# Side-stepping safeguards – Data journalists are doing science now

*An aspect of the media landscape that has been highlighted by the COVID-19 pandemic has been the increasing role of media organisations in presenting and undertaking their own, often complex, data analyses. In this cross-post* **Irineo Cabreros**, *discusses the the potential risks posed by this trend and what journalists could learn from academic researchers to safeguard the quality of their work.*

---

News stories are increasingly told through data. Witness the Covid-19 time series that decorate the homepages of every major news outlet; the red and blue heat maps of polling predictions that dominate the runup to elections; the splashy, interactive plots that dance across the screen.

As a statistician who handles data for a living, I welcome this change. News now speaks my favourite language, and the general public is developing a healthy appetite for data, too.

But many major news outlets are no longer just visualizing data, they are analysing it in ever more sophisticated ways. For example, at the height of the second wave of Covid-19 cases in the United States, The New York Times ran a piece declaring that surging case numbers were not attributable to increased testing rates, despite President Trump's claims to the contrary. The thrust of The Times' argument was summarized by a series of plots that showed the actual rise in Covid-19 cases far outpacing what would be expected from increased testing alone. These weren't simple visualizations; they involved assumptions and mathematical computations, and they provided the cornerstone for the article's conclusion. The plots themselves weren't sourced from an academic study (although the author on the by-line of the piece is a computer science Ph.D. student); they were produced through "an analysis by The New York Times."

> many major news outlets are no longer just visualizing data, they are analysing it in ever more sophisticated ways

The Times article was by no means an anomaly. News outlets have asserted, on the basis of in-house data analyses, that Covid-19 has killed nearly half a million more people than official records report; that Black and minority populations are overrepresented in the Covid-19 death toll; and that social distancing will usually outperform attempted quarantine. That last item, produced by The Washington Post and buoyed by in-house computer simulations, was the most read article in the history of the publication's website, according to Washington Post media reporter Paul Farhi.

In my mind, a fine line has been crossed. Gone are the days when science journalism was like sports journalism, where the action was watched from the press box and simply conveyed. News outlets have stepped onto the field. They are doing the science themselves.

Of course, there's nothing sacred about academic institutions. Even world-class scientists can make mistakes that culminate in faulty conclusions. These errors sometimes slip past the peer-review process, the main tool of quality control for scientific journals. Occasionally, even high-profile findings are retracted after being cited hundreds of times. As of this writing, more than 100 articles on Covid-19 have been retracted, according to Retraction Watch. But even in light of the imperfections of traditional research, the environment of a newsroom strikes me as a particularly dangerous place to produce science.

Date originally posted: 2021-04-26

Permalink: https://blogs.lse.ac.uk/impactofsocialsciences/2021/04/26/side-stepping-safeguards-data-journalists-are-doing-science-now/

Blog homepage: https://blogs.lse.ac.uk/impactofsocialsciences/

For one thing, the pace of production in a newsroom is blistering. During the summer of 2018, I had the opportunity to write on the science desk at Slate Magazine while pursuing my Ph.D. in applied mathematics. Four years into my graduate studies at the time, I had published one scientific article, and another was grinding its way through the peer-review process. At Slate, I published nine articles in 10 weeks, and even that was a snail's pace relative to the professional journalists around me.

Another clear difference between newsroom science and academic science lies in the training of the people who perform it. Academic papers are typically authored by teams of career scientists and their apprentices. By contrast, although some news analyses are co-written by journalists trained in quantitative methods, many include not a single by-line from a contributor with a higher degree in a scientific field.

I have no problem with research being produced faster than the clunky gears of the scientific apparatus can churn. Nor do I think that research performed by intelligent individuals outside of the scientific community is necessarily wrong. But those two ingredients, combined with the massive platform of a major news outlet, make for a potentially dangerous concoction. They create a situation where the quickest and least-vetted science also wields the megaphone.

