

# Statistical Modeling, Causal Inference, and Social Science

---

## Handy statistical lexicon

Posted by Andrew on 24 May 2009, 10:29 pm

These are all important methods and concepts related to statistics that are not as well known as they should be. I hope that by giving them names, we will make the ideas more accessible to people:

Mister P: Multilevel regression and poststratification.

The Secret Weapon: Fitting a statistical model repeatedly on several different datasets and then displaying all these estimates together.

The Superplot: Line plot of estimates in an interaction, with circles showing group sizes and a line showing the regression of the aggregate averages.

The Folk Theorem: When you have computational problems, often there's a problem with your model.

The Pinch-Hitter Syndrome: People whose job it is to do just one thing are not always so good at that one thing.

Weakly Informative Priors: What you should be doing when you think you want to use noninformative priors.

P-values and U-values: They're different.

Conservatism: In statistics, the desire to use methods that have been used before.

The Backseat Driver Principle: Even if the advice or criticism is annoying, it makes sense to listen.

WWJD: What I think of when I'm stuck on an applied statistics problem.

Theoretical and Applied Statisticians, how to tell them apart: A theoretical statistician calls the data  $x$ , an applied statistician says  $y$ .

The Fallacy of the One-Sided Bet: Pascal's wager, lottery tickets, and the rest.

Alabama First: Howard Wainer's term for the common error of plotting in alphabetical order rather than based on some more informative variable.

The USA Today Fallacy: Counting all states (or countries) equally, forgetting that many more people live in larger jurisdictions, and so you're ignoring millions and millions of Californians if you give their state the same space you give Montana and Delaware.

Second-Order Availability Bias: Generalizing from correlations you see in your personal experience to correlations in the population.

The "All Else Equal" Fallacy: Assuming that everything else is held constant, even when it's not gonna be.

The Self-Cleaning Oven: A good package should contain the means of its own testing.

The Taxonomy of Confusion: What to do when you're stuck.

The Blessing of Dimensionality: It's good to have more data, even if you label this additional information as "dimensions" rather than "data points."

Scaffolding: Understanding your model by comparing it to related models.

Ockhamite Tendencies: The irritating habit of trying to get other people to use oversimplified models.

Bayesian: A statistician who uses Bayesian inference for all problems even when it is inappropriate. I am a Bayesian statistician myself.

Multiple Comparisons: Generally not an issue if you're doing things right but can be a big problem if you sloppily model hierarchical structures non-hierarchically.

Taking a Model Too Seriously: Really just another way of not taking it seriously at all.

God is in Every Leaf of Every Tree: No problem is too small or too trivial if we really do something about it.

As They Say in the Stagecoach Business: Remove the padding from the seats and you get a bumpy ride.

Story Time: When the numbers are put to bed, the stories come out.

The Foxhole Fallacy: There are no X's in foxholes (where X = people who disagree with me on some issue of faith).

The Pinocchio Principle: A model that is created solely for computational reasons can take on a life of its own.

The Statistical Significance Filter: If an estimate is statistically significant, it's probably an overestimate.

Arrow's Other Theorem (weak form): Any result can be published no more than five times.

Arrow's Other Theorem (strong form): Any result *will* be published five times.

The Ramanujan Principle: Tables are read as crude graphs.

The Paradox of Philosophizing: If philosophy is outlawed, only outlaws will do philosophy.

Defaults: What statistics is the science of.

Default, the greatest trick it ever pulled: Convincing the world it didn't exist.

The Methodological Attribution Problem: The many useful contributions of a good statistical consultant, or collaborator, will often be overly attributed to the statistician's methods or philosophy.

The John Yoo Line: The point at which nothing you write gets taken seriously, and so you might as well become a hack because you have no scholarly reputation remaining.

The Chris Rock Effect: Some graphs give the pleasant feature of visualizing things we already knew, shown so well that we get a shock of recognition, the joy of relearning what we already know, but seeing it in a new way that makes us think more deeply about all sorts of related topics.

The Freshman Fallacy: Just because a freshman might raise a question, that does not make the issue irrelevant.

The Garden of Forking Paths: Multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time.

The One-Way Street Fallacy: Considering only one possibility of a change that can go in either direction.

