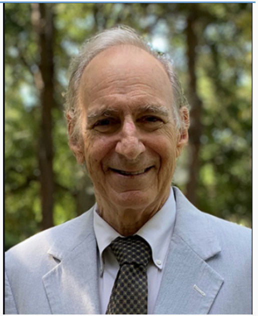
Three Rs – Reliability, Reproducibility, Replicability

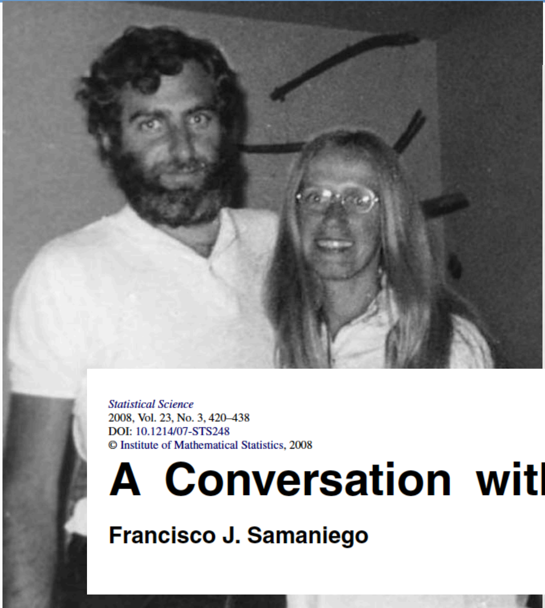
The Interplay Between Statistical Science and Data Science

Nancy Reid
University of Toronto

October 30 2020



The seventies

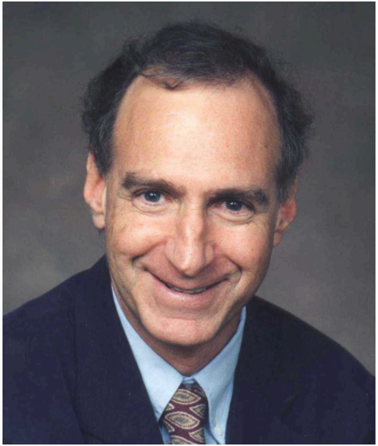


Statistical Science
2008, Vol. 23, No. 3, 420–438
DOI: 10.1214/07-STS248

A Conversation with Myles Hollander

Francisco J. Samaniego

Department Chair



Nonparametrics

AN ASYMPTOTICALLY DISTRIBUTION-FREE MULTIPLE COMPARISON PROCEDURE-TREATMENTS VS. CONTROL[1]

By Myles Hollander

The Florida State University

1. Introduction and summary. Let X_{i0} and X_{ij} $(i = 1, \cdots, n; j = 1, \cdots, k)$ be the independent measurements on the control and jth treatment in the ith block, with $P(X_{ij} \leqq x) = F_j(x - b_i)$ Here b_i is the block i nuisance parameter and the F_j ; $j = 0, \cdots, k$, are assumed continuous. Nemenyi [5] suggests treatment-control comparisons based on the statistic $T = \max_j T_{0j}$ where T_{0j} is defined by (2.1). It is shown here that, under the null hypothesis

$$H_0: F_j = F \text{ (unknown)}, \qquad j = 0, \cdots, k, \tag{1.1}$$

ON THE ASYMPTOTIC EQUIVALENCE OF TWO RANKING METHODS FOR K-SAMPLE LINEAR RANK STATISTICS

By James A. Koziol[1] and Nancy Reid

University of British Columbia and Stanford University

Two methods of ranking K samples for rank tests comparing K populations are considered. The first method ranks the K samples jointly; the second ranks the K samples pairwise. These procedures were first suggested by Dunn (1964), and Steel (1960), respectively. It is shown that both ranking procedures are asymptotically equivalent for rank-sum tests satisfying

Reliability

Reliability

Statistics How To
Statistics for the rest of us!

"Reliability is a measure of the stability or consistency of test scores"

StatisticsSolutions
Advancement Through Clarity

"Reliability refers to the extent to which a scale produces consistent results"

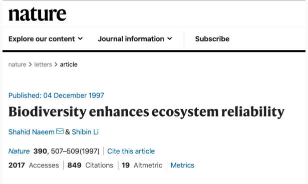
HealthKnowledge
EDUCATION, CPD AND REVALIDATION FROM PHAST

"... overall consistency of a measure ... also known as reproducibility or repeatability"

Reliability

- social sciences … consistency between raters
- medical science … consistency of a test or measurement
- physical science … “an experiment is reliable if it gives the same result when you repeat the entire experiment” link

- ecology … “ the probability that a system will provide a consistent level of performance over a given unit of time”


nature

Explore our content Journal information Subscribe

nature > letters > article

Published: 04 December 1997

Biodiversity enhances ecosystem reliability

Shahid Naeem & Shibin Li

Nature **390**, 507–509(1997) | Cite this article

2017 Accesses | **849** Citations | **19** Altmetric | Metrics

- health care … “The Institute for Healthcare Improvement uses a three-step model for applying principles of reliability to health care systems”

Institute for Healthcare Improvement

System reliability

Statistical Science
2004, Vol. 19, No. 4, 644–651
DOI 10.1214/088342304000000521
© Institute of Mathematical Statistics, 2004

Nonparametric Methods in Reliability

Myles Hollander and Edsel A. Peña

also Noether Award Lecture, several papers with Proschan, Samaniego...

The systems view: Reliability in a system of components is dependent on both the functioning of the components, and the effect of component failures on the system

Is it helpful to view statistical science as a system?

If so, what are its components?
How do we ensure reliability of the components, and the system?

Statistical systems

- social system: students; professional statisticians; academic researchers; collaborations across disciplines
- reliability through, e.g. standards of professional ethics

ASA AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

Ethical Guidelines for Statistical Practice

Prepared by the Committee on Professional Ethics of the American Statistical Association
Approved by the ASA Board in April 2018

- a data system – collection, preparation, analysis, conclusions, ...
- reliability through reproducibility

  with the same data and the same analysis the numerical findings can be reproduced

- a scientific system – foundations, theory, applications, collaborations
- reliability through replicability

  scientific findings can be verified in new experiments

Reproducibility

Reproducibility and Replicability

"A study is reproducible if you can take the original data and the computer code used to analyze the data and reproduce all of the numerical findings from the study"

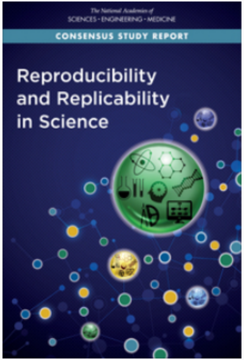
Peng 2016

" Replicability : This is the act of repeating an entire study, independently of the original investigator without the use of original data (but generally using the same methods)."

