

# Radical and Not-So-Radical Principles and Practices: Discussion of Meng<sup>☆</sup>

RONALD L. WASSERSTEIN\*, ALLEN L. SCHIRM, AND NICOLE A. LAZAR

---

## Abstract

We highlight points of agreement between Meng’s suggested principles and those proposed in our 2019 editorial in *The American Statistician*. We also discuss some questions that arise in the application of Meng’s principles in practice.

KEYWORDS AND PHRASES: ATOMIC principles, Hypothesis testing, Statistical inference.

---

When one of us (Wasserstein) invited Xiao-Li Meng to be “radical” in his presentation at the 2017 Symposium on Statistical Inference, we knew that he would take us up on the offer. He joined Andrew Gelman and Marcia McNutt in a rousing closing plenary session to the conference. We are now grateful to see that presentation in print.

In this brief discussion, we put Meng’s vision in the context of a set of principles presented in our editorial [7] that called for ending the use of statistical significance and summarized the 43 papers published in a 2019 special issue of *The American Statistician*. While we endorse many of the principles espoused by Meng, we do not agree with some of his recommendations about how they might be best achieved. Rather than delineating disagreements, however, we focus on commonalities and, specifically, the close alignment between the principles underlying Meng’s radical ideas and those that we have advocated. Before concluding with suggestions on how to adhere to the principles that we have proposed, we raise some questions for the statistical community about the implications for practice of Meng’s principles.

## 1. COMMON PRINCIPLES

The principles set forth in our editorial can be remembered easily by the acronym ATOM: **A**ccept uncertainty and be **T**houghtful, **O**pen, and **M**odest. The acronym extends to ATOMIC because of the need for **I**nstitutional **C**hange. Briefly, the principles are as follows:

- Some of the challenges that we face as statisticians come from the misunderstanding that statistics eliminate the uncertainty in results. We know better, of course, but it is possible that we contribute to the perpetuation of this myth. We need to avoid “uncertainty laundering” [4], encouraging the treatment of statistical results as “being much more incomplete and uncertain than is currently the norm” [2].

- Our advocacy for “statistical thoughtfulness,” does not imply that thoughtlessness is pervasive. Rather, it implies the importance of greater emphasis on, for example, beginning with clearly stated objectives, understanding the scientific context and nature of the research (for example, is it exploratory?), investing in the quality of the data, and considering multiple analytical approaches.
- Being open means understanding the role of expert judgment and the existence of inherent subjectivity in research decision making; reporting methods and results transparently and thoroughly; and adopting other open science practices [3].
- When we embrace uncertainty and the need for thoughtfulness and openness, modesty follows naturally as we understand and convey the limitations of our work.
- And, finally, there is the need for editorial, educational, and other institutional practices to change as we apply these principles instead of rigidly focusing on dichotomizing p-values according to arbitrary thresholds and declaring results to be statistically significant or not.

Meng’s challenge to statisticians is to be trustworthy, to “deliver what they promise.” That promise refers to statistical procedures doing what they say they are doing – reliably attaining the stated coverage and error rates. In our view, the promise extends further to all of Meng’s suggestions and, indeed, to the entire enterprise of statistics. For example, the promise entails clearly communicating what the statistical results reveal and openly and modestly acknowledging what remains uncertain. The promise also involves equipping nonstatisticians to do likewise through, for example, the sorts of educational changes advocated by Meng to enhance understanding and appreciation of uncertainty.

When Meng calls on us to double the variance – in the right circumstances – he is calling on us to deliver on all

---

<sup>☆</sup>Main article: <https://doi.org/10.51387/22-NEJSDS6>.

\*Corresponding author.

four of the principles in ATOM. He cites the need for our statistical “products” to do what they say they do, and his example of product expiration dates is a helpful illustration. Opening a container of spoiled food will quickly sour one on future purchase of that product. Ninety-five percent confidence intervals that fall well short of 95% coverage leave a sour trail of scientific results that do not replicate. Recognizing and addressing this requires the modesty to acknowledge the limitations of our statistics. Meng’s call to ensure “quality at every step” – to practice “quality-guaranteed statistics” – requires us to be thoughtful. That is, care is required when choosing to double the variance, just as care is required for every fork [5] we choose on the path to deriving and presenting statistical results. And quality requires transparency (openness). Reporting and justifying our strategies for dealing with variance – perhaps by doubling the variance in certain situations – is “a reminder of always being transparent about the data and process that lead to our findings.” Of course, the very reason for doubling the variance is an acceptance that uncertainty is part of the quality guarantee.

Meng’s notion of principled corner-cutting – to “dirtyfy Bayes,” for example – calls on us to be thoughtful about the important properties of the problem at hand. Where can we give up a little (in the “Car Talk” example, we are missing the false negative rate) and still get a reasonable answer? Meng’s three parts to principled corner-cutting, which distinguish between “good, bad and ugly studies,” encapsulate a thoughtful approach whereby we understand and acknowledge the tradeoffs in the choices that we make in any data analysis. As he also notes, however, most of us were not explicitly trained in principled corner-cutting and are not teaching it. Institutional change – reforms to how we teach statistics and what we emphasize in introductory classes – can bring us all closer to feeling comfortable with giving up a little (in the right places) in order to take an analysis further.

