On de-bunking "Fake News" in the post-truth era
How to reduce statistical error in research

Flyvbjerg, Bent; Ansar, Atif; Budzier, Alexander; Buhl, Søren; Cantarelli, Chantal; Garbuio, Massimo; Glenting, Carsten; Holm, Mette Skamris; Lovallo, Dan; Molin, Eric

DOI
10.1016/j.tra.2019.06.011

Publication date
2019

Document Version
Final published version

Published in
Transportation Research Part A: Policy and Practice

Transportation Research Part A 126 (2019) 409–411

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

On de-bunking "Fake News" in the post-truth era: How to reduce statistical error in research☆

Bent Flyvbjerg[a,*], Atif Ansar[a], Alexander Budzier[a], Søren Buhl[b], Chantal Cantarelli[c], Massimo Garbuio[d], Carsten Glenting[e], Mette Skamris Holm[f], Dan Lovallo[d], Eric Molin[g], Arne Rønnest[h], Allison Stewart[i,a], Bert van Wee[g]

[a] The University of Oxford, Saïd Business School, United Kingdom
[b] Aalborg University, Department of Mathematical Sciences, Denmark
[c] Sheffield University Management School, United Kingdom
[d] University of Sydney Business School, Australia
[e] Viegand Maagøe Inc., Copenhagen, Denmark
[f] Aalborg Municipality, Denmark
[g] Delft University of Technology, Faculty for Technology, Policy, and Management, the Netherlands
[h] The National Center for Coastal Fishing and Angling, Assens, Denmark
[i] Infrastructure Victoria, Melbourne, Australia

The editors of *Transportation Research Part A* recently received a comment by Peter Love, Lavagnon Ika and Dominic Ahiaga-Dagbui (2019) as a rejoinder to our critique (Flyvbjerg et al., 2018) of a paper by Love and Ahiaga-Dagbui (2018). However, as the rejoinder came in the form of a new paper, the editors kindly offered us the possibility of a re-rejoinder, which follows below.

1. One error covertly corrected, 13 overtly ignored

In Flyvbjerg et al. (2018: 177-78, tables 1 and 2) we rejected four postulated myths, one by one, and identified 14 serious statistical errors in the earlier paper by Love and Ahiaga-Dagbui (2018). Two mathematical statisticians helped us identify the mistakes. If we are right about these errors, then the conclusions of Love and Ahiaga-Dagbui are statistically invalid. If we are wrong, then we would expect Love and Ahiaga-Dagbui to point out where and argue why, as is common in academic discourse, so that the best validity claims may gain preference and the discourse move to a higher, more informed level. Instead, they stay silent and neglect their academic obligation to address our critique. We take this to mean that they cannot counter it.

We are happy to note, however, that by their actions if not by their words, Love et al. (2019) admit to our critique and have taken it to heart, at least in one instance. On page 23 of the accepted manuscript of Love and Ahiaga-Dagbui (2018) they wrote:

> "At a minimum, an estimate [of cost] for a large infrastructure project should include the estimated uncertainty *measured by the relative standard deviation*" (our emphasis).

In a short forerunner to Flyvbjerg et al. (2018), invited by the editor of *Local Transport Today*,[1] Flyvbjerg identified the above statement as a statistical error:

☆ We would like to dedicate this comment to the memory of our colleague Daniel Lunn, who was Professor of Statistics at the University of Oxford. Lunn co-authored our first comment (Flyvbjerg et al., 2018) and would have been on the team for the present follow-up had he not died suddenly in February 2019. We missed him in preparing this note and will miss him going forward.

* Corresponding author.
E-mail address: bent.flyvbjerg@sbs.ox.ac.uk (B. Flyvbjerg).

