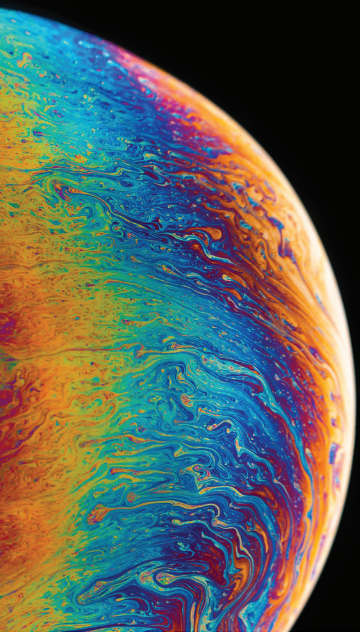


REPRODUCIBILITY NOTES

Different worlds
Confirmatory versus exploratory research

Simon Schwab and **Leonhard Held** explain the differences between confirmatory and exploratory research and the dangers of confusing the two concepts

Christopher Columbus discovered America, but he did not set out to do so. Sailing off to the west, he aimed to find a new route to India, China and Japan, but arrived at a new land mass instead. Was Columbus's mission a success or a failure? Many would call it a success. However, if his voyage of discovery was thought of as a piece of *confirmatory research*, it would have been considered a failure. As Stephen Senn put it: "A medical statistician will not accept that Columbus discovered America because he said he was looking for India in the trial plan."[1]

Senn's point is that, in confirmatory research – such as clinical trials – you cannot go looking for something ("If I sail west, will I find a new route to India?") and then change your mind about the goal of the research when you find something else ("Look! A new continent!"). Doing so opens the door to a questionable research practice known as HARKing (Hypothesising After the Results are Known).[2] HARKing is like cheating in a game of darts, where the target is drawn around a dart only *after* it has been thrown. In science, such behaviour can produce a large number of false positives. To avoid these problems, it is important to understand the distinction between confirmatory and exploratory research.

Confirmatory research

Generally, confirmatory research starts with a clear hypothesis and then collects data that may or may not support that hypothesis. For example, one might start with the hypothesis that a new drug or therapy is a more effective treatment than an existing drug or therapy. So, how is HARKing avoided?

Prior to conducting a clinical trial, a trial plan (a study protocol with all the important details) is written and sent to a clinical trials registry; for example, the US ClinicalTrials.gov or the EU Clinical Trials Register (EU-CTR). After registration, everything is set in stone. In principle, it will not be possible to alter the hypothesis or "cherry-pick" findings that are more interesting to the researchers while ignoring other findings that do not align with the researchers' expectations. The published results can simply be compared to the registered study protocol to see if researchers changed their minds and analysed something else or tried to hide some unexpected findings. In 2004, leading medical journals stated that registration in a public trials registry is a requirement for publication.[3]

Simon Schwab is a postdoctoral fellow in the Center for Reproducible Science at the University of Zurich (UZH).

Leonhard Held is a professor of biostatistics at UZH and director of the Center for Reproducible Science at UZH.

When the two concepts get confused, and the two worlds collide, the fallout can be disastrous

However, not all research is strictly confirmatory. There is also exploratory research, which deals with problems that have not yet been studied at length.

Exploratory research

The word "research" is a descendant of "search", so for the scientific "explorers" out there, following a strict plan may not match their ideal of what research should be. And they may point to the many great discoveries that were the result of happy accidents, such as the discovery of penicillin by Alexander Fleming.

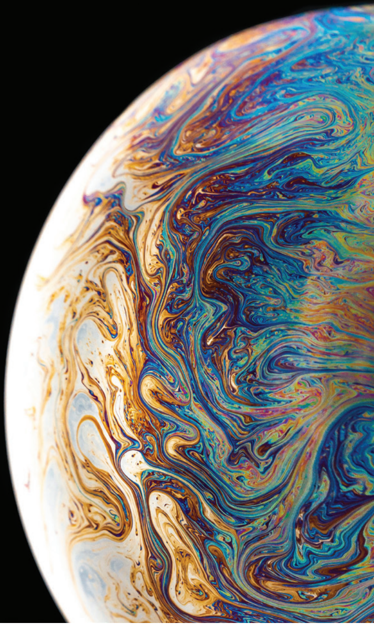
If your goal is to find the unexpected, exploratory research is far more suitable than confirmatory research (for a comparison, see the table on page 9). Unlike the confirmatory kind, exploratory research may start with no hypothesis, or only loosely formulated hypotheses, and fewer details about the design (including required study size). But that leaves exploratory research open to potentially unscrupulous manipulation.