The Pluralist's Dilemma: How to recognize that my philosophy is just one among many, that my own embrace of this philosophy is contingent on many things beyond my control, while still expressing the reasons why I prefer my philosophy to the alternatives (at least for the problems I work on).

More Vampirical Than Empirical: Those hypotheses that are unable to be killed by mere evidence. (from Jeremy Freese)

Statistical Chemotherapy: It slightly poisons your key result but shifts an undesired result above the .05 threshold. (from Jeremy Freese)

Tell Me What You Don't Know: That's what I want to ask you.

Salad Tongs: Not to be used for painting.

The Edlin Factor: How much you should scale down published estimates.

Kangaroo: When it is vigorously jumping up and down, don't use a bathroom scale to weigh a feather that is resting loosely in its pouch.

The Speed Racer Principle: Sometimes the most interesting aspect of a scientific or cultural product is not its overt content but rather its unexamined assumptions.

Uncertainty Interval: Say this instead of confidence or credible interval.

What would you do if you had all the data?: Rubin's first question.

What were you doing before you had any data?: Rubin's second question.

The Time-Reversal Heuristic: How to think about a published finding that is followed up by a careful preregistered replication.

Clarke's Law: Any sufficiently crappy research is indistinguishable from fraud.

The wedding, never about the marriage: With scientific journals, what it's all about.

The problem with peer review: The peers.

The “What does not kill my statistical significance makes it stronger” fallacy: The belief that statistical significance is particularly impressive when it was obtained under noisy conditions.

Reverse Poe: It’s evidently sincere, yet its contents are parodic.

The (Lance) Armstrong Principle: If you push people to promise more than they can deliver, they’re motivated to cheat.

The Chestertonian Principle: Extreme skepticism is a form of credulity.

The most important aspect of a statistical method: not what it does with the data but rather what data it uses.

The Pandora Principle: Once you’ve considered a possible interaction or bias or confounder, you can’t un-think it.

The Paradox of Influence: Anticipated influence becomes valueless if you end up saying whatever it takes to keep it.

Cantor’s Corner: Where you want to be.

Correlation: It does not even imply correlation.

The Javert Paradox: Suppose you find a problem with published work. If you just point it out once or twice, the authors of the work are likely to do nothing. But if you really pursue the problem, then you look like a Javert.

Eureka bias: When you think you made a discovery and then you don’t want to give it up, even if it turns out you interpreted your data wrong.

A picture plus 1000 words: Better than two pictures or 2000 words.

The Piranha Problem: These large effects can’t all coherently coexist.

The Australia principle: Build the parts of the model you need, as you need them.

Just because something is counterintuitive: Doesn’t mean it’s true.

Honesty and transparency: They’re not enough.

Breadcrumbs: I need that trail.

Random in: Random out.

16: You need this much more of a sample size to estimate an interaction that is half the size of a main effect.

The horse: Keep beating it; it’s never really dead.

The 80% power lie: None of this should be a surprise.

The causal identification Kool-Aid: The attitude by which any statistically significant difference is considered to represent some true population effect, as long as it is associated with a randomized treatment assignment, instrumental variable analysis, or regression discontinuity.

Strongest-link Fallacy: The idea that a chain of reasoning is as strong as its strongest link.

Evidence and Truth: They're different.

Assumptions and Rigor: You can't have one without the other.

Deference Trap: Without meta-science, what we get caught in.

I know there are a bunch I'm forgetting; can you all refresh my memory, please?  
Thanks.

P.S. No, I don't think I can ever match Stephen Senn in the definitions game.

Filed under Literature, Miscellaneous Statistics, Zombies  
| [Permalink](#)

## 36 Comments

1. *marcel* says:

May 25, 2009 at 12:54 pm



In WWJD, you say, "My quick answer is, Yeah, I think it would be excellent for an econometrics class if the students have applied interests. Probably I'd just go through chapter 10 (regression, logistic regression, glm, causal inference), with the later parts being optimal."

So just skip the earlier parts?

2. *Andrew Gelman* says:

May 25, 2009 at 3:23 pm



Marcel: When I say "through chapter 10," I mean, "from chapters 1 through 10." And in the last sentence above, I meant "optional," not "optimal." I'll fix that.

3. *jonathan* says:

May 26, 2009 at 7:22 am



Mister P, huh? Isn't that reflective of the old male dominant paradigm?

