

CERN Colloquium

Excerpts from "Crisis? Surely you must be joking"

Andrea Saltelli Centre for the Study of the Sciences and the Humanities, University of Bergen, and Open Evidence Research, Open University of Catalonia

Thursday 7 Jun 2018, $16:30 \rightarrow 17:30$ Main Auditorium (CERN)

Where to find this talk: www.andreasaltelli.eu



Crisis in statistics?

Statistics is experiencing a quality control crisis



Effect or no effect?





As debate rumbles on about how and how much poor statistics is to blame for poor reproducibility, Nature asked influential statisticians to recommend one change to improve science. The common theme? The problem is not our maths, but ourselves. CORRESPONDENCE · 16 JANUARY 2018



Fixing statistics is more than a technical issue

Andrea Saltelli & Philip Stark

https://www.nature.com/articles/d41586-018-00647-9

CORRESPONDENCE · 16 JANUARY 2018



Integrity must underpin quality of statistics



https://www.nature.com/articles/d41586-018-00648-8

The great paradox of science is that **passionate** practitioners must carefully produce dispassionate facts (J. Ravetz Scientific Knowledge and its Social Problems Oxford Univ. Press; 1971). Meticulous technical and normative judgement, as well as morals and morale, are necessary to navigate the forking paths of the statistical garden (Saltelli and Stark, 2018)

All users of statistical techniques, as well as those in other mathematical fields such as modelling and algorithms, need an effective societal **COmmitment** to the maintenance of quality and integrity in their work. If imposed alone. technical or administrative solutions will only breed manipulation and evasion (Ravetz, 2018)

Statistics in the fray

The discipline of statistics has been going through a phase of critique and selfcriticism, due to mounting evidence of poor statistical practice of which misuse and abuse of the P-test is the most visible sign



AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND P-VALUES

Provides Principles to Improve the Conduct and Interpretation of Quantitative

Science March 7, 2016

+ twenty 'dissenting' commentaries

Wasserstein, R.L. and Lazar, N.A., 2016. 'The ASA's statement on p-values: context, process, and purpose', The American Statistician, DOI:10.1080/00031305.2016.1154108.

See also Christie Aschwanden at http://fivethirtyeight.com/features/not-even-scientists-can-easily-explain-p-values/

P-hacking (fishing for favourable p-values) and HARKing (formulating the research Hypothesis) After the Results are Known); Desire to achieve a sought for – or simply publishable – result leads to fiddling with the data points, the modelling assumptions, or the research hypotheses themselves

Leamer, E. E. Tantalus on the Road to Asymptopia. J. Econ. Perspect. 24, 31-46 (2010).

Kerr, N. L. HARKing: Hypothesizing After the Results are Known. Personal. Soc. Psychol. Rev. 2, 196–217 (1998).

A. Gelman and E. Loken, "The garden of forking paths: Why 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," 2013.

An existential crisis?

Most observers have noted that the crisis has technical as well as ethical and behavioural elements which interact with one another – e.g. the 'publish or perish' obsession has an impact on selection bias – the tendency to favour positive over negative results

Is modelling 'breaking bad'?

Unlike statistics, mathematical modelling is not a discipline, hence the lack of appropriate internal antibodies to fight a possible infection in the form of quality standards, disciplinary fora and journals and recognized leaders The heterogeneous nature of the modelling and simulation community prevents the emergence of consolidated paradigms \rightarrow

Serification and verification procedures are a rather trial and error business

This is a survey involving 283 responding modellers in J. J. **Padilla**, S. Y. Diallo, C. J. Lynch, and R. Gore, "Observations on the practice and profession of modeling and simulation: A survey approach," Simulation, vol. I14, 2017

Most users unaware of limitations, uncertainties, omissions and subjective choices in models → over-reliance in the quality of model-based inference

Modellers oversimplify or overelaborate, obfuscating model use

A large review of several existing checklists model quality: A. J. **Jakeman**, R. A. Letcher, and J. P. Norton, "Ten iterative steps in development and evaluation of environmental models," Environ. Model. Softw., vol. 21, no. 5, pp. 602–614, 2006.

