

Election polls are 95% confident but only 60% accurate

Aditya Kotak & Don A. Moore

Method & Analysis

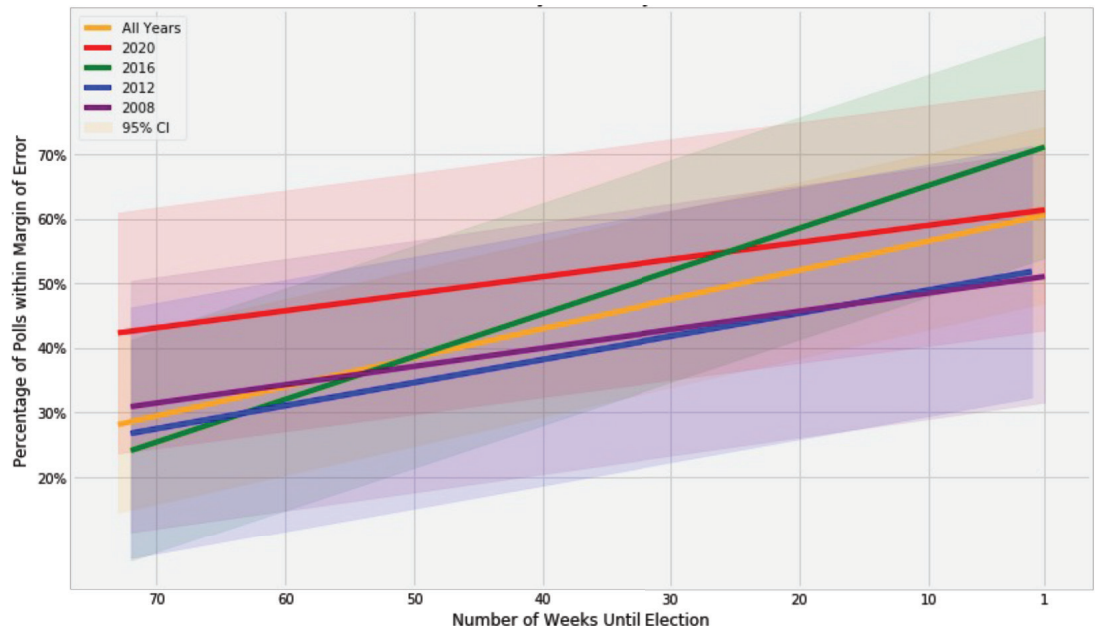
More Detail on Study 1

Our preregistration, data, and code are available at <https://osf.io/65za7/>.

Our preregistered plan called for three analyses, all of which test the accuracy of polling results as a function of time to the election: (a) an analysis regressing the probability of the forecast likelihood of the election outcome on days between the poll and the election, (b) an analysis regressing the deviations of the poll outcome from the election result on days between the poll and the election, and (c) a logistic regression testing the rate at which polls “hit” (by getting the election result inside the margin of error) as a function of days to the election. The preregistered analyses are posted at <https://osf.io/keswd/>.