4. *Ken Williams* says:

March 16, 2010 at 8:15 am



I'm not grokking what "WWJD" stands for. "What Would Jennifer Do"?

5. *Andrew Gelman* says:

March 16, 2010 at 9:42 am



y

6. *The New Dogs of Politics | ansurs* says:

April 28, 2015 at 2:42 pm

[...] analysis, and concomitant immersion in the internet. I landed on Andrew Gelman's stat blog and remembered that 'humor' is a great approach and natural response to dealing with [...]

7. *New Media in Psychology and Philosophy - Daily Nous* says:

September 22, 2016 at 11:11 am

[...] using abundant researcher degrees of freedom. It's the paradigm of the theory that in the words of sociologist Jeremy Freese, is "more vampirical than empirical—unable to be killed by [...]"

8. *Proposal of a new term: "DOCO" | Stephen R. Martin says:*  
September 21, 2017 at 3:25 pm

[...] am proposing a new term: DOCO. I will, in spirit, add it to the already impressive list of useful terminology. DOCO stands for Data(or datum) Otherwise Considered [...]

9. *Zack says:*  
November 30, 2017 at 11:25 am



I can't decide if I'm very happy or very annoyed that this exists.

On the one hand, I love learning about ALL of this stuff, especially the more subtle fallacies.

But on the other hand, my list of things to read just exploded exponentially.

So, thank you. Jerk.

10. *Sameera Daniels says:*  
May 11, 2018 at 12:38 pm



One can just relegate thinking to the dustbin of history b/c much thinking, more generally is constituted from these concepts & methods. Statistics if enabling such thinking will be futile. That's what I myself have been trying to convey to my circles. I think we are due for new epistemics/epistemology. I can visualize some dimensions already. But how to communicate it is my challenge.

I have identified some individuals who I think can make superb contributions. This forum too can be helpful.

11. *Sameera Daniels says:*  
May 12, 2018 at 2:05 pm



Andrew,

It would be great if you got John Ioannidis here to debate the p-value debate. What is its disposition? Everyone goes off on leaving just shy of making an impact debate wise. Is one to conclude that this debate on backburner?

12. *"Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science" « Statistical Modeling, Causal Inference, and Social Science says:*  
September 17, 2019 at 9:01 am

[...] in the meantime, decisions need to be made, and are being made, every day. This is related to the Chestertonian principle that extreme skepticism is a form of [...]

13. *"Less Wow and More How in Social Psychology" « Statistical Modeling, Causal Inference, and Social Science says:*  
October 1, 2019 at 10:21 am

[...] mistake, which is to just assume that the claims of the 1996 study are correct. Remember the time-reversal heuristic? Pretend the large, careful study

with its null finding came first, followed by the small, [...]

14. *Schoolmarms and lightning bolts: Data faker meets Edge foundation in an unintentional reveal of problems with the Great Man model of science* « *Statistical Modeling, Causal Inference, and Social Science* says:  
October 2, 2019 at 9:04 am

[...] You can see how this could create big problems for Hauser. To start with, if you think all that matters are the lightning bolts of intuition, then you're putting yourself under a lot of pressure to stand in just the right place in that rain cloud, to be where the voltage is highest so you can throw that lightning bolt. Second, once you become a celebrated Harvard professor, then you're under even more pressure, either to come up with that damn bolt of lightning, or to play the part and act as if you've already discovered it. Remember the Armstrong principle.  
[...]

15. *Advice for a Young Economist at Heart* « *Statistical Modeling, Causal Inference, and Social Science* says:  
February 6, 2020 at 9:48 am

[...] Again, though, expect that most things will not be statistically significant—remember 16—but that doesn't mean they're not important. Instead of thinking of your study as [...]

16. *Evidence-based medicine eats itself* « *Statistical Modeling, Causal Inference, and Social Science* says:  
February 10, 2020 at 9:39 am

[...] effects for individuals or population subsets is difficult. A quick calculation finds that it takes 16 times the sample size to estimate an interaction as a main effect, and given that we are lucky if [...]

17. *Sander Greenland* says:  
February 14, 2020 at 10:12 am



"16: You need this much more of a sample size to estimate an interaction than to estimate a main effect."

