Specification Curve Analysis Protocol

S. Stanley Young[1] and Warren B. Kindzierski[2]

[1] CGStat, Raleigh, NC, USA
[2] Independent Consultant, St Albert, Alberta, Canada
Correspondence: Stanley Young, Email: genetree@bellsouth.net

Background
Even if a research claim in a scientific study is accepted as true, it may, in fact, be false. There are many contradictory claims in the published science literature, and some are taken as true and canonized (1,2). These situations can be considered research puzzles. Which of the two opposing claims is correct? Maybe neither is correct. How these puzzles come about and what might be done to see how puzzle parts fit together is an important current question.

A specification curve analysis proceeds by computing many analyses to get an estimated effect with confidence limits for each analysis and then plotting those results in order from smallest to largest.

This research protocol provides parts/steps for identifying and hopefully resolving false claims. Claim A or B could be true, or both could be true but under different circumstances.

Part 1
The first step is to identify an important question in research and a complex data set that can be used to "answer" the question. Two examples in literature – one in the field of social sciences (Breznau, 3) and the other in the field of nutrition (Wang, 4) are noted here. The data sets are very complex in both cases, with many possible ways to compute an analysis and answer the question. This calls to mind numerous studies that have highlighted a growing realization of many research claims being routinely mistaken, they are wrong (e.g., 5−7).

Part 2
The next step is to examine a key study that supports a research claim of interest. Secure the data set used in the key study. Note that researchers should make their data set freely available to other researchers to allow independent reproduction of the original research findings. Science works by research or experiments that can be repeated. When they are repeated, they must give the same answer and if research or an experiment does not replicate, something is wrong (5). From the data compute multiple analyses and plot the results.

Often in health research, personal identity protection is offered as a reason to withhold data sets. In almost all cases, an analysis data set can be constructed that protects personal identity. One simple method is to assign cases a random number to protect identity. In other cases, microaggregation can be used to protect identity (8).

Part 3

One side of a research claim, usually the positive side, often predominates in literature. For example, consumption of red meat is positively associated with greater numbers of some types of cancer. Yet there may be a few studies in literature which do not support this claim. Why might there be many studies supporting a claim and only a few rejecting it?

Journal publications are an important, if not a principal factor for career advancement in academia. One side of a claim may offer an easier path to publications or funding, e.g., a positive claim. Once a positive claim is made in literature, there are opportunities for studying conditions supporting the claim. A negative claim, on the other hand, disagrees with the positive claim and – if the particular field of science operated honestly – it would be a showstopper for that line of research (9).

Another idea is there is natural selection of bad science in literature supporting a positive claim (1). Students of researchers using poor methods will go out into the academic world. The expected result is the natural selection of bad science (1). It can happen that a false claim becomes accepted as true, canonized (1,2).

Comments: Why can there be both positive and negative cause–effect outcomes in published literature? The answer lies here. In 2022, seventy-three research teams were given the same complex data set and the same research question to investigate (3). Their findings varied from significant negative to significant positive cause–effect associations and everything in between. Analysis choices for each team were recorded but these choices explained little of the analysis variability, less than 5%. One analysis or another of a complex data set can lead to any result with, at this point in time, no rhyme or reason other than randomness as an explanation.

Part 4

Specification curve analysis tests and visualizes associations between two variables of interest across many different model specifications and analytic decisions. This method can be used to examine a research question with contradictory claims in literature. The method would involve the following steps:

1. For the data set of interest, list the number of analysis options: outcomes, predictors, and covariates. There could be many thousands of possible analyses.
2. Select representative analyses to compute using expert opinion (3,4), clustering (SAS JMP), or space-filling design (SAS JMP).
3. Compute the selected analyses.
4. Display the results for the analysis options in terms of effects (3, 4) or a volcano plot of p-values (plot the -Log10(p-value) versus the observed effect).

Part 5 Interpretation

If results of the specification curve analysis cover a wide range of effects, i.e., consistent with Breznau (3) and Wang (4), then one can take the position that “any result is possible” and therefore any result could be wrong. Here we note that such a finding may speak more to the overall challenge and limitation of current scientific methods to examine weak associations between variables (10). This research area is relatively new, so specification curve analysis could be misguided.