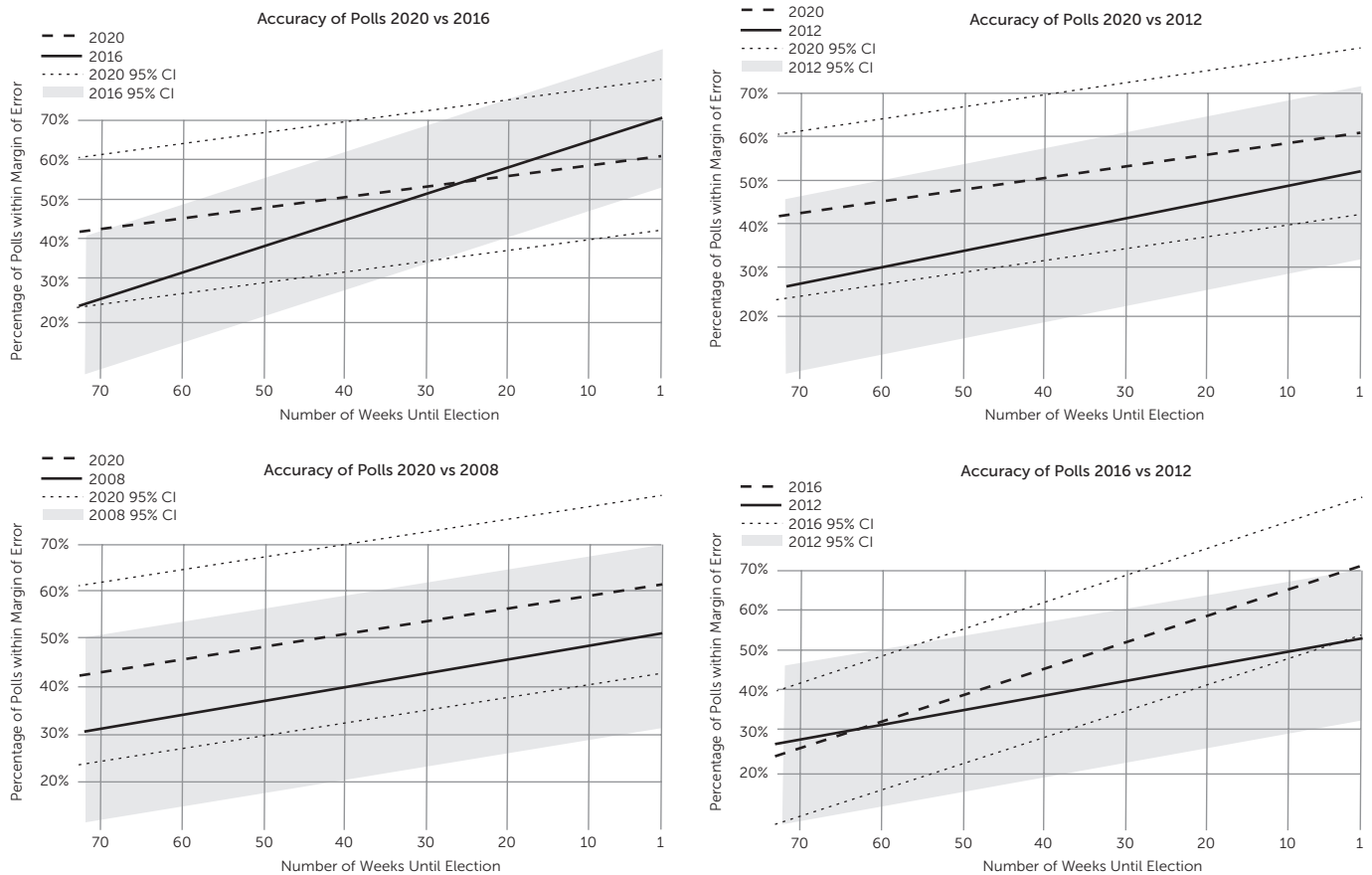
Find our preregistration plan to assess responses to seven different poll reporting styles at <https://osf.io/9qhmf>.

Figure S1. Summary of poll accuracy by year



Note. CI = confidence interval. Shaded regions indicate the 95% CI centered around a given year.

Figure S2. Selected comparison of poll accuracy by year



Additional Data Displays Showing the Accuracy of Polls by Year

Because errors within a poll are correlated, we conducted our key analysis again at the level of the poll. We used two approaches. First, we identified a hit for the poll if the average hit rate of the poll's candidate-level forecasts was greater than 50%. This approach produced results consistent with the conclusions presented in the article. Figure S3 shows the accuracy of poll-level forecasts versus candidate-level forecasts of our results.

