

BAD BAR CHARTS ARE PERVASIVE ACROSS BIOLOGY

How scientific ‘shorthand’ could be introducing data distortions into published papers. **By Amanda Heidt**

Bar charts are ubiquitous in the life-science literature, yet a study suggests that they’re often used in ways that can misrepresent research findings. The preprint¹, which was first posted on bioRxiv in September and has yet to be peer reviewed, found that in a collection of nearly 3,400 papers from 2023 that included at least one bar chart, almost one-third distorted the data in some way, highlighting a need for increased data literacy among scientists and for a system of checks throughout the writing and publishing processes.

“Data are getting more complex all the time, and data literacy doesn’t always keep pace with that. Even still, I do think it’s surprising how common some of these mistakes are,” says Rebecca Goldin, a mathematician at George Mason University in Fairfax, Virginia. “It’s good to see more attention being paid to how we visualize our work.”

To quantify the amount of data distortion she often saw in published bar charts, Markita Landry, a nanobiotechnology researcher at the University of California, Berkeley, and her colleague Teng-Jui Lin analysed 3,387 life-science papers published in 15 journals last year, including several *Nature* and *Science* journals, as well as *Cell*, *Bioengineering & Translational Medicine* and *ACS Nano*.

The pair found that 88% of the papers contained at least one bar chart; of those papers, 29% had a bar chart with some form of data distortion. The most common types of distortion included failing to start the y axis at zero, as well as mistakes that related to logarithmic axes. The former often inflates the difference between two values to make small disparities look larger, and the latter can minimize differences because our brains are prone to perceive scales as linear. Papers with multiple co-authors were most likely to include these distortions, the team found. “More collaborators may make it easier for mistakes to fall through the cracks,” Landry explains.

There’s no suggestion that these distortions represent intentional attempts to deceive, Goldin says, but they might make it harder for non-specialists to understand the studies. However, in some cases, there are defensible reasons for making these choices, which Landry calls a form of scientific shorthand. A spokesperson for *Nature*, which the study



flagged as having a high proportion of distorted figures in its articles, echoes this sentiment: “Ensuring the optimal presentation of data to aid interpretation and understanding may sometimes justify non-zero starts for y axes as well as the use of log axes.”

Cautious approach

Helena Jambor, a data-visualization scientist at the University of Applied Sciences of the Grisons in Chur, Switzerland, says that although the study taps into real issues with how science is communicated, some of its conclusions are a bit alarmist, and the paper could have benefited from input from data scientists. “These authors are correctly pointing out that many people could misunderstand what is being stated,” she says. “But that does not mean that it was necessarily incorrect or that two scientists talking about the data would misunderstand one another.”

A study² published in 2021, however, found that roughly one in five readers do misunderstand bar charts, often by interpreting the top of the bar as the upper limit of a range, rather than the mean. Zen Faulkes, an independent biologist who authored a November preprint on graph readability³, says that a lack of proper labelling often exacerbates this issue. In a survey of 200 bar charts and box plots shared at the 2020 meeting of the Animal Behavior Society, he found that nearly 90% did not have enough information on the slide to interpret them. “Conferences don’t necessarily enforce

the same rigours that a journal might, but at the very least, a bar graph should show a mean value and error bars and identify the measure being illustrated by the error bars,” he says.

In response to Jambor’s feedback and other comments, Landry says that she and Lin plan to expand the work from a commentary into a full manuscript, to include more background and discussion. The two also plan to investigate the decision-making processes that lead to these distortions. Hoping to establish a set of best practices, the pair’s preprint includes a flowchart to help to identify what type of figure is best for various forms of data, as well as recommendations for journal editors and peer reviewers assessing work for clarity.

“As researchers, we’re not only responsible for making our work accurate for researcher peers, but also have to be cautious of the use of our work by the general public – can it be misinterpreted?” Landry says. “The burden of preventing misinterpretation is not on the audience, but on the ones who make the graphics.”

Amanda Heidt is a freelance writer and editor in Moab, Utah.

1. Lin, T.-J. & Landry, M. P. Preprint at bioRxiv <https://doi.org/10.1101/2024.09.20.609464> (2024).
2. Kerns, S. H. & Wilmer, J. B. *J. Vision* **21**, 17 (2021).
3. Faulkes, Z. Preprint at bioRxiv <https://doi.org/10.1101/2024.11.08.622642> (2024).