Padilla et al. call for a more structured, generalized and standardized approach to verification

Jakeman et al. call for a 10 points participatory checklist including NUSAP and J. R. Ravetz's process based approach

For NUSAP: Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science and Policy. Kluwer, Dordrecht

J. R. Ravetz, "Integrated Environmental Assessment Forum, developing guidelines for 'good practice', Project ULYSSES.," 1997.http://www.jvds.nl/ulysses/eWP97-1.pdf

Modelling as a craft rather than as a science for Robert Rosen

Modelling as distinct from physical laws which can be falsified for Naomi Oreskes

R. Rosen, Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life. Columbia University Press, 1991.

N. Oreskes, K. Shrader-Frechette, and K. Belitz, "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences," Science, 263, no. 5147, 1994.

N. Oreskes, "Prediction: science, decision making, and the future of nature" in D. Sarewitz, R. A. Pielke, Jr., and R. Byerly, Jr. Eds. in Prediction, Science, Decision Making and the future of Nature, Island Press, 2010. Egregious modelling failure from Pilkey and Pilkey-Jarvis (from AIDS to coastal erosion…)

For John Kay modelling needs as input information which we don't have (The case of WEBTAG and knowing car passengers number decades into futures)

O. H. Pilkey and L. Pilkey-Jarvis, Useless Arithmetic: Why Environmental Scientists Can't Predict the Future. Columbia University Press, 2009.

J. A. Kay, "Knowing when we don't know," 2012, https://www.ifs.org.uk/docs/john_kay_feb2012.pdf

Economics

Paul Romer's Mathiness = use of mathematics to veil normative stances

Erik Reinert: scholastic tendencies in the mathematization of economics

P. M. Romer, "Mathiness in the Theory of Economic Growth," Am. Econ. Rev., vol. 105, no. 5, pp. 89–93, May 2015.

E. S. Reinert, "Full circle: economics from scholasticism through innovation and back into mathematical scholasticism," J. Econ. Stud., vol. 27, no. 4/5, pp. 364–376, Aug. 2000.

Cooping with uncertainty or quantification hubris

The main issue in existing practices of mathematical modelling is in the management of uncertainty in model-based inference. Modelling studies can be seen which tend to overestimate certainty, pretending to produce crisp numbers precise to the third decimal digits even in situation of pervasive uncertainty or ignorance



How uncertainty is downplayed in modelling studies: the case of sensitivity analysis

An engineer's vision of UA, SA



Can one lie with sensitivity analysis as one can lie with statistics?



Saltelli, A., Annoni P., 2010, How to avoid a perfunctory sensitivity analysis, Environmental Modeling and Software, 25, 1508–1517.

In 2014 out of 1000 papers in modelling 12 have a sensitivity analysis and < 1 a global SA; most SA still move one factor at a time



TOT SA/TOT MOD (%)

TOT GSA/TOT MOD (%)

Ferretti, F., Saltelli A., Tarantola, S., 2016, Trends in Sensitivity Analysis practice in the last decade, Science of the Total Environment, http://dx.doi.org/10.1016/j.scitotenv.201 6.02.133

OAT in 2 dimensions



Area circle / area square =?

~ 3/4

OAT in 3 dimensions



Volume sphere / volume cube =?

~ 1/2

OAT in 10 dimensions; Volume hypersphere / volume ten dimensional hypercube =? ~ 0.0025





Once a sensitivity analysis is done via OAT there is no guarantee that either uncertainty analysis (UA) or sensitivity analysis (SA) will be any good:

→ UA will be non conservative

SA may miss important factors

Why ethics of quantification?

Just as per the case of statistics, no solution is possible without careful appraisal of the social and cultural dimensions of the problem. We suggest that the situation calls an ethics of quantification to be developed, analogous to what is happening in the field of algorithms and big data.

Symbiotic relationship between quantification and trust

Theodor M. Porter





Porter's story: Quantification needs judgment which in turn needs trust …without trust quantification becomes mechanical, a system, and systems can be gamed





Big data and algorithms

Algorithms decide upon an ever-increasing list of cases, such as recruiting, carriers – including of researchers, prison sentencing, paroling, custody of minors…

Alexander, L. Is an algorithm any less racist than a human? | Technology | The Guardian. Available at https//www.theguardian.com/technology/2016/aug/03/algorithm-racist-human-employers-work (2016) (Accessed: 30th August 2017).

Abraham C. Turmoil rocks Canadian biomedical research community. Statnews, Available at https://www.statnews.com/2016/08/01/cihr-canada-research/ (2016) (Accessed: 30th August 2017).

R. Brauneis and E. P. Goodman, "Algorithmic Transparency for the Smart City," Algorithmic Transpar. Smart City, vol. 20, pp. 103–176, 2018.

