a pre-specified level of confidence, conventionally set at 90 or 95 percent in polling circles. A confidence interval is created by taking the MOE and adding it to and subtracting it from the estimate. The interpretation of the confidence interval is as follows (assuming a 95 percent level of confidence for illustrative purposes): *If the poll was repeated a large number of times using the same probability sampling method, we could expect that 95 percent of the confidence intervals would include the true value somewhere inside the confidence interval.* This classical statistical method neither requires nor relies on the specification of a prior statistical model to generate valid estimates (in contrast to a Bayesian approach). It only relies on the data and design information from a poll that used a probability sample design (plus some assumptions used for weighting adjustments). In a sense, probability based polls enjoy a sturdiness (i.e., robustness) over approaches that relies on the validity of their statistical model. It is for this reason that *probability based polling* has been considered the gold standard in the polling community for the last half century.

Why would pollsters compute credibility intervals using a Bayesian approach? In today’s society it is increasingly difficult and more expensive to conduct probability sample surveys. Wireless telephone technology and concerns about privacy are only two of several reasons why people are increasingly more difficult to reach. And once you make contact, people are increasingly reluctant to participate in a poll. Opt-in online panel surveys/polls are intended to offer an economical and timely solution because of the ease and low costs of implementation, as well as their typically large sample size. Currently, it is impossible to develop statistically valid margins of sampling error from nonprobability surveys, such as opt-in, online polls. The Bayesian approach is a logical choice since it can be applied irrespective of how the sample is drawn. It does, however, elevate the importance of correctly choosing the statistical model that links the sample to the target population through which the polling data are filtered and adjusted to produce results.

AAPOR continues to be concerned with the underlying biases associated with nonprobability opt-in online polls. As noted in AAPOR’s opt-in online Task Force report, such samples are seldom if ever representative of the target population (e.g., not representative of likely voters in pre-election polling studies), and a model’s attempt to make this link based on known variables may not be sufficient.

The emergence of credibility intervals is a genuine attempt to provide consumers of *nonprobability polls* with a measure of statistical uncertainty, albeit one that is tempered by pollster's professional expertise in choosing an appropriate model. But the inherent problems, challenges and cautions outlined in AAPOR's original 2010 report remain as relevant today as they were at the time of release.

Is the bottom line really as simple as “margins of sampling error are better than credibility intervals because of the need to choose a statistical model that cannot be validated”? Actually, things are not as simple as they seem. On the one hand, probability based polls (which can produce MOEs) encounter high levels of nonparticipation, and consequently adjustments must be made to (hopefully) correct for that. Moreover, some (though not all) probability based polls further suffer from high levels of noncoverage in which whole segments of the target population do not receive any chance of selection (e.g., newly registered voters in a sample drawn from voter registration lists; persons with cell telephones in a sample of landline phones). Noncoverage in probability sample poll also requires that weighting adjustments be made to compensate for the loss of representation. To the extent that these adjustments inadequately correct for biases, the polling estimates could be erroneous in a way that the MOEs do not or cannot reflect. And while much scientific research has gone into the development and testing of adjustments to correct for nonparticipation and noncoverage in probability sample polls, some uncertainty always remains about the effectiveness of correcting for potential biases.

The reality is that not all probability samples are designed well, and even the best design cannot compensate for serious nonparticipation by the public. Indeed, there are many instances of probability samples performing poorly because they were poorly designed and/or executed. It is for this reason that AAPOR promotes its ***Transparency Initiative*** (see http://www.aapor.org/Transparency_Initiative/5363.htm). Only through full knowledge of the design and methods (including how the questions were asked) can a proper assessment be made of the quality of a poll.

Conclusion. AAPOR urges caution using a new quantity that is appearing with some opt-in online polling results called the ***credibility interval***. The credibility interval is ***not*** the margin of sampling error that the public has come to expect to judge the statistical uncertainty of scientific (i.e., polls based on probability sampling) polls. Instead, credibility intervals reflect the uncertainty of a polling estimate by assuming that the pollster chose a valid underlying statistical model. The interpretation of a credibility interval differs as well. It represents the probability (with 90% or 95% being the conventional levels used) that the interval will include the true target population value *provided that the adopted model is valid*. Similarly, classical margins of sampling error are accurate to the extent that the assumed sampling and nonresponse models are correct. The size of the reported credibility interval can vary according to the pollster's model choice. Unfortunately, there is no easy way to validate such models. As a result, the public cannot rely on the credibility interval in the same way that it traditionally has with the margin of sampling error. (Note that the margin of sampling error, too, is not as reliable as it used to be due to increasing difficulty in public participation.) AAPOR continues to caution researchers in interpreting inferences from nonprobability samples when their objective is to estimate population values. In short: “Consumer, be aware.”

Prepared by: *AAPOR Ad Hoc Committee on Credibility Interval* (Robert Santos, Trent Buskirk, Andrew Gelman)