As an antidote to this fallacy go to this exchange:

<https://statmodeling.stat.columbia.edu/2020/02/10/evidence-based-medicine-eats-itself/#comment-1242382>

- *Andrew* says:  
February 14, 2020 at 10:21 am



Sander:

It's not a fallacy, it's an assumption! But I agree that assumptions should be clear. So I've rewritten that entry; it now says, "16: You need this much more of a sample size to estimate an interaction that is half the size of a main effect."

I agree that the entry as written was potentially misleading, so thanks for giving me the push to fix it.

- *Sander Greenland* says:  
February 14, 2020 at 11:34 am



Great!

Now, if you'd only stop claiming that "confidence" and "credibility" intervals (CI) are "uncertainty intervals" we might approach stat nirvana. Until then, that "uncertainty" label is conning the reader and ourselves. Why? Because CI do NOT capture total uncertainty (outside of the highly idealized examples that characterize the toy universe of math stat). That means calling either kind of CI an "uncertainty interval" is part of the usual stat sales gimmick of empty quality assurance (AKA "error control").

Look, we always compute CI from a data model. In 100% of my work (and I bet about the same % of yours) there's serious uncertainty about the underlying physical data-generating process. That process has important features not captured by our model, like measurement errors and selection biases. In that case the CI flopping out of our software (whether SAS, Stan or Stata) are OVERCONFIDENCE intervals, and should not be assigned anything near either the numeric confidence or credibility shown alongside them.

Unless you carry out the arduous task of including all important uncertainty sources in the model, CI do NOT account for our actual uncertainties about the mechanisms producing the data. And that uncertainty can far exceed any uncertainty from the "random variation" allowed by the assumed model; see for example Greenland, S. (2005). Multiple-bias modeling for analysis of observational data (with discussion). J Royal Statist Soc A, 168, 267-308.

Note well: Model averaging and so-called "robust" (another con word) methods don't address this uncertainty problem. Those methods only address uncertainty about the "best" mathematical form for combining the observations, not problems with the observations like measurement error, selection bias, and (in allegedly causal analyses) uncontrolled confounding.

At best then, we can only say that CI show us a range of good-fitting models (models "highly compatible with the data") within the very restricted model family used to combine the observations.

- *Andrew says:*  
February 14, 2020 at 11:46 am



Sander:

Statistical interval estimates are used in different ways, including to express confidence in a conclusion, to express a range of credible values, and to express uncertainty about an inference. In that sense, all three terms, "confidence interval," "credibility interval," and "uncertainty interval," are reasonable, as they represent three different goals that are served by interval estimation. Separating these concepts can help, as there are examples of confidence intervals that do not include credible values and do not summarize uncertainty, there are examples of credible intervals that do not convey confidence and do not capture uncertainty, and there are examples of uncertainty intervals that are not interpretable as confidence or credibility statements.



Regarding your point: all three of these concepts—"confidence interval," "credibility interval," and "uncertainty interval"—are model-based, and all of our models are wrong. So, sure, I agree, except in some rare cases, uncertainty intervals do not capture total uncertainty. But the same is the case for confidence and credibility intervals. Except in some rare cases, confidence intervals do not have the claimed confidence properties, and, except in some rare cases, credible intervals can exclude credible values and include incredible values.

If you want to call the term "uncertainty interval" a "sales gimmick," fine. I'd prefer to say it's a mathematical statement conditional on a model, which is what I'd also say of "confidence interval" or "credibility interval." I don't see how calling it a "CI" solves this problem.

- *Sander Greenland says:*  
February 14, 2020 at 12:55 pm



Thanks Andrew! I'd like to think we are getting closer...

For this iteration, in response:

First: "CI" is just an abbreviation for "confidence", "coverage", "credible", "compatibility" etc. (e.g., "crap") interval. It solves only a speed-typing problem. What they share is that none of them capture uncertainty outside of stylized (and in my work, unrealistic) examples. Otherwise we should face the fact that the interval estimates in research articles and textbooks do not deserve labels as strong as "confidence", "coverage", or "credible". The key question is: Why should we care about uncertainty (or coverage, confidence, or credibility) given unrealistic models? At best we are only getting compatibility with those models (distinguished from the other Cs only in that it is not a hypothetical conditional; see Greenland & Chow, <http://arxiv.org/abs/1909.08583>).