Weapons of Math Destruction

O'Neil, C. Weapons of math destruction : how big data increases inequality and threatens democracy. (Crown/Archetype, 2016).

Algorithmic audit in New York city

Dwyer J. Showing the Algorithms Behind New City Services - The New York Times. New Yo Times Aug. 24, (2014).



Mathematical modelling does not make it to the headlines but is possibly in a worse shape Statistical modelling

Mathematical

modelling

Algorithms

Blurring lines:

"what qualities are specific to rankings, or indicators, or models, or algorithms?"

E. Popp Berman and D. Hirschman, The Sociology of Quantification: Where Are We Now?, Contemp. Sociol., vol. in press, 2017.

Ethics of quantification; a new grammar for mathematical modelling?

1. Uncertainty and sensitivity analysis (never execute the model once)

2. Sensitivity auditing and quantitative storytelling (investigate frames and motivations)

Saltelli, A., Guimarães Pereira, Â., Van der Sluijs, J.P. and Funtowicz, S., 2013, 'What do I make of your latinorum? Sensitivity auditing of mathematical modelling', Int. J. Foresight and Innovation Policy, (9), 2/3/4, 213–234.

Saltelli, A., Does Modelling need a reformation? Ideas for a new grammar of modelling, available at https://arxiv.org/abs/1712.06457

3. Replace 'model to predict and control the future' with 'model to help mapping ignorance about the future' ...

... in the process exploiting and making explicit the metaphors embedded in the model

J. R. Ravetz, "Models as metaphors," in Public participation in sustainability science : a handbook, and W. A. B. Kasemir, J. Jäger, C. Jaeger, Gardner Matthew T., Clark William C., Ed. Cambridge University Press, 2003, available at http://www.nusap.net/download.php?op=getit&lid=11

END



Extra slides

Statistics as a garden of forking paths even with no explicit HARKing





Jorge Luis Borges

Andrew Gelman

http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

S|S|S

The future(s) of open science

Social Studies of Science 2018, Vol. 48(2) 171–203 © The Author(s) 2018 Reprints and permissions: sagepub.co.uk/journalsPermissions.nav DOI: 10.1177/0306312718772086 journals.sagepub.com/home/sss





John J. Reilly Center, University of Notre Dame, Notre Dame, IN, USA

Abstract

Almost everyone is enthusiastic that 'open science' is the wave of the future. Yet when one looks seriously at the flaws in modern science that the movement proposes to remedy, the prospect for improvement in at least four areas are unimpressive. This suggests that the agenda is effectively to re-engineer science along the lines of platform capitalism, under the misleading banner of opening up science to the masses.

EC impact assessment guidelines: sensitivity analysis & auditing

European Commission		
Home	Better Regulation Guidelines	Share C and
 REFI Stakeholder consultations Roudmaps / Inception Impact Assessments Impact Assessment Evaluation Regulatory Scrutiny Board Guidelines Better Regulation Guidelines Better Regulation Toolbox" Key documents 	These guidelines explain what Better Regulation is and how it should be applied in to day practices when preparing new initiatives and proposals or managing existing policies and legislation	the day Search
	They cover the whole policy cycle, from policy preparation and adoption to implementation and application, to evaluation and revision of EU law. For each of these phases there are a number of Better Regulation principles, objectives, tools and procedures to make sure that the EU has the best regulation possible. These relate to planning, impact assessment, stakeholder consultation, implementation and evaluation. The <u>Better Regulation Cuidelines</u> are structured into chapters which cover each of the instruments of the law-making process. The corresponding toology gives more detailed and technical information.	histion re are as thent. Latest documents
		he alled <u>19/05/2015 - Better Regulatio</u> Package
	Better Regulation Guidelines are based on the outcomes of public consultation exe carried out in 2013 and 2014.	rcises Help us Improve
	 Public consultation on the revision of the Commission's Impact Assessment Quidelines 	Find what you wanted?
	 Stateholder Consultation Guidelinss Consultation on the draft Commission Evaluation Policy Guidelines 	Yes INO What were you looking for?
		Any suggestions?

http://ec.europa.eu/smart-regulation/guidelines/docs/br_toolbox_en.pdf



Philip Mirowski devotes a full chapter in Never Let a Serious Crisis Go to Waste to disparage the OVer-reliance on DSGE (Dynamic Stochastic General Equilibrium) models