Reproducibility and Replicability

"Reproducibility means computational reproducibility – obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis"

"Replicability means obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data "



Results Replicated acm V1.1

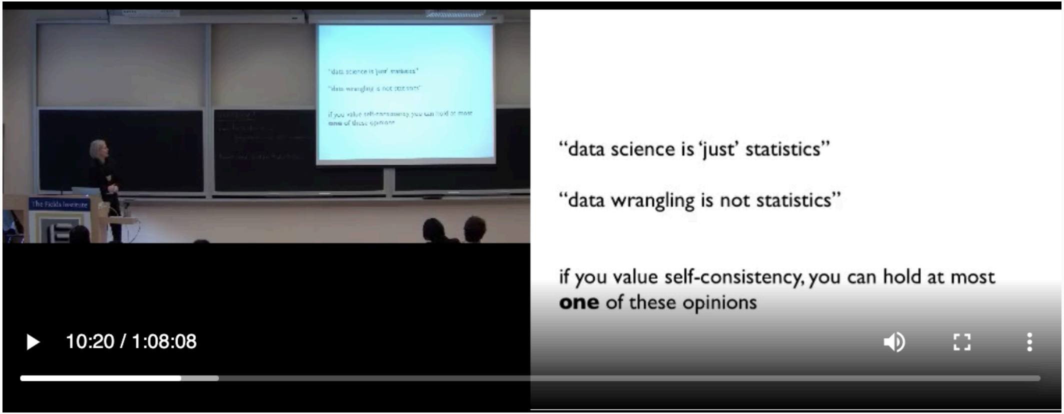
Association for Computing Machinery; Version 1.1

Reproducibility = Replicability ?

WIKIPEDIA
The Free Encyclopedia

"The replication crisis (or replicability crisis or reproducibility crisis) is, as of 2020, an ongoing methodological crisis in which it has been found that many scientific studies are difficult or impossible to replicate or reproduce."

Statistical and data science



"It's important that we build a really big tent"

FieldsLive, 2015

HARVARD DATA SCIENCE REVIEW

Ten Research Challenge Areas in Data Science

Jeannette M. Wing

Challenges and Opportunities in Statistics and Data Science: Ten Research Areas

Xuming He[1], Xihong Lin[2]

Ten research challenge areas

Ten Research Challenge Areas in Data Science

Jeannette M. Wing

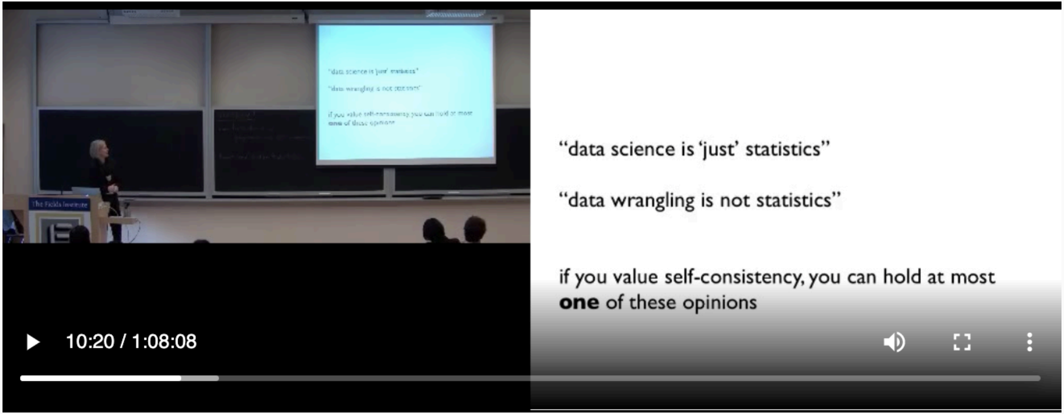
1. understanding algorithms
2. causal reasoning
3. precious data
4. multiple heterogeneous data
5. inferring from noisy/incomplete data
6. trustworthy AI
7. computing systems for data-intensive apps
8. automating front end strategies
9. privacy
10. ethics

Challenges and Opportunities in Statistics and Data Science: Ten Research Areas

Xuming He[1], Xihong Lin[2]

1. quantitative precision
2. fair and interpretable learning
3. postselection inference
4. statistical/computational efficiency
5. scalable/distributed inference
6. design for reproducibility/replicability
7. causal inference for big data
8. integrative analysis types/sources data
9. statistical analysis of privatized data
10. emerging data challenges

←——— ethics ———→



“It’s important that we build a really big tent”

FieldsLive, 2015

... statistical science

Cox & Donnelly 2011

- start with a scientific question
- assess how data could shed light on this
- plan data collection
- consider of sources of variation and how careful planning can minimize their impact

- develop strategies for data analysis: modelling, computation, methods of analysis
- assess the properties of the methods and their impact on the question at hand

- communicate the results: accurately

but not pessimistically

- visualization strategies, conveyance of uncertainties

and data science

- data acquisition
- making data trustable and usable
- management of data

- modelling and analysis

- dissemination and visualization
- data and analysis preservation

security , privacy , ethics , policy , impact

… statistical science and data science

- scientific question, data
- plan data collection
- sources of variation

- data analysis: modelling, computation, methods of analysis
- properties of the methods and their impact, replicability

- communicate
- visualization strategies, conveyance of uncertainties

- data acquisition
- making data trustable and usable
- management of data

- modelling and analysis
- computational efficiency

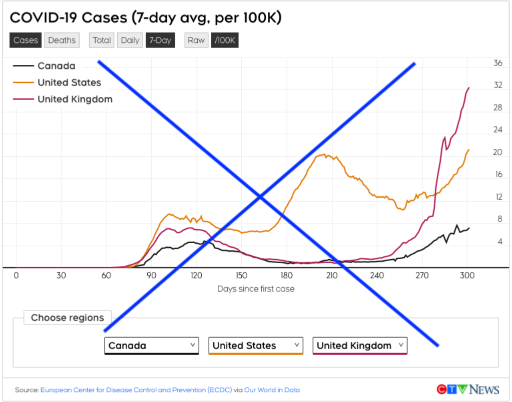
- data and analysis preservation , reproducibility