Quality introspection, or as Meng delightfully describes it, not selling what we refuse to buy, reminds us to be thoughtful, open, and modest. He provides ideas for reward systems to bring about the institutional change needed, especially in publications. When we analyze data and publish the results, do we stand by what we write? Here, Meng calls on us to question what we do and think carefully about the implications of our analysis choices – to not “impose on others risks we are not willing to take ourselves.” Meng provides an “introspection checklist” to encourage thoughtfulness – to help us think about the quality of our data and data analysis. This is a crucial step towards both openness and modesty; it is harder to be immodest while fully and candidly recognizing and acknowledging the shortcomings of our work. On the other hand, as Meng also rightly observes, such introspection is not currently rewarded by the incentive systems under which many (most?) of us work. He proposes changes that can be made at the institutional level. For example, universities can change how they evaluate publications for decisions about promotion, tenure, and

salary increases. One of Meng’s radical suggestions is to reduce salary by a percentage corresponding to the p-value significance cutoff used in a paper for findings that are later found to be incorrect. While admittedly extreme, the idea of shifting the incentives to quality control (and, more broadly, to quality over quantity) holds some appeal. Likewise, regulatory agencies could encourage “self-quality control.”

Still, the biggest institutional change needed is changing the way that we teach statistics. Meng sees the need to “plant the seeds” early in the curriculum. Current practice at the pre-college level – and even beyond – emphasizes rules and increasing mathematical complexity, rather than an understanding of and appreciation for the variability and intrinsic uncertainty that are at the core of our field. Yet, as Meng notes, we can learn to appreciate uncertainty as the “other side of the coin” from information. He provides delightful examples of data visualizations that elementary school children create when left to their own devices and imaginations, and calls for changes to encourage and develop such thinking. If we are to effectively train the next generation to be comfortable – and, even more, fluent – in the languages of uncertainty and variation, it is important to start at the youngest ages. Children are naturally curious about the world around them. They will commonly collect data about their friends’ preferences for ice cream flavors, pets, or pizza toppings, providing an entry point in the curriculum to thinking and teaching about variability and instituting change from the very start of the educational journey.

In summary, we find Meng’s radical ideas stimulating and thoroughly ATOMIC. Meng says we must: deliver on our promises and promise no more than we can deliver; learn that some corners can be cut, albeit in principled ways; be ready to stand by what we write, acknowledging the limitations of our work; and plant seeds by teaching the languages of variation and uncertainty early in and throughout the statistics curriculum. Readers will undoubtedly find much to debate in Meng’s paper, and that is surely his intention. But whether one agrees with doubling variance, for example, the goals behind that suggestion – that our statistics deliver what they claim to deliver, that there is quality at every step, etc. – are goals that we can all get behind.

## 2. PRINCIPLES AND PRACTICES: SOME QUESTIONS FOR ALL OF US

We have argued elsewhere [7] that statistical practice has evolved over many decades such that dichotomizing p-values and declaring results significant/not significant does far more harm than good and should be abandoned. We welcome further debate about that proposal, which because it has been widely mischaracterized, we emphasize again is about abandonment of dichotomized p-values, not continuous p-values. Rather than focusing our discussion of Meng’s paper on whatever differences exist between us and him, we

have concentrated, instead, on the principles advanced by him and their substantial overlap with our ATOMIC principles. Before suggesting some best practices for adhering to the ATOMIC principles, we will raise a few questions about practices for adhering to Meng's principles and the implications of those principles for practice.

### On single studies, hypothetical replications, and soft elimination

- Outside of seemingly rare (but important) domains like industrial quality control in which there are actual repeated applications of statistical procedures – not just “hypothetical replications we conceive to be relevant” – what does a coverage or error rate mean for any particular study? What promise is delivered by a single study?
- What is gained/what principle is maintained by using findings from a single study to eliminate “from further consideration” – as though wholly incompatible with the data, rather than just less compatible – a value barely outside the estimated interval when its p-value is only slightly less than that for a retained value barely inside (or even much farther inside) the estimated interval?

### On corner cutting and the choice of $\alpha$

- What principle – besides convention – motivates the ubiquitous, corner-cutting choice of 0.05 as a threshold for declaring results statistically significant (or not significant)?
- Would Meng's principles of delivering what is promised and not selling what one is not willing to buy be better served by less corner cutting and more well-justified, application-specific balancing of the rates and consequences of Type I and II errors, at least for studies with potentially actionable results?

### On promises, tradeoffs, and disclaimers

- Do power calculations imply a promise about Type II errors that is always much less important than the promise about Type I errors implied by the chosen significance level, and if so (or if not), what is the guiding principle?
- What disclaimer should be made about the “quality guarantee” implied by power calculations in general and, especially, when power is sacrificed because an “extra protection against exceeding Type-I errors” premium has been paid, e.g., by doubling the variance?