P. Mirowski, Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Meltdown. Verso, 2013.

Rules for sensitivity analysis

1. Never run a model just once

 Sensitivity analysis is not "run" on a model but on a model once applied to a question

3. Sensitivity analysis should not be used to hide assumptions

4. If SA shows that a question cannot be answered change either the question or the model (don't shave the uncertainties)

5. SA shows that there is always one more bug! (Lubarsky's Law of Cybernetic Entomology)

6. Never run a SA where each factors has a 5% uncertainty range

The rules of sensitivity auditing

- Check against rhetorical use of mathematical modelling;
- 2. Adopt an "assumption hunting" attitude; focus on unearthing possibly implicit assumptions;
- 3. Check if uncertainty been instrumentally inflated or deflated.

4. Find sensitive assumptions before these find you; do your SA before publishing;

5. Aim for transparency; Show all the data;

6. Do the right sums, not just the sums right; frames; → quantitative storytelling

7. Perform a proper global sensitivity analysis.

The importance of frames Quantitative storytelling Frames; The expression 'tax relief' is apparently innocuous but it suggests that tax is a burden, as opposed to what pays for road, hospitals, education ...



George Lakoff



Lakoff, G., 2010, Why it Matters How We Frame the Environment, Environmental Communication: A Journal of Nature and Culture, 4:1, 70-81.

Lakoff, G., 2004–2014, Don't think of an elephant: know your values and frame the debate, Chelsea Green Publishing.





For Akerlof and Shiller against what the 'invisible hand' would contend - economic actors have no choice but to exploit frames to 'phish' people into practices which benefit the actors not the subject phished



George Akerlof



Robert R. Shiller

Quantitative storytelling tests frames/narratives for:

- Internal contradictions
- Feasibility (outside human control);
- Viability (under human control); and
- Desirability (normative; plurality of actors)

An example: Sensitivity auditing of the OECD PISA study



Q. Search analysis: research, academics.

Arts + Culture Business + Economy Cities Education Environment + Energy FactCheck Health + Medicine Politics + Society Science + Technology



Chemistry class at the Dong Tien Secondary School, Thai Nguyen Province, Vietnam, Asian Development Berlinflich, CC EVEA.

L. Araujo, A. Saltelli, and S. V. Schnepf, "Do PISA data justify PISA-based education policy?," Int. J. Comp. Educ. Dev., vol. 19, no. 1, pp. 20-34, 2017.

Saltelli, A., International PISA tests show how evidence-based policy can go wrong, The Conversation, June 12, 2017

With PISA the OECD gained the centre-stage in the international arena on education policies, which led to important controversies

http://www.theguardian.com/education/2014/may/06/oecdpisa-tests-damaging-education-academics

theguardian

OECD and Pisa tests are damaging education worldwide - academics

In this letter to Dr Andreas Schleicher, director of the OECD's Programme for International Student Assessment, academics from around the world express deep concern about the impact of Pisa tests and call for a halt to the next round of testing



① School children in Sichuan province in China. Academics say the OECD should develop alternatives to league tables and find more meaningful ways of reporting assessment, taking account of different cultures. Photograph: James Zeng Huang/Corbis Sygma

Critical remarks by 80 signatories of the letter:

- Flattening of curricula (exclusion of subjects)
- Short-termism (teaching to the test)
- Promoting "life skills to function in knowledge societies"
- Stressing the student
- ••• \rightarrow Stop the test!
- Ask for more participation in design

Figure 1

Present value of Scenario I (improve student performance in each country by 25 points on the PISA scale) in billion USD (PPP)



Note: Discounted value of future increases in GDP until 2090 due to reforms that improve student performance in each

http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/t hehighcostofloweducationalperformance.htm

"If every EU Member State achieved an improvement of 25 points in its PISA score as Germany and Poland did over the last decade, the GDP of the whole EU would increase by between 4% and 6% by 2090; such a 6% increase would correspond to 35 trillion Euro"

Woessmann, L. (2014), "The economic case for education", EENEE Analytical Report 20, European Expert Network on Economics of Education (EENEE), Institute and University of Munich.

We find both technical and normative issues:

1) Non response bias (which students are excluded) PISA non-response for England: the bias turned out to be twice the size of the OECD declared standard error in 2003

2) Non open data, which makes SA impossible

3) Flattening curricula (do all countries wish to prosper by becoming knowledge societies?)

4) Power implications: OECD (unelected officers and scholars) becoming a global superministry of education