[1] The editor sent Flyvbjerg the accepted manuscript of Love and Ahiaga-Dagbui (2018). The manuscript was also available at ResearchGate.

https://doi.org/10.1016/j.tra.2019.06.011

Available online 05 July 2019

"Distributions of cost overrun for large infrastructure projects are asymmetrical and fat-tailed … *For such distributions the standard deviation is not a good measure of uncertainty*. The standard deviation ignores fat tails and gives the impression that distributions are symmetric, i.e. that overruns and underruns around the central value are equally likely, which is emphatically not the case for large infrastructure projects" (Flyvbjerg, 2018a: 3, emphasis added).

Love and Ahiaga-Dagbui evidently read Flyvbjerg's (2018a) critique, which was printed before the final publication of their manuscript, and took it to heart, because in their published paper the following words from their accepted manuscript had mysteriously disappeared:

"measured by the relative standard deviation,"

changing their final text to read instead:

"At a minimum, an estimate for a large infrastructure project should include the estimated range of uncertainty" (Love and Ahiaga-Dagbui, 2018: page 366).

We see the change from accepted to published manuscript as direct acknowledgment from Love and Ahiaga-Dagbui that we were right about their error regarding the standard deviation. However, they do not verbally acknowledge this anywhere. We believe the academic debate would benefit from them explaining why they made the change, and why their original approach to measuring uncertainty is faulty. Other scholars would then be able to learn from the mistake and not repeat it, which is the way the academy is supposed to work. We further suggest that after having corrected the first error, the scholarly thing to do would have been for Love and Ahiaga-Dagbui to address each of the remaining 13 errors, like we addressed each of Love and Ahiaga-Dagbui's points of critique in Flyvbjerg et al. (2018).

Instead of furthering academic argument in this well-tried manner, Love et al. (2019) choose to present a simple repetition of what they already said in Love and Ahiaga-Dagbui (2018). We responded to this in Flyvbjerg et al. (2018) and therefore see no reason to repeat ourselves here. Love et al. (2019) also introduce an additional topic, not covered in Love and Ahiaga-Dagbui (2018), namely a repeat of a critique by Ika (2018), whom Love and Ahiaga-Dagbui have recruited as co-author. However, Ika's critique was also addressed previously, in Flyvbjerg (2018b), and will therefore be considered only briefly here.

2. Garbage in, garbage out

Love and his colleagues make much of a so-called replication of Flyvbjerg's (2016) test of Hirschman's theory of the Hiding Hand, which Ika claims to have carried out based on a study of project performance in a sample of 161 projects.[2] Even if we were to assume that the replication was valid, it would not be a replication of the original study with the same type of data and methodology, and therefore the original study would remain valid. However, the replication is not valid for the following reasons.

First, Ika's study is based on *recalled, perceived* project performance as subjectively reported by project supervisors on a simple seven-point scale. Flyvbjerg (2016), and most other academic studies in this field, are based on *observed, actual* project performance, measured by the difference between estimated and actual costs and benefits, recorded consistently across projects following standards first systematized by Pickrell (1989) for the U.S. Department for Transportation and today followed by academics, governments, and national audit offices around the world.

To compare recalled, subjectively perceived performance with observed, actual performance, as Ika does, is likely to entail error. For instance, behavioral science has shown that people, including experts, generally perceive and remember outcomes as more positive than they actually are or were (Gilovich et al., 2002). This is part of optimism bias and is likely to have influenced the results of Ika's analysis. People also intentionally and unintentionally misrepresent their answers in surveys and interviews to conform to what is perceived to be socially desirable, which leads to underreporting of undesirable failure (Krosnick, 1999). Such optimism and misrepresentation would explain the striking difference between Ika's results and the rest of the field, and why Ika's results are utterly at odds with the findings of behavioral science. Given this background, it is surprising, and disconcerting, that Ika would rely on his subjective data without any reflection on the important and well-documented sources of error and bias this entails, and what they mean to his conclusions.