Second: Fully agree that "confidence intervals" rarely have their claimed coverage properties and so are not coverage intervals; (thus their name is a confidence trick, as Bowley said upon seeing them in 1934). That's why I call them "compatibility intervals" in my work. And fully agree that "credible intervals" rarely warrant credibility near what is stated (e.g., 95%) and often contain incredible values, so that at least one modern Bayesian text (McElreath) also calls them "compatibility intervals" (albeit here the compatible models include an explicit prior).

Third: If you agree that all these CIs are model-based and thus do not capture total uncertainty, then you've made my point: "Uncertainty interval" (UI) is a very bad term for them because (apart from very special cases) CIs do not capture total uncertainty. Worse, CIs often capture only a minority of uncertainty, for the reasons I stated.

Adding those up: You have been in a leader in condemning uncertainty laundering, hence I'm baffled as to why you'd continue to promote labeling CIs as UIs. It seems obvious (to me anyway) from past researcher performance that they already take CIs as representing total uncertainty; thus relabeling CIs as "uncertainty intervals" will only dig in this misinterpretation even deeper. At best, they could be labeled as "MINIMAL-uncertainty intervals" with a massive emphasis on "minimal", but then we should caution that they may be WAY too narrow, and may be biased WAY off to an unknown side.

■ *Zad says:*

February 14, 2020 at 1:46 pm



I do like that McElreath calls Bayesian intervals "compatibility intervals", but I think it might open the doors for possible confusion with frequentist intervals. What do you think about 95% highest density posterior intervals (HDPI) for Bayesian intervals, it's one that John Kruschke uses in his Bayes book

■ *Sander Greenland says:*

February 14, 2020 at 3:14 pm



I share concerns about confusing types of intervals, but there is a sense in which they are more alike than different whenever the model is only hypothetical. In that case there is no real coverage validity (calibration), and both types of intervals are only showing compatibility of the data with their assumed model; only the "compatibility" criterion differs. This raises the possibility of other criteria, but then the resulting interval functions have usually turned out to be numerically the same as particular coverage or credibility functions (as with pure likelihood).

HDPI may be OK insofar as it sounds hard to misinterpret, but researchers are creative so may prove me wrong if they adopt it (which does not seem likely any time soon in my field).

■ *Keith O'Rourke says:*

February 14, 2020 at 6:06 pm



As Sander put it – only the "compatibility" criterion differs.

Frequentist compatibility is conditional on the specific tested parameter – how often would possible data be this or more discrepant than the observed data, with the specific tested parameter (if the specific tested parameter was true).

Bayesian compatibility is conditional on the observed data – what's the distribution of parameters that each would generate the exact same possible data as the observed data at least this often (or plausibly) or more.

Actually allows one to mitigate the degree of uncertainty laundering with Frequentist methods while introducing Bayesian inference in a way that is inoculated against uncertainty laundering using a workflow introduced first with Frequentist methods.

Or so I hope.

- *Daniel Lakeland says:*  
February 14, 2020 at 12:31 pm



Sander, shouldn't we be advocating that people actually model those often unmodeled uncertainties. I mean, for example unless your measurement apparatus is quite good, you should probably have a measurement error in your model, and unless you're doing an extraordinary job of recruiting a wide variety of patients to match the demographics of your country, you should be including some kind of sample bias or something in your model, and when there are generating process issues, you should add reasonable "width" to your likelihoods, which can be accomplished through informative priors that bias the error scales away from zero intentionally...

having done all that, we won't be perfect, but we won't be fooling ourselves either, and now, with those components in our models, we can discuss them explicitly and argue over what a good model for them is...

Anything else is I agree fooling ourselves, and like Feynman said in his cargo cult lecture, the first thing we need to do is not fool \*ourselves\*.

- *Sander Greenland says:*  
February 14, 2020 at 1:20 pm



Daniel: I agree to all that in principle. Unfortunately as with so much that is good "in principle", it's simply not practical (apart from infrequent exceptions), at least in my main application field (medical drug, device, and practice surveillance). There, few researchers can correctly interpret a P-value or CI let alone comprehend in detail the ordinary unrealistic model generating those; some high-prestige journals like JAMA even force authors to misinterpret P-values and CIs!

No surprise then that the labor involved in modeling out uncertainty sources in detail is well beyond that budgeted for analyses, and far beyond the training or competence of most teams. Worst of all, the incentives are all stacked to do no such thing, because it will inevitably lead to weaker

conclusions not even worth a press release let alone acceptance in a high-status journal.