- dissemination and visualization

security , privacy , ethics , policy , impact

Examples

a hard constraint



Examples

Example 1

- Prediction, machine learning, and individual lives: an interview with Matthew Salganik
- Measuring the predictability of life outcomes with a scientific mass collaboration
- An introduction to the special collection on Fragile Families Challenge

HARVARD DATA SCIENCE REVIEW

Special Collection: Fragile Families Challenge

Submit About Contact Journal Club Subscribe

PNAS Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author, or DO

Home Articles Front Matter News Podcasts Authors

NEW RESEARCH IN Physical Sciences Social Sciences

RESEARCH ARTICLE

Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik, Ian Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul, Jennie E. Brand, Nicole Bohme Carnegie, Ryan James Compton, Debanjan Datta, Thomas Davidson, Anna Filippova, Connor Gilroy, Brian J. Goode, Eaman Jahani, Ridhi Kashyap, Antje Kirchner, Stephen McKay, Allison C. Morgan, Alex Pentland, Kivan Polimis, Louis Raes, Daniel E. Rigobon, Claudia V. Roberts, Diana M. Stanescu, Yoshihiko Suhara, Adaner Usmani, Erik H. Wang, Muna Adem, Abdulla Alhajri, Bedoor AlShebli, Redwane Amin, Ryan B. Amos, Lisa P. Argyle, Livia Baer-Bositis, Moritz Büchi, Bo-Ryehn Chung, William Eggert, Gregory Faletto, Zhilin Fan, Jeremy Freese, Tejomay Gadgil, Josh Gagné, Yue Gao, Andrew Halpern-Manners, Sonia P. Hashim, Sonia Hausen, Guanhua He, Kimberly Higuera, Bernie Hogan, Ilana M. Horwitz, Lisa M. Hummel, Naman Jain, Kun Jin, David Jurgens, Patrick Kaminski, Areg Karapetyan, E. H. Kim, Ben Leizman, Naijia Liu, Malte Möser, Andrew E. Mack, Mayank Mahajan, Noah Mandell, Helge Marahrens, Diana Mercado-Garcia, Viola Mocz, Katariina Mueller-Gastell, Ahmed Musse, Qiankun Niu, William Nowak, Hamidreza Omidvar, Andrew Or, Karen Ouyang,

Fragile Familes Data

- Fragile Families and Wellbeing Study: longitudinal survey of
  $\sim$ 4700 births; 3600 non-marital
- stratified random sample of all US cities with 200,000 or more people
- random samples of hospitals within cities
- random samples of married and unmarried births within hospitals

Reichman et al., 2001

https://fragilefamilies.princeton.edu/documentation

- six waves of data collection: birth, ages 1, 3, 5 9, 15
- each wave had a number of data collection modules
- each module had a number of sections/topics
- in-home assessments in waves 3, 4 and 5 ages 3, 5, 9

- use data from waves 1–5 — background data
- and some data from wave 6 — labelled data
- to predict outcomes on remaining data from wave 6 (age 15) — holdout data

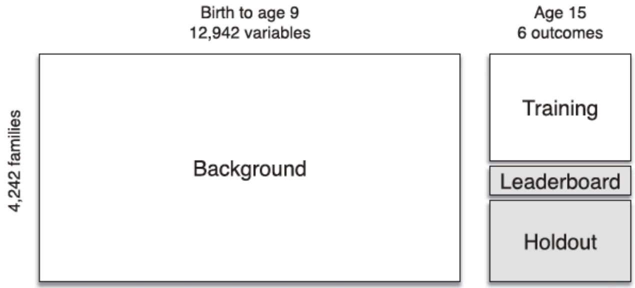


Fig. 2. Datasets in the Fragile Families Challenge. During the Fragile Families Challenge, participants used the background data (measured from child's birth to age 9 y) and the training data (measured at child age 15 y) to predict the holdout data as accurately as possible. While the Fragile

- predict any or all of 6 outcome variables — 3 continuous, 3 binary
- evaluated on relative mean-squared-error on leaderboard data
- final evaluations on holdout data at the end of the challenge — 160 teams

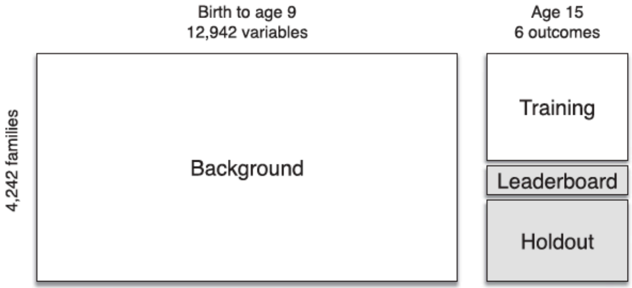
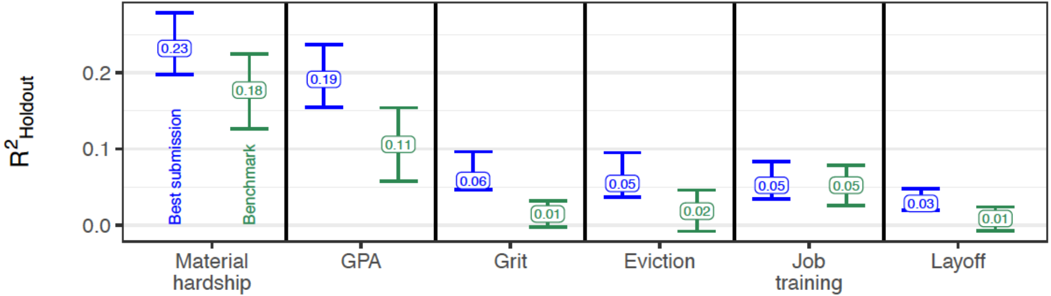


Fig. 2. Datasets in the Fragile Families Challenge. During the Fragile Families Challenge, participants used the background data (measured from child's birth to age 9 y) and the training data (measured at child age 15 y) to predict the holdout data as accurately as possible. While the Fragile Families Challenge was underway, participants could assess the accuracy

… the Results

- “even the best predictions were not very accurate
- “the best submissions were only somewhat better than … a simple benchmark model that used linear … or logistic regression with four predictor variables selected by a domain expert and a measure of the outcome [from wave 5]
- “teams used a variety of different data processing
  and statistical learning techniques
- “despite diversity in techniques, the resulting predictions were quite similar
- “within each outcome, squared prediction error was strongly associated with the family being predicted and weakly associated with the technique”

ormance of benchmark and best submissions.