Second, and worse, Ika's methodology seems manipulated, deliberately or not, to produce support for the Hiding Hand. When purporting to measure success and failure, Ika divides his seven-point scale into two separate sections, with 1–3 (three choices) signifying project failure and 4–7 (four choices) signifying project success. With this division, even if responses were distributed randomly across the seven-point scale, the outcome would be a strong bias for concluding that projects are successes on average, which – conveniently for Love et al. – happens to support the Hiding Hand. This outcome is not a characteristic of the projects Ika studied, but a result of the biased methodology that he designed for the study, which would have amplified well-known response biases (social desirability bias, acquiescence bias) that again would have led to underreporting of failure and overreporting of success. It is difficult to believe that Love et al. would not be aware of these self-made biases and how they affect their conclusions. But nowhere do they warn readers against the biases, which makes the analysis highly suspect.

[2] It is interesting to note that Love et al. seem happy with a sample of 161 projects for Ika's so-called replication, whereas the sample of Flyvbjerg's (2016) study, which is supposedly being replicated by the 161 projects, is criticized for containing "only 327″ projects, i.e., more than twice the size of Love et al.'s sample. This is yet another inconsistency with Love et al.

In sum, due to these errors and biases, Ika's so-called “data” – and the “findings” built on them – are methodological artifacts that say little about reality and much about a relaxed attitude to validity and truth that has no place in scholarship. No matter how many averages, standard deviations, and technical tables are computed on the basis of such wanting data, the results will be garbage-in-garbage-out. Ika's work unfortunately is as statistically inadequate as that of Love and Ahiaga-Dagbui. For the full argument regarding Ika (2018) we refer to Flyvbjerg (2018b).

3. How to get more signal and less noise in research

We would like to end by asking, at a more general level, how statistical errors like those committed by Peter Love and his colleagues may be reduced in research. Gigerenzer (2004), Taleb (2007), and others (McGregor, 1993, Leek et al., 2017, Gelman, 2018) have long argued that statistical incompetence is a problem in research on human affairs, and that such research therefore often produces spurious results, sometimes with negative consequences for policy and practice. It is crucial for any academic field, including transportation research and project management, to root out such incompetence. If this does not happen, it will become increasingly difficult to distinguish between signal and noise in research results, which will undermine the value of research and trust in it. We note with alarm that statistical noise is beginning to creep into research on cost overrun due to work like that of Love and his colleagues, who are not alone, to be fair, as we show in Flyvbjerg et al. (2018).

Researchers, referees, and editors alike have key roles to play in ensuring that statistically misleading results do not get published. We suggest the following simple heuristics as a first step to improve the statistical quality of published research:

(1) *If you're not a statistician, or don't have a strong background in statistical analysis, don't do statistical analyses*.[3] Or if you do, make sure to obtain quality assurance by a statistician. Having taken a few statistics courses and knowing how to run a statistical package on your computer is not enough. You would never allow your ear doctor to operate on your eyes. Similarly, you should not allow non-statisticians, including yourself, to do your statistical analyses. The most important skill of researchers regarding statistical analyses is to realize they are not statisticians. If you're a teacher, instill this attitude in your students.

(2) *If you're a journal referee, clearly state your level of statistical proficiency to the editors*. If you do not have the statistical expertise to review a paper, let the editors know so they can ensure that one of your fellow referees does.[4] Do not go light on the review of statistical analyses on the assumption that one of the other referees will do the work (unless you know for sure that is the case). Do not assume that the statistical analyses are probably okay, because often they are not.

(3) *If you're a journal editor, make sure that at least one referee is capable of reviewing the statistical and methodological aspects of a paper*.

We maintain that Peter Love and colleagues would have been well served by observing these heuristics, and by engaging in proper academic debate about their many statistical errors, which now hang over their work as a damaging and unresolved issue.

