

Margin of Sampling Error/Credibility Interval

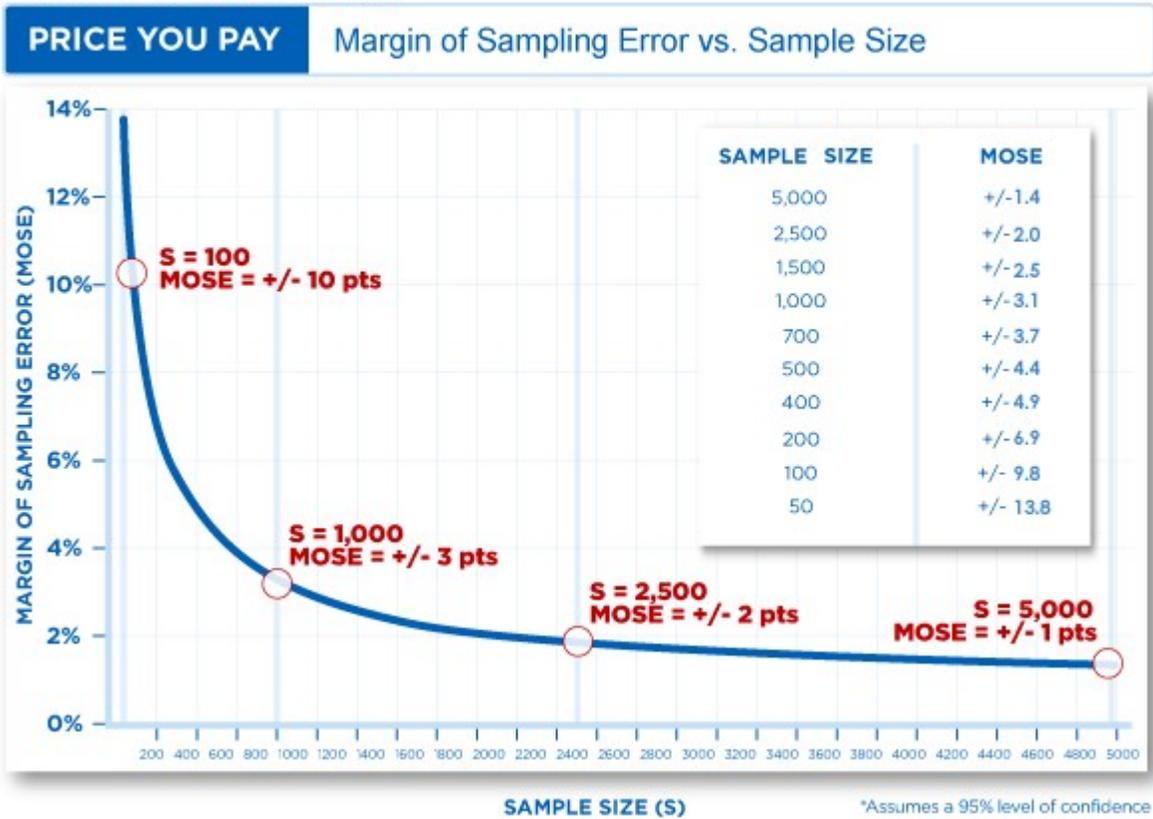
The margin of sampling error is the price you pay for not talking to everyone in the population you are targeting. It describes the range that the answer likely falls between *if* we had talked to everyone instead of just a sample. For example, if a statewide survey of adults with a margin of error of plus or minus 3 percentage points finds that 58% of the public approve of the job their governor is doing, we would be confident that the true value would lie somewhere between 55% and 61% if we had surveyed to the whole adult population in the state.

To be technically correct, we really only have *some degree* of confidence that our margin of sampling error (MOSE) falls within the calculated range. Generally, pollsters calculate the MOSE using a 95% confidence level. That is, in 95 times out of a 100, we expect that the answer we get from the survey will fall somewhere within our margin of sampling error. But about five times out of 100 it will not – one reason findings from even the best survey should be interpreted cautiously, particularly those that are significantly different from similar polls conducted at about the same time. Also, the MOSE varies depending on the percentage: It is largest when the population percentage is 50 percent and that is the figure pollsters typically use in reporting the MOSE.

Note that margin of sampling error is always expressed in percentage points, not as a percentage – for example, three percentage points and not 3%. And the margin of sampling error only applies to probability-based surveys where participants have a known and non-zero chance of being included in the sample. It does not apply to opt-in online surveys and other [non-probability based polls](#). More about this in a moment.

Sample Size and MOSE

UNDERSTANDING & INTERPRETING POLLS



When it comes to minimizing sampling error, bigger is better—up to a point. The larger the sample, the smaller the sampling error. But look at the accompanying chart, which shows the margin of sampling error for a simple random sample of a population where all people are equally likely to respond to the survey. (*Note that this condition is not the case for most polls, and is typically corrected through weighting. This does affect the MOSE – more on that below*). Notice that as the sample size increases, the margin of sampling error falls dramatically between small sample sizes of say 100 and larger samples of 1,000. But once we get to 1,000, additional sampling error falls only slightly. In fact, doubling the sample size from 1,000 to 2,000 only reduces the margin of sampling error by about a single percentage point.

But be careful: The overall sampling error applies to the total sample and not to subgroups, which have a different MOSE based on their size. Consider a survey of 1,000 adults with an overall margin of sampling error of plus or minus 3 percentage points. If that sample includes, for example, 200 Hispanics, the overall results based on the subsample of Latinos is plus or minus 6.9 percentage points—the MOSE for a sample of 200.

Even huge samples of 10,000 or more theoretically contain some error due to sampling. In the early stages of the 2016 presidential campaign small margins of error threatened

to make big news. Republican presidential hopefuls were invited by the television networks to participate in televised “prime time” presidential debates based on their ranking in an aggregation of recent polls. But as some statistically savvy GOP campaign strategists noted, the tiny differences in levels of support that separated some of the candidates—frequently measured in the tenths of a percentage point—was well within the aggregate sample’s MOSE. That made it impossible to specify with confidence the rank order of all the candidates and is one reason why AAPOR cautioned against using polls to pick debate participants.

Other factors can affect the margin of sampling error. For example, how the sample was selected and the extent to which a sample was statistically adjusted or “weighted” to bring it into line with known characteristics of the target population can affect MOSE. These “design effects” can substantially increase the margin of sampling error beyond the simple estimates reflected in the chart. These effects typically are factored into the overall margin of sampling error reported in most high-quality surveys.

The Credibility Interval

As online surveys and other types of nonprobability-based polls play a larger role in survey research, another statistic has emerged that is often confused with MOSE. It is called the “credibility interval” and it is used to measure the theoretical accuracy of nonprobability surveys. While both MOSE and a credibility interval are expressed in the familiar “plus-or-minus” language, they are very different.

The credibility interval relies on assumptions that may be difficult to validate, and the results may be sensitive to these assumptions. So while the adoption of the credibility interval may be appropriate for non-probability samples such as opt-in online polls, the underlying error associated with such polls remains a concern. Consequently, [AAPOR urges caution](#) when using credibility intervals or otherwise interpreting results from electoral polls using non-probability online panels.

One final note: There is no such thing as a measurable overall margin of error for a poll -- surveys are subject to other errors, ranging from how well the questions were designed and asked to how well the interviews were conducted. Good pollsters and researchers do everything in their power to minimize these other possible sources of errors, but they cannot be measured so one can never know the precise amount of error associated with any poll finding.