

Melissaberry/shutterstock.com

Downloaded from academic.oup.com/jrssig/article/23/6/39/8993950 by guest on 20 November 2026

Bad stats

A regular series exploring slip-ups, snafus and salutary lessons from the world of statistics

Have Australian men *really* stopped reading?

Marius Radean examines a recent news story to highlight the pitfalls of overlooking or misusing confidence intervals

When reporting research findings or policy evidence, it is not easy to find the right balance between accessibility and thoroughness. Too much detail may overwhelm and put off the reader, whereas glancing over technical details leaves the door open to misinterpretation. While the key may be a blending between the two approaches, halfway solutions can give the trappings of scientific arguments without providing a definitive answer, or, more worryingly, may suggest an incorrect one. Drawing invalid conclusions from research findings could be due to a number of factors, including overlooking the uncertainty around estimated results or misinterpreting measures of uncertainty when these are reported. Here uncertainty refers to how an estimate might differ from the unknown, true value.

The decision on how to engage with the uncertainty in public opinion surveys – or whether to acknowledge it at all – is a prime example of the balancing act between providing too much and not enough information. Since pollsters never reach out to every single individual in a country, survey results are based on the responses from a sample or a subset of the population. Because not everyone is included, those responses are only an estimate of the entire population's views. By definition, an estimate is an approximation of the true value and comes with a certain degree of uncertainty.

In practice, there are some who focus on a single value of the estimate, which is taken as absolute. As an example, it is not uncommon to have election commentators report only the average value of opinion polls (e.g., candidate A leads in the polls with 51% of the voting intention). This approach can leave the reader with a misguided sense of certainty about the strength of the finding and also about the interpretation of what the result means. Other analysts are more candid about the uncertainty around various estimates,



Marius Radean is a senior lecturer in the Government Department at the University of Essex, UK. His research focuses on democratic representation and accountability, democratic survival, and quantitative methodology.

and report a range of likely values (e.g., the confidence interval, or CI). For the polling example, the 95% CI of candidate A's electoral support could be 48% to 54%. This means that if we were to repeat the sampling many times (i.e., conduct numerous opinion polls) and built a confidence interval each time, about 95% of those intervals would contain the true value, within the interval. The latter approach is more informative, but also prone to mistakes when assessing different estimates. A common mistake is to directly compare estimates based on overlapping or non-overlapping CIs.[1]

This is not just a theoretical or purely technical issue. It can have very real consequences. To illustrate, let us examine two media articles reporting on a survey about the reading habits of Australians. Both reports use the same data from the Australian Bureau of Statistics (ABS; tinyurl.com/rwwa4ket), but they differ in terms of whether the results are reported with or without CIs.

Making the headline, an ABC News article claims the data show that "Australians ... are reading less than ever before" and that "females, on average, were reading more than males across all age groups" (tinyurl.com/yszxxchj). In support of these claims, the article presents a graph (reproduced in Figure 1) with the proportion of males and females who participate in reading across five generations. Notably, these proportions are not accompanied by CIs or any other indicator of uncertainty. Seemingly supporting the article's first claim that reading has become less popular over time, the generational difference in reading habits is substantial. More specifically, the reading rate of Generation Z (15–24 years) is about 11.2%, whereas the rate of the interwar generation (75 years and over) stands at 47.4%. The respective rates are the proportion of people in a generational group who spent at least 5 minutes on reading in a day. With respect to the second claim about the difference between females and males, the percentage of men reading is consistently lower across all five generations.

Highlighting several problems with this analysis, a follow-up review in the *Guardian* newspaper argues the data warrants neither pessimistic conclusion (tinyurl.com/24w6yp7b). First, the data precludes one from claiming a decline in the reading habits *across generations*. While the average reading rate of Generation Z is more than four times lower than the rate of the interwar generation, we cannot rule out that reading habits differ by stage of life. Put differently, it may be the case that by the time the people in Generation Z reach the age of 75, their reading rate would be similar to that currently observed for the interwar generation.