I strongly doubt the situation is any better in other health sciences or social sciences or psychology. In the face of such harsh reality, I see no alternative than to try and force honest description of conventional outputs. At least get away from terms promoting overconfidence, like "significance", "confidence", "coverage", "credibility", etc. in favor of less sensational, more modest ordinary-language descriptions, as illustrated in Chow & Greenland, <http://arxiv.org/abs/1909.08579>

■ *Daniel Lakeland says:*

February 14, 2020 at 1:52 pm



From a getting paid to do work perspective of course you are absolutely right. I'd just say that being open and up-front about those particular issues, and telling people what *\*should\** be done even if it can't be done is a task we should bend over backwards to do. of course, it doesn't make getting contracts any easier.. let me agree entirely on that. Its a pleasure when you find someone who will buy the real deal.

My wife was discussing the budget for a grant with one of her colleagues, she proposed putting something in the budget explicitly for data analysis. Her colleague just said they should find some collaborator whose lab would do it free, after all it's only a few hours of a grad student's time or something to press the buttons on the bioinformatics software and write up the results right?

:-|

18. *You don't want a criminal journal... you want a criminal journal* « *Statistical Modeling, Causal Inference, and Social Science* says:

March 5, 2020 at 9:11 am

[...] Javert paradox rears its ugly head! Call out misconduct and you get slammed for being a [...]

19. *"As a girl, she'd been very gullible, but she had always learned more that way."* « *Statistical Modeling, Causal Inference, and Social Science* says:

March 22, 2020 at 9:56 am

[...] from being gullible (or, to put it more politely, open-minded), which is related the Chestertonian principle that extreme skepticism is a form of credulity, and let's accept that instead of poking holes [...]

20. *Let's do preregistered replication studies of the cognitive effects of air pollution—not because we think existing studies are bad, but because we think the topic is important and we want to understand it better.* « *Statistical Modeling*, says:

March 27, 2020 at 9:14 am

[...] replicate," but rather to provide new estimates that we can use, following the time-reversal [...]

21. *Noise-mining as standard practice in social science* « *Statistical Modeling, Causal Inference, and Social Science* says:  
April 3, 2020 at 9:04 am
- [...] 1. The researchers seem to have completely internalized the biases arising from the statistical significance filter that lead to estimates being too high (as discussed in section 2.1 of this article), thus they came into this new experiment expecting to see a huge and statistically significant effect (recall the 80% power lie). [...]
22. *Tracking R of COVID-19 & assessing public interventions; also some general thoughts on science* « *Statistical Modeling, Causal Inference, and Social Science* says:  
April 27, 2020 at 12:49 am
- [...] always saying how statistics is the science of defaults. But there is a more general sense in which all of science is the science of defaults. In the [...]
23. *University of Washington biostatistician unhappy with ever-changing University of Washington coronavirus projections* « *Statistical Modeling, Causal Inference, and Social Science* says:  
May 5, 2020 at 4:56 pm
- [...] Model improvement is always a possibility; on important problems we're always working near Cantor's corner. But if your model is a hot steaming mess and you keep having to add new terms to keep it under [...]
24. *Here's what academic social, behavioral, and economic scientists should be working on right now.* « *Statistical Modeling, Causal Inference, and Social Science* says:  
May 13, 2020 at 10:54 am
- [...] are lots of exciting applied problems for social science to be working on. Also, God is in every leaf of every tree, so if you really push hard to solve these applied problems, the theory will come [...]
25. *Alexey Guzey's sleep deprivation self-experiment* « *Statistical Modeling, Causal Inference, and Social Science* says:  
May 26, 2020 at 9:47 am
- [...] God is in every leaf of every tree. [...]
26. *"The good news about this episode is that it's kinda shut up those people who were criticizing that Stanford antibody study because it was an un-peer-reviewed preprint. . . ."* and a P.P.P.S. with Paul Alper's line about the dead ho says:  
May 30, 2020 at 9:29 am
- [...] P.P.S. As usual, I'm open to the possibility that the conclusions in the disputed paper are correct. Just because they haven't made a convincing case and they haven't shared their data and code and people have already found problems with their data, that doesn't mean that their substantive conclusions are wrong. It just means they haven't supplied strong evidence for their claims. Remember evidence and truth. [...]