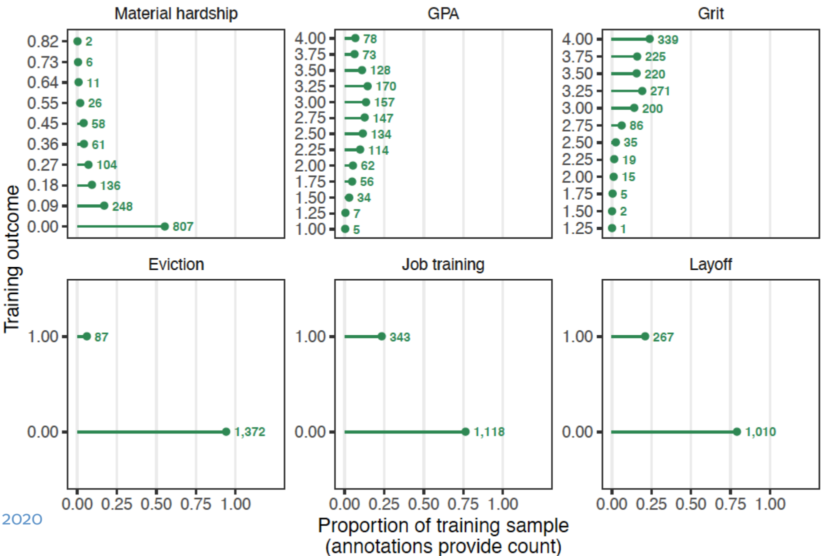


… the Conclusions

- predictive models are used in policy settings — Chouldechova et al 2018
- theory needed to address the difficulty of prediction — weather, stock market
- study can serve as a template for similar challenges — code and predictions open source
- methodology development
  - much missing data, some missing by design
  - distribution of responses quite skewed
  - "bottom up" vs "top down" approaches
  - binary vs continuous predictions
  - …

… the Conclusions

salganik et al., PNAS Supp. Fig S2



More Information

[Socius special issue]

SOCIUS

Special Collection: Fragile Families Challenge

Introduction to the Special Collection on the Fragile Families Challenge

AMERICAN SOCIOLOGICAL ASSOCIATION

Socius: Sociological Research for a Dynamic World
Volume 5: 1–21

DOI: 10.1177/2378023119871580
srd.sagepub.com
SAGE

Matthew J. Salganik[1], Ian Lundberg[1], Alexander T. Kindel[1], and Sara McLanahan[1]

Abstract

The Fragile Families Challenge is a scientific mass collaboration designed to measure and understand the predictability of life trajectories. Participants in the Challenge created predictive models of six life outcomes using data from the Fragile Families and Child Wellbeing Study, a high-quality birth cohort study. This Special Collection includes 12 articles describing participants' approaches to predicting these six outcomes as well as 3 articles describing methodological and procedural insights from running the Challenge. This introduction will help readers interpret the individual articles and help researchers interested in running future projects similar to the Fragile Families Challenge.

Keywords

life course, prediction, mass collaboration, common task method, machine learning

Examples

Example 2

Nitrous oxide a growing threat to climate

Chemical, mainly used in fertilizers, could hamper environmental efforts if emissions continue at current rate, research finds

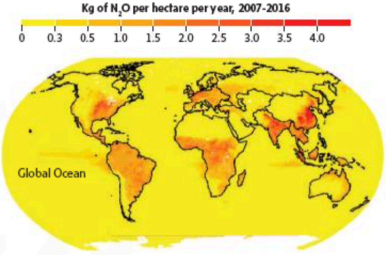
IVAN SEMENIUK
SCIENCE REPORTER

Human activity, primarily through the use of industrial fertilizers, has nearly doubled the world's atmospheric concentration of nitrous oxide and made the gas a more significant contributor to global warming than previously thought, an international research team has found.

If emissions of the gas continue to grow at their current rate, it could significantly hamper efforts by countries to meet the goal of the Paris Agreement, which seeks to limit the increase in global average temperature to less than two degrees above preindustrial levels.

NITROUS OXIDE EMISSIONS BY REGION

Emissions of the greenhouse gas nitrous oxide (N_2O) over a 10-year period are widespread around the globe. High emissions in Europe, Asia and the U.S. Midwest are associated with the use of fertilizer in agriculture. In tropical regions, natural sources play a larger role.



IVAN SEMENIUK AND JOHN SOPINSKI/THE GLOBE AND MAIL, SOURCE: NATURE

and human-caused, as well as the various processes that can remove it from the atmosphere over time.

When compared with carbon dioxide, which is responsible for lingering in the atmosphere for more than a century, or nearly 10 times longer than methane. And when it breaks down, it creates ozone-depleting byproducts. But because nitrous oxide is difficult

Tian said the new result is the culmination of a five-year effort that brought together experts in ocean, forest, soil and freshwater systems, among others, to better characterize the movement of nitrous oxide around the globe.

As expected, the study estimated significant natural sources of the gas, primarily from soils and ocean waters with emissions distributed widely around the globe. The study also provided the best estimates so far of smaller-scale emission from inland waters and production of the gas in the atmosphere by lightning.

But more important for climate scientists and policy makers is what the report reveals about human-generated or anthropogenic sources of the gas, when compared with what was previously estimated by the Intergovernmental Panel on Climate Change.

"The biggest surprise is that the rate of increase is higher than other emission scenarios that have been developed by the IPCC, and that's a cause for concern," said Nandita Basu, an associate professor of water sustainability and ing in Asia, Europe and North America, clearly stand out as hot spots, whereas high emissions in tropical regions are driven more by natural sources. Over all, human-caused emissions of the gas have increased by 30 per cent over the past four decades, the report found.

Daniel Pennock, a soil scientist and professor emeritus at the University of Saskatchewan, said the results illustrate the importance of trying to reduce agricultural emissions of nitrous oxide, adding there are several ways to do so without sacrificing crop yields through more efficient use and application of fertilizer.