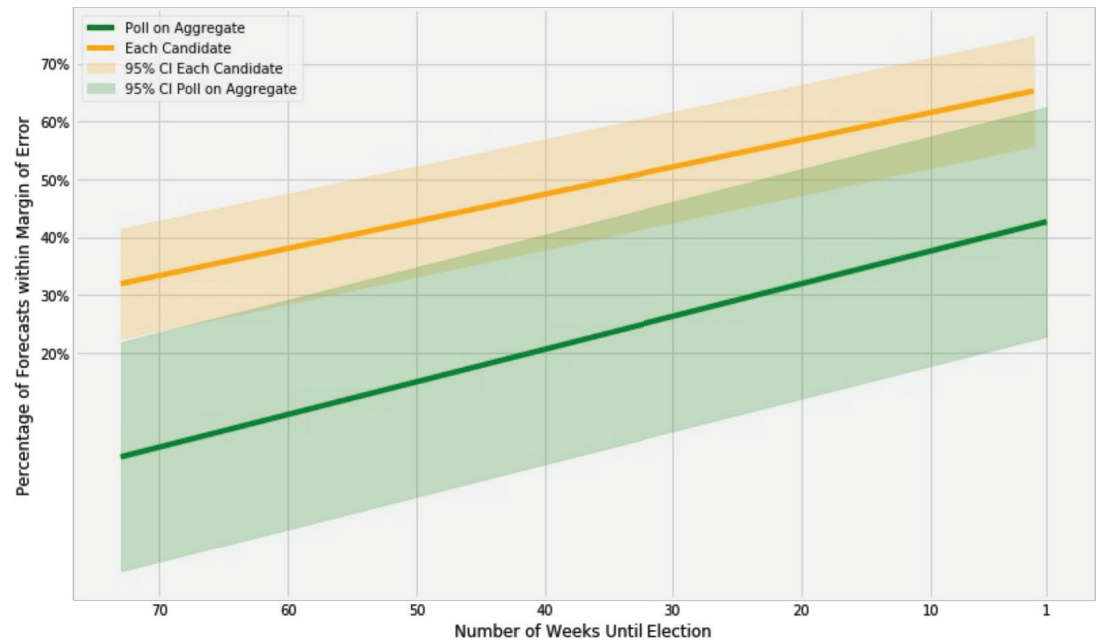
Our second approach averaged the hits within a poll and assigned that as the average hit rate for the entire poll. For example, if a poll includes four candidates and one (and only one) of them achieves a vote share inside the margin of error,

then the poll's hit rate is 25%. The result is also consistent with our original analysis, as illustrated in Figure S4.

Our next test investigated whether the caucus format by which Iowa selects candidates contributes more error than do other election formats. We conducted an analysis distinguishing between caucuses and other elections. The results do not reveal a significant difference between these two, $t(55) = 1.1$, $p = .27$. The results appear in Figure S5.

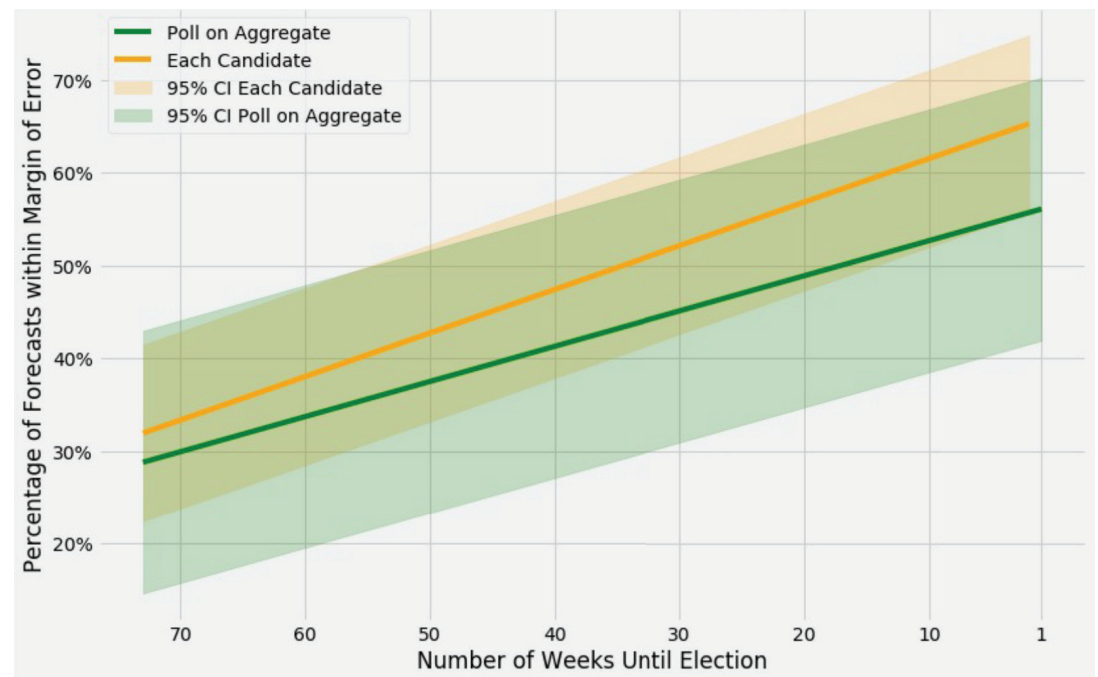
Similarly concerned about the degree to which state-level primary elections might differ from general elections, we conducted an analysis comparing them. Again, the results, shown in Figure S6, do not reveal significant differences, $t(68) = 1.6$, $p = .11$.