Second, absent accompanying indicators of uncertainty, it is hasty to conclude that the reading rates *between males and females* are statistically different. Since the reported results are based on a sample from the entire population, we do not know with certainty the actual proportion of males and females who read regularly. When accounting for this uncertainty, the ostensible difference may not be significant.

Both criticisms are sound. While the graph in Figure 1 has a strong visual impact, the information conveyed is actually limited.

Addressing the second problem, the *Guardian* article reports an updated graph where the proportion of people who engage in reading is accompanied by CIs. This graph is reproduced in Figure 2. The new analysis concludes that, once we account for the uncertainty around the compared percentages, the perceived difference in the reading rates between men and women is not statistically significant (with the exception of Generation X). The revised conclusion is based on the fact that the CIs of the compared percentages for men and women overlap. This criterion is embedded in the new graph, with the message

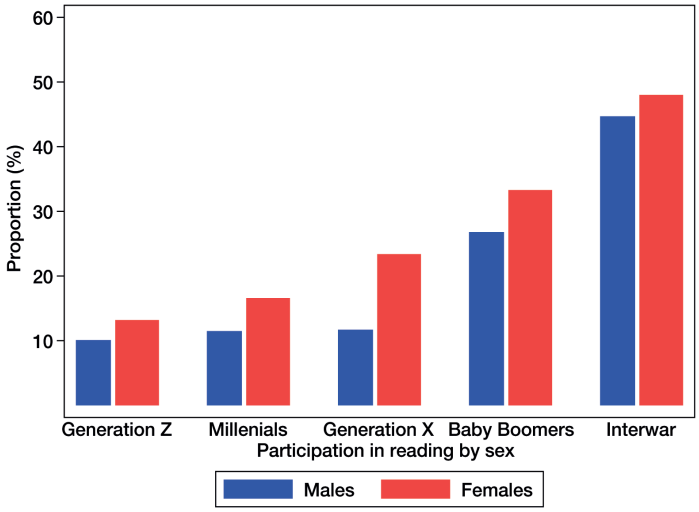

Figure 1: Research estimates with no uncertainty measurement.

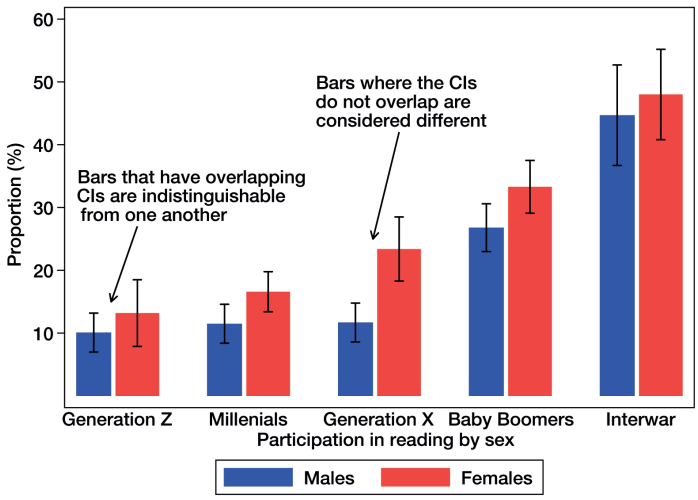

Figure 2: Research estimates with accompanying CIs.

Downloaded from academic.oup.com/jrssig/article/23/6/39/8993950 by guest on 20 November 2026

in the leftmost text box reading: "Bars that have overlapping CIs are indistinguishable from one another".

Thus, the two articles reach different conclusions, even though they use the same data. One may be inclined to give more credence to the *Guardian* analysis because of the more rigorous approach to reporting uncertain data. But relying on the CI overlap to determine whether two estimates are statistically distinct is not the touted solution. While two values whose CIs do not overlap are statistically different, it is not the case that two values with overlapping CIs are necessarily indistinguishable from one another. When the CIs overlap, the estimates may or may not be statistically different. Unless we conduct additional tests, we simply cannot tell. Since this is the case here, it is theoretically possible that men do read significantly less than women. This in turn would support the conclusion from the ABC News article, even though its reason for reaching that conclusion was incorrect (i.e., the nominal difference between male and female reading rates). As a side note, I requested the raw survey data from the ABS to conduct the standard difference in (reading) proportions test, but I have not heard back. For now, the mystery of whether there is a reading crisis affecting mainly young men continues.