How and why are irreproducible (false) research claims canonized (taken as fact)? How: natural selection of bad science (1,6). Why: the main reason is positive studies greatly outnumber negative studies, which come from inadequate or compromised peer review and editorial decisions (7,11). As stated previously, if a particular field of science operated honestly, one valid negative study is supposed to disprove any number of positive studies (9), yet there is publication bias that hides negative studies (7,12). Pointing to Breznau in his opening remarks, Briggs notes that much science is failing to replicate (13).

These results are consistent with the warning of Hayek in his 1974 Prize Lecture (14): "If man is not to do more harm than good in his efforts to improve the social order, he will have to learn that in this, as in all other fields where essential complexity of an organized kind prevails, he cannot acquire the full knowledge which would make mastery of the events possible."

Hayek was addressing social complexity; we can add data complexity to his warning. The truth is, it is unknown if many current canonized claims in literature based on the analysis of complex data sets are correct. If any answer is possible with a seemingly plausible analysis, then there is a problem.

A possible way forward is to let science be the judge. Data sets used in any complex study making a research claim important to society should be freely available for other researchers to allow independent reproduction of the original research findings. This can be done with microaggregation (8) to protect personal identity by aggregating data over time or space. Running a specification curve analysis (3, 4) on large, complex data sets could shed useful light on the situation and help put all the puzzle pieces together.

References

1. Smaldino, P.E., McElreath, R. 2016. The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384. https://doi.org/10.1098/rsos.160384

2. Nissen, S.B., Magidson, T., Gross, K., Bergstrom, C.T. 2016. Publication bias and the canonization of false facts. *eLife*, 5, e21451. https://doi.org/10.7554/elife.21451

3. Breznau, N., Rinke, E.M., Wuttke, A., Nguyen, H.H.V., Adem, M., Adriaans, J., et al. 2022. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences*, 119, e2203150119. https://doi.org/10.1073/pnas.2203150119

4. Wang, Y., Pitre, T., Wallach, J.D., de Souza, R.J., Jassal, T., Bier, D., Patel, C.J., Zeraatkarg, D. 2024. Grilling the data: application of specification curve analysis to red meat and all-cause mortality. *Journal of Clinical Epidemiology*, 168, 111278. https://doi.org/10.1016/j.jclinepi.2024.111278

5. Young, S.S., Karr, A. 2011. Deming, data and observational studies. *Significance*, 8(3), 116–20. https://doi.org/10.1111/j.1740-9713.2011.00506.x

6. Gelman, A., Loken, E. 2014. The statistical crisis in science. *American Scientist*, 102(6), 460–465. https://doi.org/10.1080/13854046.2016.1277557

7. Randall, D., Welser, C. 2018. *The Irreproducibility Crisis of Modern Science: Causes, Consequences, and the Road to Reform*. New York, NY: National Association of Scholars. https://www.nas.org/reports/the-irreproducibility-crisis-of-modern-science

8. Yan, Y., Eyeleko, A.H., Mahmood, A., Li, J., Dong, Z., Xu, F. 2022. Privacy preserving dynamic data release against synonymous linkage based on microaggregation. *Scientific Reports*, 12(1), 2352. https://doi.org/10.1038/s41598-022-06182-y

9. Feynman, R. 1964. *Messenger Lecture on "The Character of Physical Law,"* Cornell University, Ithaca, New York. November 9, 1964. https://jamesclear.com/great-speeches/seeking-new-laws-by-richard-feynman

10. Young, S.S., Kindzierski, W.B. 2019. Evaluation of a meta-analysis of air quality and heart attacks, a case study. *Critical Reviews in Toxicology*, 49(1), 85–94. https://doi.org/10.1080/10408444.2019.1576587

11. Schachtman, N. 2011. *Misplaced Reliance On Peer Review to Separate Valid Science From Nonsense*. Tortini, August 14, 2011. https://schachtmanlaw.com/2011/08/14/misplaced-reliance-on-peer-review-to-separate-valid-science-from-nonsense/

12. Ritchie, S. 2020. *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth*. New York, NY: Henry Holt and Company.

13. Briggs, W.M. 2024. Science is Broken Lecture. https://x.com/FamedCelebrity/status/1797668902304141743

14. Hayek, F. 1974. *The Pretence of Knowledge. Lecture to the memory of Alfred Nobel*, December 11, 1974. https://www.nobelprize.org/prizes/economic-sciences/1974/hayek/lecture/