References

Flyvbjerg, Bent, 2016. The fallacy of beneficial ignorance: a test of Hirschman's Hiding Hand. World Dev. 84, 176–189.
Flyvbjerg, Bent, 2018a. Critics don't understand behavioral science. Local Trans. Today, April 27, pp. 3–4, https://www.transportxtra.com/publications/local-transport-today/news/56703/in-depth-academics-clash-on-causes-of-transport-cost-overruns.
Flyvbjerg, Bent, 2018b. Planning fallacy or hiding hand: which is the better explanation? World Dev. 103, 383–386.
Flyvbjerg, Bent, Ansar, Atif, Budzier, Alexander, Buhl, Søren, Cantarelli, Chantal, Garbuio, Massimo, Glenting, Carsten, Holm, Mette Skamris, Lovallo, Dan, Lunn, Daniel, Molin, Eric, Rønnest, Arne, Stewart, Allison, van Wee, Bert, 2018. Five things you should know about cost overrun. Transport. Res. Part A: Policy Pract. 118, 174–190.
Gelman, Andrew, 2018. The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. Pers. Soc. Psychol. Bull. 44 (1), 16–23.
Gigerenzer, Gerd, 2004. Mindless statistics. J. Socio-Econ. 33 (5), 587–606.
Gilovich, Thomas, Griffin, Dale, Kahneman, Daniel (Eds.), 2002. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, Cambridge.
Ika, Lavagnon A., 2018. Beneficial or detrimental ignorance: the straw man fallacy of Flyvbjerg's test of Hirschman's Hiding Hand. World Dev. 103, 369–382.
Krosnick, Jon A., 1999. Survey research. Annu. Rev. Psychol. 50 (1), 537–567.
Leek, Jeff, McShane, Blakeley B., Gelman, Andrew, Colquhoun, David, Nuijten, Michèle B., Goodman, Steven N., 2017. Five ways to fix statistics. Nature 551, 557–559.
Love, Peter E.D., Ahiaga-Dagbui, Dominic D., 2018. De-bunking 'fake news' in a post-truth era: the plausible untruths of cost underestimation in transport infrastructure projects. Transport. Res. A: Policy Practice 113, 357–368.
Love, Peter E.D., Ika, Lavagnon, Ahiaga-Dagbui, Dominic D., 2019. On de-bunking 'fake news' in a post-truth era: why does the Planning Fallacy explanation for cost overruns fall short? Transport. Res. A: Policy Practice. 126, 397–408.
McGregor, James P., 1993. Procrustus and the regression model: on the misuse of the regression model. PS Polit. Sci. Polit. 26 (4), 801–804.
Pickrell, Don, 1989. Urban Rail Transit Projects: Forecast Versus Actual Ridership and Cost. Department of Transportation, Washington, DC: US.
Taleb, Nassim Nicholas, 2007. Black Swans and the domains of statistics. Am. Statist. 61 (3), 198–200.

[3] By statisticians we here mean people who master statistics at a level equivalent to people who were trained as statisticians. Many econometricians would qualify, for example. We make one exception for heuristic no. 1: If statistical analyses are straightforward in the sense that all assumptions of the applied techniques are met, most people who have followed a few statistics courses would be able to conduct the analyses. However, as documented by Taleb, Gigerenzer, and others (including our previous comment) for data with fat tails, like the cost overruns and benefit shortfalls discussed in the main text, the assumptions of standard statistical techniques are typically not met. Different techniques, requiring specialist knowledge, are needed in this situation to avoid statistical error.

[4] Some journals are already applying this heuristic by asking referees to what extent they have relevant statistical expertise, for instance *Journal of Management* (*JOM*), which stipulates: "Not all reviewers have deep expertise in the variety of statistical methods used across studies submitted to *JOM*. To ensure that all papers have at least one reviewer with deep knowledge of the methods used, given your expertise in the statistical methods used in this paper, please indicate your comfort/confidence in your ability to rigorously evaluate the results reported" (https://mc.manuscriptcentral.com/jom). We suggest that all journals publishing quantitative analyses should do like *JOM*.