

Subscribe (/subscribe/) Archive (/archive/) Q

de (<https://www.horizonte-magazin.ch/2021/12/02/gleiche-information-andere-schlussfolgerung/>) fr (<https://www.revue-horizons.ch/2021/12/02/memes-informations-conclusions-differentes/>)

X (https://www.twitter.com/horizons_en) @ (https://www.instagram.com/horizons_mag)

Read for Content

Horizons

The Swiss Research Magazine
 (<https://www.horizons-mag.ch/>)



SCIENCE STUDIES

Same data, different conclusions

Asserting that a study has shown something suggests that one acts with an evidence-based attitude. But studies often contradict each other. And a new study shows that this even happens when those studies analyse the exact the same data. We've taken a closer look at the problem.

Edwin Cartlidge (<https://www.horizons-mag.ch/author/edwincartlidge/>),
 02/12/2021 (<https://www.horizons-mag.ch/2021/12/02/>)

X ([/#x](#)) in ([/#linkedin](#))

+ (<https://www.addtoany.com/share?url=https%3A%2F%2Fwww.horizons-mag.ch%2F2021%2F12%2F02%2Fsame-data-different-conclusions%2F&title=Same%20data%2C%20different%20conclusions>)



Is Denis Zakaria here being discriminated against on account of his skin colour? Football can provide us with a large volume of raw data, but it is not clear what we should actually be analysing. In fact, this image can also serve to confirm our own latent prejudices – because the red card in the game between Switzerland and Spain was actually for Remo Freuler, who is not shown here. | Image: Anton Vaganov/Pool/AFP/Keystone

Give several groups of scientists the same set of data and ask them to investigate a number of hypotheses. If those scientists are scrupulous and careful, all of the groups should come to very similar conclusions. At least, that is how we normally think of the properly functioning scientific method.

Not so, according to new research from a large collaboration of social scientists, computer scientists and statisticians published in June 2021 in the journal *Organizational Behavior and Human Decision Processes*. The study provided independent analysts with nearly four million words from an online academic forum and asked them to investigate two apparently quite straightforward hypotheses regarding how gender and academic status influence contributions to the forum's discussions. The results were striking, with wild variations in both the analysts' approach to scrutinising the data and in their conclusions – in some cases arriving at diametrically opposite responses.

“With analytical steps, all the choices made need to be abundantly clear”.

– Abraham Bernstein

That outcome is not a bolt from the blue. Numerous studies in recent years have shown that many research results in fields from sociology to medicine are impossible to reproduce. But much of the ensuing discussion has focused on the pitfalls of searching for statistical significance within noisy data as well as bias in the publication of results. The new study reveals that variation in research method is also a big problem.

For the collaboration member Abraham Bernstein, a computer scientist at the University of Zurich, this means that scientists need to publish not only the data that underlie their research but also the precise analytical steps they take. “There is too much interpretational flexibility by what is meant by ‘we ran x test’”, he says. “All the choices need to be made abundantly clear”.

Comparing analysts

The new study is an example of what is known as crowdsourcing – the recruitment of numerous independent research groups, usually online, to analyse the same dataset. One early example of this, published in 2018, asked analysts to study data from football leagues to see if there was any correlation between a player’s skin colour and how many red cards they receive. That led to a wide range of different conclusions – with most analysts finding a statistically significant but not particularly marked effect while some found no correlation at all.

However, while that project stipulated a specific correlation – one involving red cards, rather than, say, offside rulings – the latest work left analysts free to define the relevant variables themselves. The analysts’ job was to assess two hypotheses regarding comments submitted over the course of nearly two decades to the website edge.org by over 700 contributors, 128 of whom were female. One, that ‘A woman’s tendency to participate actively in a conversation correlates positively with the number of females in the discussion’. And two, that ‘Higher status participants are more verbose than are lower status participants’.

The study was carried out by an international collaboration and coordinated by Martin Schweinsberg, a psychologist at the European School of Management and Technology in Berlin. It involved 19 analysts, taken from an initial pool of 49, who used a specially designed website called Dataexplained to record and explain their analytical steps – both those they ended up using and those they rejected.