Figure S3. Accuracy of polls versus time until election: Aggregate poll hits & misses



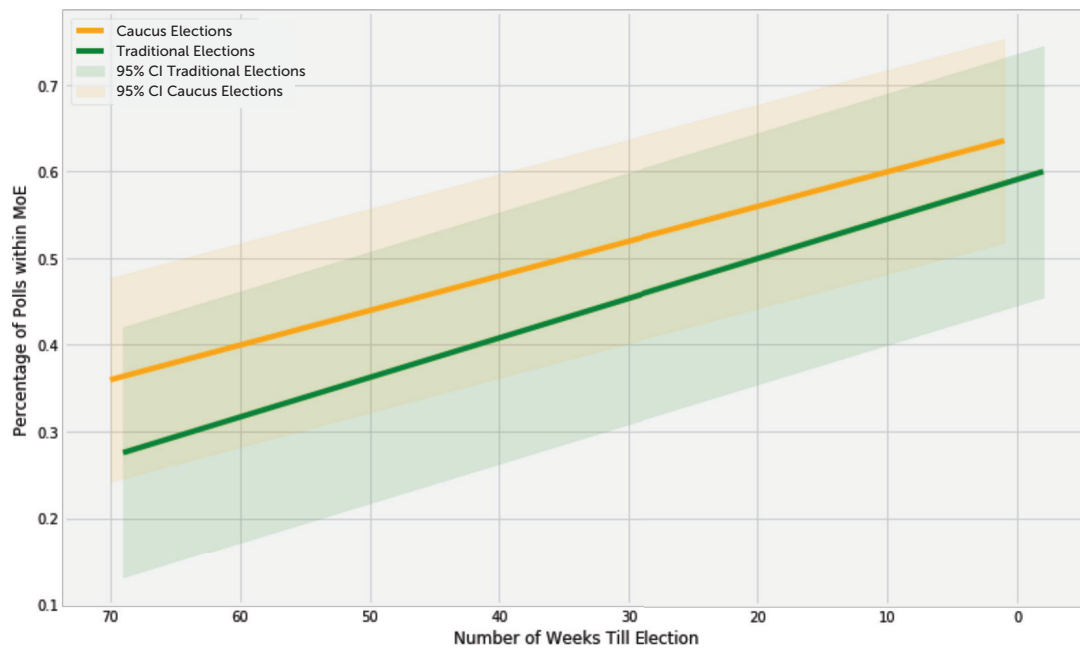
Note. CI = confidence interval. Yellow color refers to candidate-level analysis. Green color refers to poll-level analysis.