First, there is more room for "fudging".[4] A researcher may, sometimes without realising it, fudge a hypothesis to fit the data. And fudging can also lead to HARKing, particularly in situations where a hypothesis is generated *post hoc* but the researcher presents it as an *a priori* hypothesis.

An early example of fudging is the concept of epicycles by the astronomer Claudius Ptolemy.[4] In the Ptolemaic system, the Earth was believed to be

© 2020 The Royal Statistical Society

Exploratory research	Confirmatory research
No hypothesis required/hypothesis can be vague	Clear hypothesis required
Generate new hypothesis from data	Test *a priori* hypothesis with new data
High sensitivity desired, i.e. minimising the risk of false negatives	High specificity desired, i.e. minimising the risk of false positives
Suitable for making new discoveries and finding the unexpected	Suitable for establishing strong evidence and confirming the expected
For example: Testing of new compounds in mice	For example: Assessing the efficacy of a drug in humans



Daniel Olah/Unsplash.com

at the centre of the universe. However, the observed data – the motion of the planets – did not fit the model very well. Therefore, Ptolemy introduced epicycles into his model: planets move on a circle (the epicycle) that in turn is moving on another, much larger, circle around the Earth. The fudging was that Ptolemy, in retrospect, adjusted his theory to accommodate the observed data.

The second danger of exploratory research is that a researcher may only selectively report what "works" and ignore what does not, meaning that large effects and statistically significant results are more likely to be reported, whether spurious or not. Such results may then be used to support an apparently coherent hypothesis even if it is false.

Best of both worlds

The most effective research exploits both worlds of exploration and confirmation. Exploratory research is used to generate hypotheses, and confirmatory research to test them. For example, after studying the orbit of a bright comet he had observed in 1682, Edmond Halley noticed it followed a similar orbit to comets recorded in 1531 and 1607. This led him to hypothesise that all three were in fact the same object. This reflects the exploratory part. Halley went on to predict that the comet would appear again in 1758. This was the confirmatory part. Halley's prediction proved correct and the comet was named after him. Unfortunately, he had died 16 years earlier and so did not see the return of the comet.

Embracing more confirmatory research in traditionally exploratory fields can enhance the transfer of research findings into practice – a process known as "translation". For example, promising drugs or therapies in preclinical animal studies are often unsuccessful when tested in humans.[5] An explanation is that exploratory preclinical studies only reveal potentially promising directions, while human trials, in contrast, are confirmatory, and aim to kill off those directions that are really dead ends. It is essential to distinguish between the two types of research, and it is the combination of the two that creates useful evidence.

While the two worlds can work well together, scientists are on dangerous ground when they merge and confuse confirmatory and exploratory research. For example, researchers might present exploratory results as confirmatory in order to increase the probability of publication. But what they will have done here is engage in the practice of HARKing, and this increases the risk that a false-positive finding will make its way into the scientific literature, and decreases the overall likelihood of the result being reproducible, replicable, or generalisable.[6]

As a solution, the widespread adoption of preregistration beyond clinical trials has been suggested[7] as a way to improve the credibility of research – not least by compelling authors to be clear about whether their study is exploratory or confirmatory.

In conclusion, it is essential to distinguish between exploratory and confirmatory research, and they are equally important to the scientific enterprise. Finding the questions to ask is at least as crucial as answering them, if not more so.[8] But when the two concepts get confused, and the two worlds collide, the fallout can be disastrous. ■

References

1. Senn, S. (2007) *Statistical Issues in Drug Development*. Chichester: John Wiley & Sons.

2. Kerr, N. L. (1998) HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, **2**(3), 196–217.

3. De Angelis, C., Drazen, J. M., Frizelle, F. A. *et al.* (2004) Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, **351**, 1250–1251.

4. Lipton, P. (2005) Testing hypotheses: Prediction and prejudice. *Science*, **307**, 219–221.

5. Kimmelman, J., Mogil, J. S. and Dirnagl, U. (2014) Distinguishing between exploratory and confirmatory preclinical research will improve translation. *PLoS Biology*, **12**, e1001863.

6. National Academies of Sciences, Engineering, and Medicine (2019) *Reproducibility and Replicability in Science*. Washington, DC: National Academies Press.

7. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. and Mellor, D. T. (2018) The preregistration revolution. *Proceedings of the National Academy of Sciences of the USA*, **115**, 2600–2606.

8. Tukey, J. W. (1980) We need both exploratory and confirmatory. *American Statistician*, **34**, 23–25.