Dr. Pennock said Canada should be adopting policies that create incentives for farmers to take such measures. "There's been a lot of good research in this area but relatively little uptake," he said, adding that the overapplication of fertilizer has been seen as a relatively inexpensive way for farmers to reduce risk because of weather and other factors.

Darrin Qualman, who represents Farmers for Climate Solu-

Nitrous oxide

Science Daily

Science News *from research organizations*

Nitrous oxide emissions pose an increasing climate threat, study finds

Date: October 7, 2020

Source: University of East Anglia

Summary: Rising nitrous oxide emissions are jeopardizing the climate goals of the Paris Agreement, according to a major new study. The growing use of nitrogen fertilizers in the production of food worldwide is increasing atmospheric concentrations of nitrous oxide -- a greenhouse gas 300 times more potent than carbon dioxide that remains in the atmosphere for more than 100 years.

N2O EMISSIONS POSE AN INCREASING CLIMATE THREAT, FINDS BREAKTHROUGH STUDY

Published by **Communications**
On 7th Oct 2020

SHARE

AUBURN UNIVERSITY

The Newsroom

The Official Source for Auburn University News

OCM Home | News Articles | Campus Notices | Expert Answers | Calendar | Submit News

Office of Communications & Marketing / The Newsroom / News Articles / 2020 / October /

Search The Newsroom

Breakthrough study confirms global food production poses an increasing climate threat

Published: October 07, 2020

Teri Greene | School of Forestry and Wildlife Sciences

Nitrous oxide

Stanford | News

THE GLOBE AND MAIL

Bloomberg Green

Global NEWS

CBC

Search The Guardian International e

Article

A comprehensive quantification of global nitrous oxide sources and sinks

https://doi.org/10.1038/s41586-020-2780-0

Received: 28 December 2019

Accepted: 14 August 2020

Published online: 7 October 2020

Check for updates

Hanqin Tian[1]✉, Rongting Xu[1], Josep G. Canadell[2], Rona L. Thompson[3], Wilfried Winiwarter[4,5], Parvadha Suntharalingam[6], Eric A. Davidson[7], Philippe Ciais[8], Robert B. Jackson[9,10,11], Greet Janssens-Maenhout[12,13], Michael J. Prather[14], Pierre Regnier[15], Naiqing Pan[1,16], Shufen Pan[1], Glen P. Peters[17], Hao Shi[1], Francesco N. Tubiello[18], Sönke Zaehle[19], Feng Zhou[20], Almut Arneth[21], Gianna Battaglia[22], Sarah Berthet[23], Laurent Bopp[24], Alexander F. Bouwman[25,26,27], Erik T. Buitenhuis[6,28], Jinfeng Chang[8,29], Martyn P. Chipperfield[30,31], Shree R. S. Dangal[32], Edward Dlugokencky[33], James W. Elkins[33], Bradley D. Eyre[34], Bojie Fu[16,35], Bradley Hall[33], Akihiko Ito[36], Fortunat Joos[22], Paul B. Krummel[37], Angela Landolfi[38,39], Goulven G. Laruelle[15], Ronny Lauerwald[8,15,40], Wei Li[8,41], Sebastian Lienert[22], Taylor Maavara[42], Michael MacLeod[43], Dylan B. Millet[44], Stefan Olin[45], Prabir K. Patra[46,47], Ronald G. Prinn[48], Peter A. Raymond[42], Daniel J. Ruiz[14], Guido R. van der Werf[49], Nicolas Vuichard[8], Junjie Wang[27], Ray F. Weiss[50], Kelley C. Wells[44], Chris Wilson[30,31], Jia Yang[51] & Yuanzhi Yao[1]

… Nitrous oxide emissions

Tian et al. ***Nature*** **Oct 7**

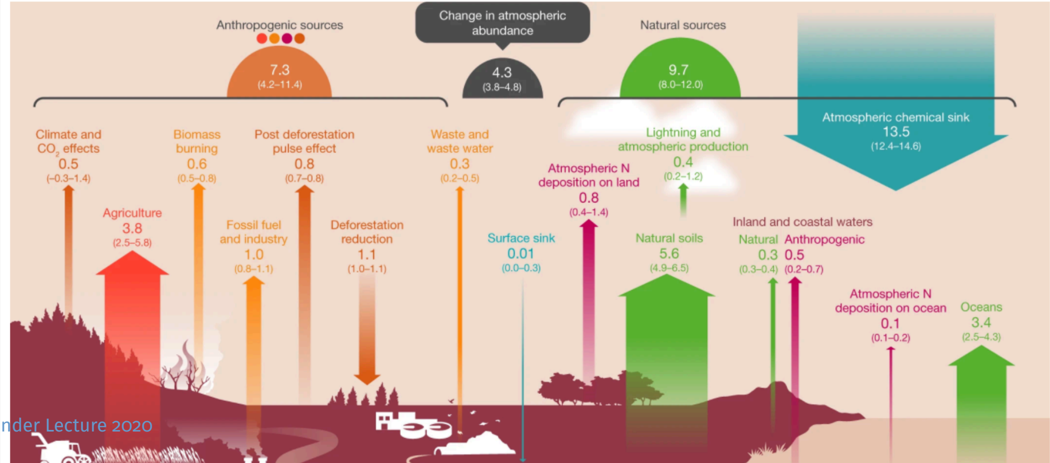
- “Dr. Tian said the new result is the culmination of a five-year effort
- “ brought together experts in ocean, forest, soil and fresh water systems, among others,
- “ The biggest surprise is that the rate of increase is higher than other emission scenarios that have been developed by the IPCC, Pr. Basu, Waterloo via G&M

- “ constructed a total of 43 flux estimates, including 30 using bottom-up approaches, 5 using top-down approaches, and 8 other estimates using observation and modelling approaches *Nature*, §1
- data synthesis from terrestial biosphere models, global ocean biogeochemistry models, dynamic land ecosystem model, various databases (FAO), published estimates of N2O from the literature, …
- atmospheric estimates from other simulations and publications, e.g. NOAA

Tian et al. *Nature* Oct 7

Fig. 1: Global N_2O budget for 2007–2016.