### On false negatives, the pufferfish/selfish test, and pandemics

- Should a new treatment for an often quickly fatal disease be forever abandoned despite a hugely or even moderately beneficial estimated effect (point estimate) because the interval estimate – maybe after doubling

the variance – includes small detrimental effects? What statistical principles or “professional ethical considerations” should guide consideration of the consequences of potential false negatives?

- Is adopting a higher  $\alpha$  (to allow shorter clinical trials) “during pandemic outbreaks” necessarily “too radical to be entertained during a normal time”? Should  $\alpha$  be adjusted only under narrowly defined extreme circumstances, or for example, could a rare but quickly fatal disease be considered sufficiently extreme (especially by those who contract it) to justify a higher  $\alpha$ ? More generally, if an  $\alpha$  is needed at all (e.g., for making a decision in high consequence situations), does a generalized selfish test – only sell what you are willing to buy and only renounce what you are willing to forego – imply that  $\alpha$  should be context-specific (fit for purpose)?

## 3. ADHERING TO THE ATOMIC PRINCIPLES

In addition to the radical but thoroughly ATOMIC ideas presented by Meng at the Symposium on Statistical Inference and in his subsequent paper, many ATOMIC ideas – not always as radical – were proffered by other symposium presenters and by authors of papers in the special issue of *The American Statistician*. As we have had the opportunity to speak to groups from many disciplines about the ATOMIC principles, six relatively easy-to-implement practices have seemed to resonate. They are not all as radical as some of Meng's, but they are in the spirit of delivering on what we promise and adhering to the ATOMIC principles:

1. Lead with and focus on effect sizes and related measures of uncertainty, such as interval estimates.
2. Focus on the substantive implications of those estimates. For example, do not focus on whether the interval contains zero, but on whether the interval bounds have qualitatively different practical consequences.
3. Interpret confidence intervals as compatibility intervals, that is, describing how compatible the data are with your hypothesized model [1].
4. When presenting p-values, present them as continuous values (not categorized into significant or not), and along with the standard p-value for the null hypothesis, report p-values for other pre-specified hypotheses. For example, instead of assuming no effect, assume the minimum meaningful effect size.
5. Interpret p-values as uncertain descriptive measures of compatibility with the model. Recognize that the p-value is impacted by not only the assumption of the null hypothesis, but also the many other assumptions and choices that data analysts make [2].
6. Do not focus on the statistical measure alone (for example, the p-value) but also consider related prior evidence, plausibility of mechanism, study design and data quality, real world costs and benefits, novelty of finding,

and other factors that are relevant to the scientific and practical context and might vary by research domain [6].

Debate about these proposals and Meng’s is needed. While often radical but sometimes not so radical, Meng’s paper makes an important contribution to the debate by calling on the statistics community to challenge itself to use statistical tools better and communicate better what the results do – and do not – say.

## DISCLAIMER

The findings and conclusions in this paper were made by Wasserstein in his personal capacity. The opinions expressed in the paper are the author’s own and should not be construed to represent any American Statistical Association statement or policy.

*Accepted 7 September 2022*

## REFERENCES

- [1] AMRHEIN, V., GREENLAND, S. and MCSHANE, B. (2019). Scientists rise up against statistical significance. *Nature* **567** 305–307.
- [2] AMRHEIN, V., TRAFIMOW, D. and GREENLAND, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don’t expect replication. *The American Statistician* **73:sup 1** 262–270. <https://doi.org/10.1080/00031305.2018.1543137>. MR3925731
- [3] BANKS, G. C., FIELD, J. G., OSWALD, F. L., O’BOYLE, E. H., LANDIS, R. S., RUPP, D. E. and ROGELBERG, S. G. (2019). Answers to 18 questions about open science practices. *Journal of Business and Psychology* **34** 257–270.
- [4] GELMAN, A. (2016). The problems with p-values are not just with p-values. *The American Statistician (online supplement)* **70** 129–133. <https://doi.org/10.1080/00031305.2016.1154108>.
- [5] GELMAN, A. and LOKEN, E. (2014). The statistical crisis in science. *American Scientist* **102** 460–465.
- [6] MCSHANE, B., GAL, D., GELMAN, A., ROBERT, C. and TACKETT, J. L. (2019). Abandon statistical significance. *The American Statistician* **73:sup 1** 235–245. <https://doi.org/10.1080/00031305.2018.1527253>. MR3925729
- [7] WASSERSTEIN, R. L., SCHIRM, A. L. and LAZAR, N. A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician* **73:sup 1** 1–19. <https://doi.org/10.1080/00031305.2019.1583913>. MR3925703

Ronald L. Wasserstein. American Statistical Association, USA.  
E-mail address: [ron@amstat.org](mailto:ron@amstat.org)

Allen L. Schirm. Mathematica Policy Research, Retired, USA.  
E-mail address: [allenschirm@gmail.com](mailto:allenschirm@gmail.com)

Nicole A. Lazar. Department of Statistics, Penn State University, USA.  
E-mail address: [nf15182@psu.edu](mailto:nf15182@psu.edu)