To some extent, the misuse of CIs in media accounts is understandable. In fact, there is evidence from multiple academic disciplines that even trained analysts (i.e., authors of peer-reviewed articles) use the CI overlap method to examine significance of differences.[1–3] This being the case, it is unsurprising that non-technical journalists would make that mistake. But this does not mitigate the far-reaching impact of inaccurate reporting. Given that broad audiences usually get their information from print or electronic media, "[t]he journalists and editors who decide what analysis warrants coverage and *how to report it* therefore have considerable power to influence societal perspectives".[4]

So, what lessons can we learn? First, we should always consider the uncertainty around individual point estimates. Otherwise, we project confidence about data that is inherently uncertain. Second, we should not use the CI overlap to determine whether two estimates are statistically different. While visually appealing, this is not an appropriate method. This approach can reveal *some* cases when the compared estimates are statistically distinct, but not all (estimates with overlapping CIs may still be statistically different). And it can never clarify whether two estimates are statistically indistinguishable. Thus, this approach is uninformative at best and misleading at worse. To complicate matters, simply adding a warning or disclaimer is unlikely to be effective. Practically, the "use of a single visualization with overlapping and nonoverlapping confidence intervals leads many to draw such [wrong] conclusions, despite the best efforts of statisticians toward preventing users from reaching such conclusions".[5]

While two values whose CIs do not overlap are statistically different, it is not the case that two values with overlapping CIs are necessarily indistinguishable from one another

There are two main alternatives to judge the significance of differences. First, we may use specialised uncertainty intervals that are designed to compare alternative values, such as the significance of differences interval (SDI).[3] When using this approach, overlapping SDIs indicate that the compared estimates are not statistically different, whereas the non-overlapping SDIs indicate that the estimates are statistically distinct. Second, one may compute and examine the difference between the compared values with its CI. If the associated CI does not include 0, we may conclude that the difference in estimates is statistically significant. Put differently, this would mean there is evidence that the compared estimates are distinct.

To finish on a lighter note, before adopting any of the above suggestions, you should check that it does not contravene the law in your country. If the warning made you smile, you would be forgiven for thinking that the presentation of research findings is determined only by the analyst's preferences or the established practice. In some contexts, this is a matter of law.[4] In the US, for instance, the Congressional Budget Act of 1974 mandates the Congressional Budget Office to provide *point predictions* about the likely impact on the budget of proposed legislation. As a result, for the last fifty years, the estimates of the budgetary implications of various bills have been presented to Congress without any measure of uncertainty. Of course, five- or ten-year forecasts about the financial costs of a piece of legislation are highly uncertain and tentative at best.

In contrast, the UK legislation requires similar impact assessments for draft legislation to be accompanied by "a range for [the] costs and benefits estimates".[4]

The distinct legal requirements may reasonably lead to different voting and policy outcomes. When only the raw score is available, budgetary affordability and comparisons between competing pieces of legislation can only be made based on the bills' *nominal* price tag. When CIs are reported alongside the point predictions, there is additional information to consider. For example, is an individual bill still affordable if we consider the upper limit of the CI range? Or, are two budgetary estimates statistically different once we account for the uncertainty around the point estimates? Since it has real-world implications, the decision to report and correctly interpret CIs is more than a statistical quirk. ■

Downloaded from academic.oup.com/jrssig/article/23/6/39/8993950 by guest on 20 November 2026

References

1. Schenker, N. and Gentleman, J. F. (2001) On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, **55**(3), 182–186.

2. Belia, S., Fidler, F., Williams, J. and Cumming, G. (2005) Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, **10**(4),389–396.

3. Radean, M. (2023) The significance of differences interval: Assessing the statistical and substantive difference between two quantities of interest. *Journal of Politics*, **85**(3), 969–983.

4. Manski, C. F. (2013). *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press.

5. Wright, T., Klein, M. and Wieczorek, J. (2019) A primer on visualizations for comparing populations, including the issue of overlapping confidence intervals. *The American Statistician*, **73**(2), 165–178.