The results of the exercise show just how varied analytical approaches can be. In terms of method, individuals employed a wide range of statistical techniques and an even wider range of variables. When it came to encoding an individual’s ‘status’, for example, they used, among other parameters, academic job title, possession of a PhD, number of citations and a number reflecting how many highly cited articles a person has published, known as the h-index.

“There are problems inherent in the freedom enjoyed by researchers in exploratory analyses”.

– Leonhard Held

That variety in method was then reflected in the wide range of outcomes. About two-thirds of analysts concluded that women did tend to contribute more in the presence of other women, but over a fifth came to the opposite conclusion. For the question of status, the division was even starker – with 27 percent in favour of the hypothesis and 20 percent opposed, while the remaining analysts came up with statistically insignificant results.

Too much discretion?

Some outside the collaboration argue that caution is needed in interpreting these results. Leonhard Held of the University of Zurich points out that – for all their freedom in defining variables and statistical techniques – the analysts did not themselves choose a particular online forum to answer the research questions in hand. He reckons their exclusion from this process could have influenced results, given that sample size can affect statistical significance. He also questions how realistic the analyses were as attempts to generalise about group dynamics, given the use of that one forum.

Nevertheless, Held welcomes the study, saying it “clearly illustrates the inherent problems of the degree of freedom enjoyed by researchers in exploratory analyses”. Anna Dreber Almenberg of the Stockholm School of Economics in Sweden is also enthusiastic, describing the research as “super important” in trying to further enhance reproducibility. She reckons it highlights a limitation of pre-registration – the requirement that researchers specify their method and statistical tests before they collect and analyse their data. While pre-registration can improve the reliability of results, she points out that it can’t dictate what specific analysis to employ.

Indeed, Schweinsberg and colleagues argue that variations in analysis pose “a more fundamental challenge for scholarship” than either p-hacking or peeking at data before they are tested. These other problems, they say, can be addressed either by pre-registration or by employing a blind analysis – a researcher in the latter case, they point out, being unable either consciously or unconsciously to choose an analysis that will yield a desired signal.

In contrast, they say, natural variations in analysts’ knowledge, beliefs and interpretations will lead to different research results even when the individuals involved act transparently and in good faith. “Subjective choices and their consequences, often based on prior theoretical assumptions, may be an inextricable aspect of the scientific process”, they write.

Make all choices explicit

One project designed to make such variation explicit is ‘Many Paths’. Run by a group of five academics from Germany, the Netherlands and Switzerland, it aims to expose ‘the messy middle’ of the research process by bringing practitioners from different disciplines to work together on joint projects. It has started the ball rolling by soliciting political scientists, philosophers, psychologists and others to discuss the age-old question: ‘Does power corrupt?’ – having so far enrolled around 40 experts.

Many Paths uses a tool called Hypergraph to document research in a modular, step-by-step way. As project member Martin Götz of the University of Zurich explains, the idea is to “dissect the classic scientific paper” so that individuals can work on the parts that most interest them, be that theory, data gathering or meta-analyses, for example. The ultimate aim, he says, is to “replace the current publishing system” and what he sees as its incentives for headline-grabbing rather than robust research.

“We should avoid nihilism”.

– John Ioannidis

Although many experts agree on the need to improve reproducibility, some caution against too much pessimism. One of them is John Ioannidis, a medical scientist and epidemiologist at Stanford University, USA. He argues that, although there will be some research where any result is possible, in most cases multiple analyses will reveal some results to be more plausible than others. “We should avoid nihilism”, he says.

Bernstein too remains upbeat, maintaining that the research process will be robust as long as scientists are clear about the choices they make during the course of their work. The important thing, he says, “is that those choices can be made explicit and so held to scrutiny”.

M. Schweinsberg et al.: Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes* (2021)
(<https://doi.org/10.1016/j.obhdp.2021.02.003>)

CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>)

✕ [\(/#x\)](#) in [\(/#linkedin\)](#)

✚ <https://www.addtoany.com/share?url=https%3A%2F%2Fwww.horizons-mag.ch%2F2021%2F12%2F02%2Fsame-data-different-conclusions%2F&title=Same%20data%2C%20different%20conclusions>)

Subscribe to Newsletter

Send

Subscribe to Horizons