From: A comprehensive quantification of global nitrous oxide sources and sinks

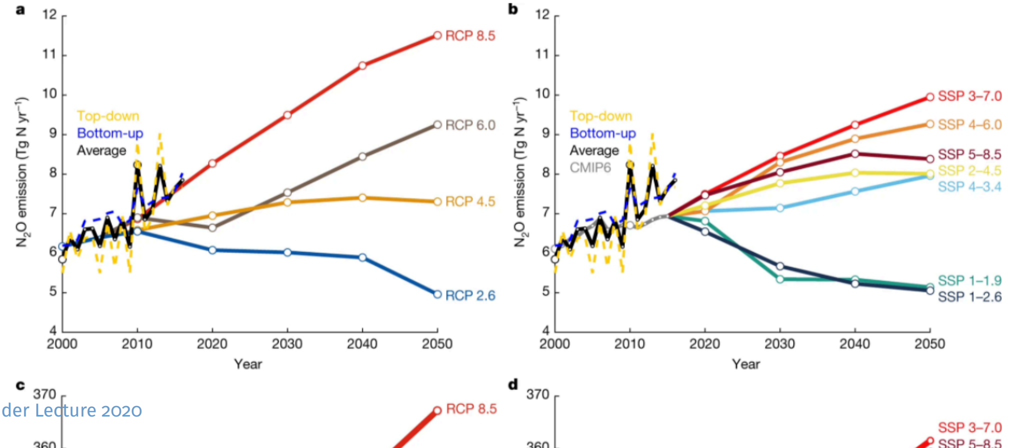


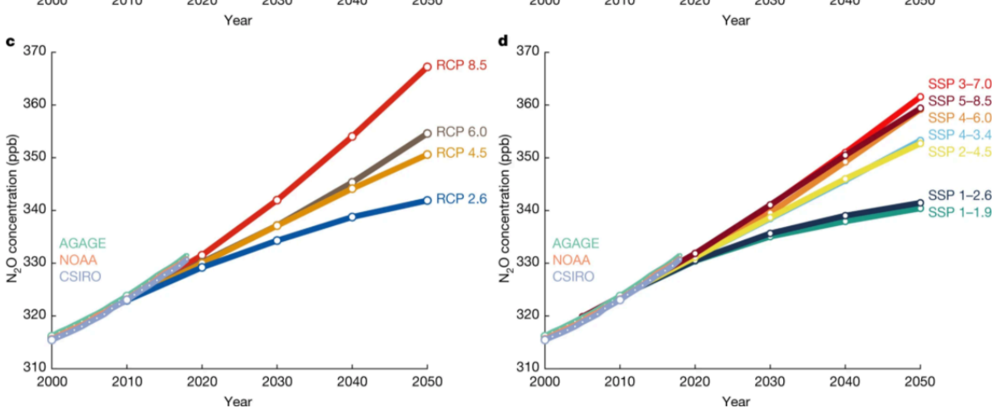
… Nitrous oxide emissions

Tian et al. *Nature* Oct 7

Fig. 4: Historical and projected global anthropogenic N_2O emissions and concentrations.

From: A comprehensive quantification of global nitrous oxide sources and sinks





a–**d**, Global anthropogenic N_2O emissions (**a**, **b**) and concentrations (**c**, **d**) compared to the four RCPs in the IPCC assessment report 5 (**a**, **c**; ref. [2]) and the new marker scenarios based on the SSPs used in CMIP6 (**b**, **d**; ref. [48]). The historical emissions data are represented as the mean of the bottom-up and top-down estimates of anthropogenic N_2O emissions, whereas the historical atmospheric concentration data are from the three available observation networks: AGAGE, NOAA, and CSIRO. Top-down anthropogenic emissions were calculated by subtracting natural fluxes derived from bottom-up approaches. To aid the comparison, the four RCPs were shifted down so that the 2005 value is equal to the 2000–2009 average of the mean of top-down and bottom-up estimates. The SSPs are harmonized[3] to match the historical emissions used in CMIP6[49]; Extended Data Fig. 10 shows the unharmonized data.

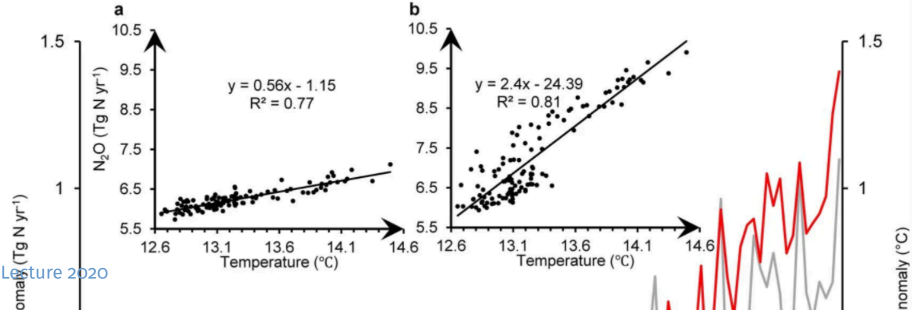
global climate models, terrestial biosphere models, observational data bases, atmospheric physics models, ...

multiscale

statistical methods for sanity checks, e.g. N2O interaction with global warming

Extended Data Fig. 8: Global simulated N_2O emission anomaly due to climate effect and global annual land surface temperature anomaly during 1901–2016.

From: A comprehensive quantification of global nitrous oxide sources and sinks



Ten research challenge areas

Ten Research Challenge Areas in Data Science

Jeannette M. Wing

1. understanding algorithms
2. causal reasoning
3. precious data
4. multiple heterogeneous data
5. inferring from noisy/incomplete data
6. trustworthy AI
7. computing systems for data-intensive
8. automating front end strategies
9. privacy
10. ethics

Challenges and Opportunities in Statistics and Data Science: Ten Research Areas

Xuming He[1], Xihong Lin[2]

1. quantitative precision
2. fair and interpretable learning
3. postselection inference
4. statistical/computational efficiency
5. scalable/distributed inference
6. design for reproducibility/replicability
7. causal inference for big data
8. integrative analysis types/sources data
9. statistical analysis of privatized data
10. emerging data challenges

ethics

Statistical Theory

- nonparametric Bayesian inference — Dirichlet process mixtures, Indian buffet processes, ...
- concentration inequalities for posterior summaries — means, modes, ...
- consistency, contraction rates, infinite-dimensional Bernstein-vonMises theorems

- sequence model $Y_i \sim N(\theta_i, 1), \quad i = 1, \ldots, n$ — high-dimensional regression

