

## Low-quality papers are surging by exploiting public data sets and AI

Paper mills are also likely contributing to “false discoveries”

14 MAY 2025 • 4:15 PM ET • BY CATHLEEN O'GRADY



SARA GIRONI CARNEVALE

SHARE:

LISTEN TO THIS ARTICLE

▼

Last year, Matt Spick began to notice oddly similar papers flooding in for peer review at *Scientific Reports*, where he is an associate editor. He smelled a rat. The papers all drew on a publicly available U.S. data set: the National Health and Nutrition Examination Survey (NHANES), which through health exams, blood tests, and interviews has collected dietary information and other health-related measurements from more than 130,000 people. “I was getting so many nearly identical papers—one a day.

## Science Adviser

Get Science's award-winning newsletter with the latest news, commentary, and research, free to your inbox daily.

[SUBSCRIBE](#)

[NOT NOW](#)

X

Such free data sources allow almost anyone to take a known research method and swap in new variables to create fresh “findings” in a kind of “research Mad Libs,” says Reese Richardson, a metascientist at Northwestern University who was not involved with the work. Other researchers have found similar “explosions” in a range of topics, he says, including various kinds of genetic studies as well as analyses of bibliometrics or gender disparities in different scientific disciplines.

## SIGN UP FOR THE AWARD-WINNING SCIENCEADVISER NEWSLETTER

The latest news, commentary, and research, free to your inbox daily

[SIGN UP >](#)

The NHANES papers Spick was receiving all followed the same formula: They chose a health condition, an environmental or physiological factor that could be associated with it, and a population group—perhaps looking at the link between vitamin D levels and depression in men over age 65, or poor dental health and diabetes in women between the ages of 18 and 45. “It felt like every possible combination was being worked through by someone,” Spick says.

To get a better understanding of how prevalent these studies are, he and his team searched two major databases of scientific papers, PubMed and Scopus, for studies using NHANES data that looked at single associations. They found 341 of these papers published in 147 journals, including *Scientific Reports*, *BMC Public Health*, and *BMJ Open*. Between 2014 and 2021, an average of four such papers were published per year—but a rapid increase kicked off in 2022, with 190 papers published in 2024 up to October, when the researchers did their search. The rise far outstripped the growth in health studies using large data sets generally, the authors report, suggesting some additional factor underlying the swell of NHANES studies.

The timing points to the widespread availability of AI chatbots such as ChatGPT that can generate readable text from simple prompts and uploaded information. They may have been used to rephrase the same basic NHANES findings endlessly to avoid plagiarism detection, says Jennifer Byrne, a molecular biologist at the University of Sydney who peer reviewed the *PLOS Biology* paper. It’s not possible to conclude with certainty that paper mills—[commercial entities that sell authorship on fraudulent or low-quality papers](#)—produced the papers, she says, but the “timing and scale of the increase make you think there has to be some kind of coordination behind this.”

Many of the more recent NHANES studies selectively analyzed portions of its data set without a clear rationale—for example, authors limited their analysis to certain years, or certain ages of people in the survey. That suggests the authors were on the hunt for statistically significant results to generate easy publications, Spick says. But fishing for results in such a huge data set is bound to come up with many false positive findings. When the team took a closer look at the 28 NHANES studies that had explored depression, they found that only 13 of the results survived a statistical adjustment that corrects for the risk of finding false positives.

Spick and his team think their analysis may drastically underestimate the problem. Their search only looked for NHANES studies that fit the formula Spick had been seeing, but a broader search finds that papers using the data set increased from 4926 in 2023 to 7876 in 2024. And other big health data sets—such as the Global Burden of Disease study—may also be vulnerable, Spick says. These data sets make it easy for researchers to interact with their information using coding languages such as Python or R, but this also makes them easy to exploit: His team was easily able to write code that could pull all the data from NHANES and “chug through the combinations” of diseases and health variables. The “industrialization” of low-quality research overwhelms the literature with useless findings, Spick says. “Honestly, I got really hopping mad about it.”

These papers reflect broad problems in both scientific publishing and how research is rewarded, Richardson says. “All of the publishers named in the article accepted fees, likely on the order of \$1000 each, to publish this junk,” he notes. (Open-access journals, including *PLOS Biology*, generally charge author fees to make papers freely available.) And researchers are incentivized to publish more papers, rather than higher quality papers, in order to advance in their careers, Richardson adds. The problem, he warns, “will only get worse unless we radically restructure incentives around scientific publication.”

doi: 10.1126/science.zgawnij

# Science Adviser

Get Science's award-winning newsletter with the latest news, commentary, and research, free to your inbox daily.