This can be particularly problematic when newsroom research butts heads with academic research. For example, in 2016 ProPublica published an influential article claiming that a risk algorithm used to predict criminal recidivism was marred by racial bias. The piece prompted a robust academic response. Some scientists pointed out methodological shortcomings of the original ProPublica research. For instance, one study found that ProPublica made incorrect assumptions about how age was incorporated into the proprietary algorithm. When age was appropriately accounted for, the case for algorithmic racial bias dwindled. Other critiques were more subtle. For instance, researchers argued that the racial bias ProPublica identified was impossible to remove without introducing another equally pernicious type of racial bias.

Had the ProPublica investigation taken place in a traditional academic setting, there is no guarantee that all of these shortcomings would have been identified and pre-empted. But they likely wouldn't have reached such a wide audience. And the ProPublica argument has stuck in the collective consciousness in a way that the scientific rebuttals — published in academic outlets with comparatively small readerships — have not. It continues to be widely cited today, both in news articles and in scholarly work.

Date originally posted: 2021-04-26

Permalink: https://blogs.lse.ac.uk/impactofsocialsciences/2021/04/26/side-stepping-safeguards-data-journalists-are-doing-science-now/

Blog homepage: https://blogs.lse.ac.uk/impactofsocialsciences/

To be sure, scientific analysis has a place in the newsroom, encouraging the public to see issues of societal importance through a scientific lens. But how can it be done at the speed that the modern news cycle requires without jeopardizing the accuracy that science demands?

There are a couple of ways to fit this square peg into the round hole. The first is to increase transparency. Science produced by news outlets should be completely reproducible. This means that an article should link to the data and code used to produce its analysis, allowing anyone to look directly under the hood. If serious errors are identified, the original article should be amended or retracted, and notice of the amendment or retraction should be visible to all future visitors of the article. The danger of speedy publication can be mitigated by speedy correction.

> Efforts at reproducibility have been made by some outlets, but they have been inconsistent and often incomplete.

Efforts at reproducibility have been made by some outlets, but they have been inconsistent and often incomplete. For instance a New York Times article on "the pandemic's hidden death toll" includes a link to the relevant dataset, but not to any of the code used to analyse it. Meanwhile, another article at the Times provides almost no documentation at all. To ProPublica's credit, the outlet did publish extensive documentation of its methods for the 2016 recidivism investigation, allowing scientists to easily critique the work. And while ProPublica published a direct response to some of the critiques, the publication has not made any substantive changes to the original article.

Newsrooms should also implement a process of peer review. Although data stories may be fact-checked, and reporters often consult scientists for input on in-house analyses, the ultimate gatekeeper is typically the editor, who is seldom a scientist. A more rigorous peer-review process for data-driven stories would be to send the story to an appropriate outside expert for review, make adjustments to the story based on feedback until the reviewer affirms the analysis, and then name the reviewer as co-author on the by-line. That last step would both incentivize timely reviews and hold reviewers accountable for their work.

Science and journalism are kindred pursuits. Both seek to gather information, understand it, and convey its meaning to others. It makes sense that journalists would be drawn to do science, and such mergers have often resulted in truly compelling work. The ProPublica article on recidivism, despite its shortcomings, was arguably responsible for energizing an entire field of academic research and galvanizing scientists around a topic of real societal import. Similarly, The Covid Tracking Project, initiated by The Atlantic, collected national Covid-19 data that governments and researchers relied on throughout the pandemic.

Newsrooms *are* doing science, and I hope that trend continues. But before they shout their bold conclusions from the rooftops, they need to make sure it's done right.

---

*This article was originally published on Undark. Read the original article.*

*Note: This review gives the views of the author, and not the position of the LSE Impact Blog, or of the London School of Economics.*

*Image Credit: Ono Kosuki, via Pexels.*

---

Date originally posted: 2021-04-26

Permalink: https://blogs.lse.ac.uk/impactofsocialsciences/2021/04/26/side-stepping-safeguards-data-journalists-are-doing-science-now/

Blog homepage: https://blogs.lse.ac.uk/impactofsocialsciences/