Scandinavian Journal of Statistics Theory and Applications

Most Recent | Most Cited

A Reproducing Kernel Hilbert Space log-rank test for the two-sample problem

Tamara Fernández, Nicolás Rivera

First Published: 17 October 2020

Abstract | PDF

Nitrous Oxide

Tian et al *Nature*

Methods

Terminology

This study provides an estimation of the global N_2O budget considering all possible sources and all global change processes that can perturb the budget. A total of 18 sources and three sinks of N_2O are identified and grouped into six categories (Fig. 1, Table 1): (1) natural fluxes in the absence of climate change and anthropogenic disturbances including soil emissions, surface sink, ocean emissions, lightning and atmospheric production, and natural emission from inland waters, estuaries, coastal zones (inland and coastal waters); (2) perturbed fluxes from climate/CO_2/land cover change including the effect of CO_2, the effect of climate, the post-deforestation pulse effect, and the long-term effect of reduced mature forest area; (3) direct emissions from nitrogen additions in the agricultural sector ('agriculture') including emissions from direct application of synthetic nitrogen fertilizers and manure (henceforth 'direct soil emissions'), manure left on pasture, manure management and aquaculture; (4) indirect emissions from anthropogenic nitrogen additions including atmospheric nitrogen deposition (NDEP) on land, atmospheric NDEP on ocean, and effects of anthropogenic loads of reactive nitrogen in inland waters, estuaries and coastal zones; (5) other direct anthropogenic sources including fossil fuel and industry, waste and waste water, and biomass burning; and (6) two estimates of stratospheric sinks obtained from atmospheric chemistry transport models and observations, and one tropospheric sink (Table 1, Extended Data Fig. 2).

forms). We combined the estimate from lightning with that from atmospheric production into an integrated category denoted 'Lightning and atmospheric production'. We make a simplification by considering the category 'Lightning and atmospheric production' as purely natural; however, atmospheric production is affected to some extent by anthropogenic activities through enhancing the concentrations of the reactive species NH_2 and NO_2. This category is in any case very small and the anthropogenic enhancement effect is uncertain. Lightning produces NO_x, the median estimate of which is 5 Tg N yr^{-1} (ref. [53]). We assumed an emission factor of 1% (ref. [54]) and a global estimate of 0.05 (0.02–0.09) Tg N yr^{-1} from lightning. Atmospheric production of N_2O results from the reaction of NH_2 with NO_2 (refs. [55,56]), N with NO_2, and from the oxidation of N_2 by $O(^1D)$[57], all of which constitute an estimated source of 0.3 (0.2–1.1) Tg N yr^{-1}. The estimate of the 'Surface sink' was obtained from ref. [58] and ref. [59].

The anthropogenic sources include four sub-sectors:

(a) Agriculture. This consists of four components: 'Direct soil emissions', 'Manure left on pasture', 'Manure management' and 'Aquaculture'. Data for 'Direct soil emissions' were obtained as the ensemble mean of N_2O emissions from an average of three inventories (EDGAR v4.3.2, FAOSTAT and GAINS), the SRNM/DLEM models and the NMIP/DLEM models. The statistical model SRNM covers only cropland N_2O emissions, the same as the NMIP. Thus, we add the DLEM-based estimate of pasture N_2O emissions into the two estimates in cropland to represent direct agricultural soil emissions (that is, SRNM/DLEM or NMIP/DLEM). The 'Manure left on pasture' and 'Manure management'

Nitrous Oxide

Tian et al *Nature*

Comparison with the IPCC guidelines

The IPCC has provided guidance to quantify N_2O emissions, which is widely used in emission inventories for reporting to the UNFCCC. Over time the recommended approaches have changed, which is critical for estimating emissions from agricultural soils, the largest emission source. Previous global N_2O assessments[52,82,83] based on the IPCC 1996 guidelines[84] attributed about 6.3 Tg N yr^{-1} to the agricultural sector, including both direct and indirect emissions. This estimate is notably larger than our results (Fig. 1, Table 1) derived from multiple methods, and is also larger than the most recent estimates from global inventories (EDGAR v4.3.2, FAOSTAT and GAINS) that are based on the IPCC 2006 guidelines[14]. The main reason is that indirect emissions from leaching and groundwater were overestimated in previous studies[85]. Correspondingly, projections of atmospheric N_2O concentrations that are based on these overestimated emissions[82] led to biased estimates. For example, in ref. [82], atmospheric N_2O concentrations were expected to be 340–350 ppb in the year 2020, instead of 333 ppb[5] as observed. The 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories has recently been published[86] and adopts the same approach for nitrogen application on soils, but considers the effects of different climate regimes. The new guidelines, which are based on a wealth of new scientific literature, proposed much smaller emissions from grazing animals, by a factor of 5–7. Our preliminary calculations indicate that global soil emissions based on these new guidelines may decrease by 20%–25%. Integrating estimates that rely on the IPCC methodology with estimates from process-based models provides for a more balanced assessment in this paper. We also added information from assessments[87,88] that derived agricultural emissions as the difference between atmospheric terms and other emissions such as combustion, industry and nature, and they gave comparable magnitudes (4.3–5.8 Tg N yr^{-1}) to our bottom-up results.

Uncertainty

Current data analysis and synthesis of long-term N_2O fluxes are based on a wide variety of top-down and bottom-up methods. Top-down approaches, consisting of four inversion frameworks[89–92], provide a wide range of estimates largely because of systematic errors in the modelled atmospheric transport and stratospheric loss of N_2O. In addition, the emissions from top-down analyses are dependent on the magnitude and distribution of the prior flux estimates to an extent that is strongly determined by the number of atmospheric N_2O measurements[18]. Inversions are generally not well constrained (and thus rely heavily on a priori estimates) in Africa, Southeast Asia, southern South America, and over the oceans, owing to the paucity of observations in these regions. The improvement of atmospheric transport models, more accurate priors, and more atmospheric N_2O measurements would reduce uncertainty in further top-down estimates, particularly for ocean and regional emissions.