[SUBSCRIBE](#)

[NOT NOW](#)



## ABOUT THE AUTHOR



Cathleen O'Grady [✉](#) [𝕏](#)

Author

## MORE FROM NEWS



15 MAY 2025

**['About as close to aliens as we'll ever get.'](#)**  
**[Can AI crack animal language?](#)**

BY CHRISTA LESTÉ-LASSERRE



15 MAY 2025

**[Massive DNA sequencing effort reveals how colonization shaped Brazil's genetic diversity](#)**

BY RODRIGO PÉREZ ORTEGA



15 MAY 2025

**[Indian police are trying to 'read minds' of suspects, over neuroscientists' objections](#)**

BY JONATHAN MOENS

[VIEW MORE >](#)

Got a tip for Science's news department?

[CONNECT >](#)

### Sign up for ScienceAdviser

Get Science's award-winning newsletter with the latest news, commentary, and research, free to your inbox daily.

[SUBSCRIBE >](#)

## SCIENCEINSIDER •



14 MAY 2025 | BY JOCELYN KAISER, MEREDITH WADMAN

**[Kennedy, Trump's health chief, confronts criticism and praise from U.S. lawmakers](#)**



14 MAY 2025 | BY CATHLEEN O'GRADY

**[Low-quality papers are surging by exploiting public data sets and AI](#)**



13 MAY 2025 | BY JEFFREY MERVIS

**[NSF board member resigns, saying Trump's policies have harmed the agency](#)**

# ScienceAdviser

Get Science's award-winning newsletter with the latest news, commentary, and research, free to your inbox daily.

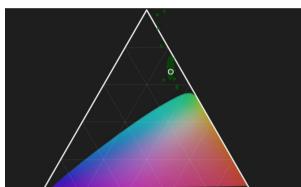
[SUBSCRIBE](#)

[NOT NOW](#)



**SIFTER** ⓘ

25 APR 2025 | BY CHRISTA LESTÉ-LASSERRE

**Ice age couture: Ancient people tailored form-fitting clothing from reindeer pelts**

18 APR 2025 | BY CHRISTIE WILCOX

**Superprecise lasers show people a whole new color**

10 APR 2025 | BY JENNA AHART

**Medieval manuscripts were bound in baby seal fur**

11 MAR 2025 | BY JOCELYN KAISER

**Gene editing fixes DNA—and treats incurable lung and liver disease**

VIEW MORE &gt;

**RECOMMENDED**

6 APR 2017 | BY DALMEET SINGH CHAWLA

**Now free: citation data from 14 million papers, and more might come**

27 JUL 2017 | BY LINDSAY MCKENZIE

**Sci-Hub's cache of pirated papers is so big, subscription journals are doomed, data analyst suggests**

NEWS OF THE WEEK | SEPTEMBER 2004

**NIH Proposes 6-Month Public Access to Papers**

POLICY FORUM | JANUARY 2019

**Toward unrestricted use of public genomic data**

SPONSORED WEBINAR | SCIENCE AND LIFE | 22 MAY 2025

**Redefining Connection: How young people are shaping the future of the rare disease community through technology and innovation**e  
OpticsScience  
SignalingScience  
Translational  
Medicine

Science

Science  
AdvancesScience  
Immuno**Science Adviser**

Get Science's award-winning newsletter with the latest news, commentary, and research, free to your inbox daily.

SUBSCRIBE

NOT NOW



## NEWS

[All News](#)  
[ScienceInsider](#)  
[News Features](#)  
[Subscribe to News from Science](#)  
[News from Science FAQ](#)  
[About News from Science](#)  
[Donate to News](#)

## COMMENTARY

[Opinion](#)  
[Analysis](#)  
[Blogs](#)

## AUTHORS & REVIEWERS

[Information for Authors](#)  
[Information for Reviewers](#)

## ADVERTISERS

[Advertising Kits](#)  
[Custom Publishing Info](#)  
[Post a Job](#)

## ABOUT US

[Leadership](#)  
[Work at AAAS](#)  
[Prizes and Awards](#)

## CAREERS

[Careers Articles](#)  
[Find Jobs](#)  
[Employer Hubs](#)

## JOURNALS

[Science](#)  
[Science Advances](#)  
[Science Immunology](#)  
[Science Robotics](#)  
[Science Signaling](#)  
[Science Translational Medicine](#)  
[Science Partner Journals](#)

## LIBRARIANS

[Manage Your Institutional Subscription](#)  
[Library Admin Portal](#)  
[Request a Quote](#)  
[Librarian FAQs](#)

## RELATED SITES

[AAAS.org](#)  
[AAAS Communities](#)  
[EurekAlert!](#)  
[Science in the Classroom](#)

## HELP

[FAQs](#)  
[Access and Subscriptions](#)  
[Order a Single Issue](#)  
[Reprints and Permissions](#)  
[TOC Alerts and RSS Feeds](#)  
[Contact Us](#)