Figure S4. Accuracy of polls versus time until election: Average poll hit rate



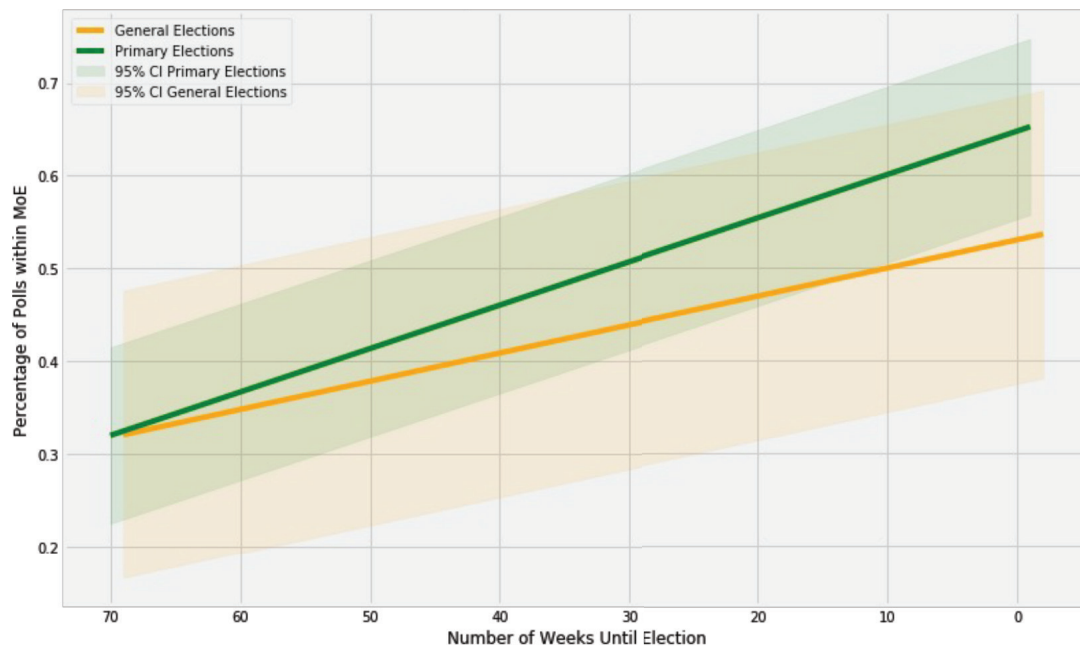
Note. CI = confidence interval. Yellow color refers to candidate-level analysis. Green color refers to poll-level analysis.

Figure S5. Accuracy of polls versus time until election:
Caucuses versus traditional primaries



Note. CI = confidence interval; MoE = margin of error.

Figure S6. Accuracy of polls versus time until election:
Primaries versus general elections



Note. CI = confidence interval; MoE = margin of error.

More Details About Study 2 Results

As predicted, average confidence ($M = 59.9\%$) exceeded average accuracy ($M = 49.7\%$), $t(216) = 6.81$, $p < 10^{-11}$. But confidence in a poll's predictive accuracy varies by time horizon and reporting style. A 3 (time horizon) \times 7 (poll reporting style) mixed analysis of variance (ANOVA) with repeated measures on the second factor reveals a main effect of time horizon, $F(2, 1491) = 30.1$, $p < 10^{-13}$, and a main effect for reporting style, $F(6, 1491) = 7.6$, $p < 10^{-7}$. Their interaction is not significant, $F(12, 1491) = 1.38$, $p = .17$. Please note that we erroneously preregistered only the one-way ANOVA examining the within-subjects effect of reporting style. This within-subjects main effect is significant, but the omnibus ANOVA we report is more useful.

Although we found that reported confidence exceeded historical accuracy across all reporting styles, this was not universally true for all time horizons. Style 2 represents one of the most common styles: a point estimate, plus or minus a margin of error. For reporting Style 2, confidence exceeds accuracy only for the one-year time horizon, $t(74) = 8.5$, $p < 10^{-11}$, and not for three months, $t(71) = 1.1$, $p = .29$, or for one day, $t(69) = 2.39$, $p = .02$. Figure 5 displays results that underscore both people's excessive faith in polls' predictive accuracy and the challenge of correcting it. Providing more information about polls' poor record of accuracy in reporting Styles 6 and 7 failed to bring confidence in line with historical accuracy.