

## **What Exit Pollsters Can Learn From Empirical Social Science**

*Subjecting pollsters to the scrutiny, transparency and demands of modern-day scientific research will help inject some much-needed scientific rigour into exit polls that are otherwise largely an exercise in showbiz.*

As the election results began trickling in on June 4, it became apparent that the exit polls had got it wrong and how. The charges against the pollsters ranged from being casual, fly-by-night operators to being obsequious agents of the ruling dispensation, with the accusation that their modus operandi lacked transparency tucked somewhere in that spectrum. Soon after, much drama unfolded in the television studios. Some claimed they had got the vote share correct, but the vote share-to-seat conversion was always tricky. Others grudgingly admitted they had called all the states, save a few ones, right. One even broke down in the face of persistent questioning.

Astonishingly, the question that never gets asked is: given that the results are out and the new government has been formed, can the polling agencies release the exit poll data in public domain? Can they now inform the public about their methodology for arriving at the exit poll numbers? At the very minimum, can they let the public know the sample characteristics based on which the predictions were made? What reasons can there be for not releasing the data, now that even the negotiations for cabinet berths are over and ministries assigned?

### **Why the data should be released**

There are several benefits to releasing the data. First, independent researchers, with little skin in the game, can examine the nature of the sample used for prediction and test to what extent the sample characteristics deviate from the population in terms of caste, religion, gender, socio-economic status, etc.

Second, Nobel Prize-winning economist Roland Coase mordantly remarked, "If you torture the data long enough, it will confess to anything." Releasing the data now could help researchers 'torture' it in this Coasian sense and examine to what extent the predicted numbers would differ if another methodology were used. This is particularly important after the allegations that the exit poll numbers are fudged. After all, it is much easier to fudge the numbers using a 'favourable' method than outrightly doctoring the data.

Third, it will help researchers understand whether human judgements have been used to arrive at the headline predictions over and above the purely data-driven predictions.

While releasing the data and methodology now is a starting point, the credibility crisis of exit and opinion polling is unlikely to be mitigated by this alone. If pollsters claim that theirs is a

scientific process, they must adopt the best practices of related sciences. Pre-registration and pre-analysis plans are two principal means through which sciences and social sciences have sought to enhance the transparency of research practices.

### **How does science protect itself against half-truths?**

Imagine a researcher interested in testing whether an FM radio-based campaign can increase voter turnout in Mumbai, where the turnout rate in the 2024 parliamentary election stood at about 52%. The researcher designs a detailed research protocol and collects data from all six constituencies of Mumbai. When the turnout data arrives, our researcher realises the campaign had no effect in Mumbai South and Mumbai South Central, but had a small impact on the remaining four constituencies. She decides to write a report claiming that the radio-based campaign positively affected voter turnout, by excluding the data from Mumbai South and Mumbai South Central. While this does not represent an outright research fraud, it does qualify as an egregious half-truth, which can sometimes be as dangerous as a lie, if not more.

How does science protect itself against such half-truths? First, journals demand that researchers pre-register their research hypotheses. Researchers describe their research protocol in a pre-specified and open online platform and 'lock' it before they begin to collect the data. At the very minimum, researchers must write down the hypotheses they intend to test, commit to sample size, and declare some well-defined sample characteristics they wish to achieve. For the example above, this will mean proposing to test the hypothesis that the radio campaign does affect voter turnout in Mumbai, setting up a sampling framework to collect data from all the constituencies of Mumbai, and declaring what proportion of women, Dalits, Muslims and other communities will be sampled. Importantly, this protocol is 'locked' and made public before data collection begins.

Second, researchers are often required to commit to a pre-analysis plan. They write down the methodology for analysing the data and post it publicly. The pre-analysis plan ties a researcher's hand by forcing her to analyse the data only using the pre-specified methodology and, therefore, does not allow her to fish for results by 'torturing' the data using different methods. While none of these commitment devices prevents the more hideous forms of research fraud, they have come to occupy an essential element in the toolkit of empirical social scientists, particularly those that use the experimental approach.

Inspired by these, the Press Council of India, for lack of a more appropriate body, may create a platform where polling agencies can publicly post their plan before they begin to collect data. This plan should include the proposed sample size and how the sample would look in terms of gender, caste, religion, and various other demographics. They should specify which methodology will be employed to convert the vote share to the number of seats. Finally, the

polling agencies should commit to releasing the data after the results are out. Fulfilling these essential criteria would then result in an accredited exit poll.

Psephology may well be an inexact science, perhaps more so in a country as diverse as India. However, surely, it is not the only form of inexact science. One may argue that the inexact science of predicting the polls has far greater impact on people than other inexact sciences whose reach remains largely academic. The stock market volatility in the aftermath of the election results and the consequent loss of investor wealth is a case in point. Therefore, it stands to reason that the level of scrutiny and transparency that ought to be demanded of the pollsters should far exceed the level of scrutiny on, let's say, a university professor's research on India's state of democracy. Subjecting themselves to the scrutiny, transparency and demands of modern-day scientific research will help inject some much-needed scientific rigour into what is otherwise largely an exercise in showbiz.

*Ritwik Banerjee is an associate professor of economics at IIM Bangalore.*