Bottom-up approaches are subject to uncertainties in various sources from land[16] and oceans[32]. For process-based models (for example, NMIP and ocean biogeochemical models), the uncertainty is associated with differences in model configuration as well as process parameterization[16,32]. The uncertainty of estimates from NMIP could be reduced in several ways[16]. First, the six models in NMIP exhibited different spatial and temporal patterns of N_2O emissions even though they used the same forcings. Although these models have considered essential biogeochemical processes in soils (for example, biological nitrogen fixation, nitrification/denitrification, mineralization/immobilization, etc.), some missing processes such as freeze–thaw cycles and ecosystem disturbances should be included in terrestrial biosphere models to reduce uncertainties. Second, the quality of input datasets—specifically the amount and timing of nitrogen application, and spatial and

national and global N_2O flux measurement networks[17] could be used to validate model performance and to constrain large-scale model simulations. Data assimilation techniques could be used to improve model accuracy.

Current remaining uncertainty in global ocean model estimates of N_2O emission includes the contribution of N_2O flux derived from the tropical oceanic low oxygen zones (for example, the eastern Equatorial Pacific, the northern Indian ocean) relative to the global ocean. These low oxygen zones are predominantly influenced by high yield N_2O formation processes (for example, denitrification and enhanced nitrification). Regional observation-based assessments have also suggested that these regions may produce more N_2O than is simulated by the models[32]. The current generation of global ocean biogeochemistry models are not sufficiently accurate to represent the high N_2O production processes in low-oxygen zones and their associated variability (see refs. [34,94,95] for more detail). Thus, precisely representing the local ocean circulation and associated biogeochemical fluxes of these regions could further reduce the uncertainty in estimates of global and regional oceanic N_2O emissions.

Regardless of the tier approach used, greenhouse-gas inventories for agriculture suffer from high uncertainty in the underlying agriculture and rural data and statistics used as input, including statistics on fertilizer use, livestock manure availability, storage and applications, and nutrient, crop and soils management. For instance, animal waste management is an uncertain aspect, because much of the manure is either not used, or is used as a fuel or building material, or may be discharged directly to surface water[96], with important repercussions
for the calculated emission
ries using default emission
to global scales, especially f
poorly captured dependenc
climate, management, and s
It is well known, for example
greenhouse-gas inventories
by using the alternative em
environmental factors and m
of emission factors have bee
sions[31,39,62,87], and long-term
and marine aquaculture are s
N_2O sources that have not y
date, robust estimates of N_2
tion are still lacking, althoug
to the drainage of organic so
GAINS databases[28,41]. Recent
and the freeze–thaw cycle
however, are not well establi
N_2O budget.

Statistics

The Mann–Kendall test in R-3.4.4 was used to assess the significance of trends in annual N_2O emissions from each sub-sector based on the bottom-up approach.

Data availability

The relevant datasets of this study are archived in the box site of the International Center for Climate and Global Change Research at Auburn University (https://auburn.box.com/). Researchers that are interested in using the results made available in the repository are encouraged to contact the original data providers.

Code availability

The relevant codes used in this study are archived in the box site of

Statistics

The Mann–Kendall test in R-3.4.4 was used to assess the significance of trends in annual N_2O emissions from each sub-sector based on the bottom-up approach.

The role of theory

R & Cox, 2013

- how to get from data to conclusions
- with generalizable strategies
- what principles do we use to develop these strategies
- how are these strategies to be evaluated

  efficiency, precision
- a long history of the subject; using probability to both develop statistical methods and to evaluate their performance

  Bayes, Laplace, Gauss; Student, Fisher, Neyman, Pearson, Jeffreys, ...
- leading to confidence intervals, p-values, estimates and standard errors, etc.

Those pesky *p*-values

David Spiegelhalter
@d_spiegel

This paper motivates the call for the end of significance. A 25% mortality reduction, but because P=0.06 (two-sided), they declare it 'did not reduce' mortality. Appalling.
jamanetwork.com/journals/jama/…

ASA News
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

732 North Washington Street, Alexandria, VA 22314 • (703) 684-1221 • Toll Free: (888) 231-3473 • www.amstat.org • www.twitter.com/AmstatNews

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

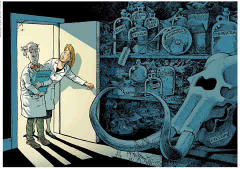
COMMENT

EVOLUTION Cooperation and conflict from ants and chimps to us p.308

HISTORY To fight denial, study Galileo and Arendt p.309

CHEMISTRY Three more unsung women — of astatine discovery p.311

PUBLISHING As well as ORCID ID and English, list authors in their own script p.311



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

When was the last time you heard a seminar speaker claim there was 'no difference' between two groups because the difference was 'statistically non-significant'?

If your experience matches ours, there's a good chance that this happened at the last talk you attended. We hope that at least someone in the audience was perplexed if, as frequently happens, a plot or table showed that there actually was a difference.

How do statistics so often lead scientists to deny differences that those not educated in statistics can plainly see? For several generations, researchers have been warned that a statistically non-significant result does not 'prove' the null hypothesis (the hypothesis that there is no difference between groups or no effect of a treatment on some measured outcome)[1]. Nor do statistically significant results 'prove' some other hypothesis. Such misconceptions have famously warped the literature with overstated claims and, less famously, led to claims of conflicts between studies where none exists.

We have some proposals to keep scientists from falling prey to these misconceptions.

PERVASIVE PROBLEM

Let's be clear about what must stop: we should never conclude there is 'no difference' or 'no association' just because a *P* value is larger than a threshold such as 0.05

21 MARCH 2019 | VOL 567 | NATURE | 305

HDSR
HARVARD DATA SCIENCE REVIEW

P-Values on Trial: Selective Reporting of (Best Practice Guides Against) Selective Reporting

by Deborah Mayo

Those pesky p-values

- science is a process
- learning is incremental
- probability expresses uncertainty
- either epistemically or empirically
- for scientific advances, empirical behaviour of procedures is key
- for decision-making, personal probabilities have an important role

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

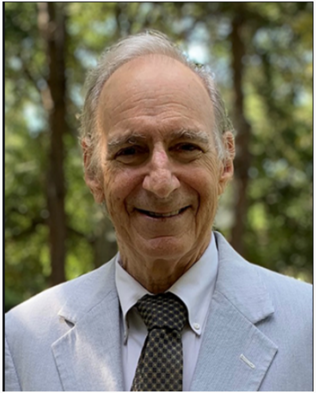
Ten Simple Rules for Effective Statistical Practice

Robert E. Kass[1], Brian S. Caffo[2], Marie Davidian[3], Xiao-Li Meng[4], Bin Yu[5], Nancy Reid[6]*

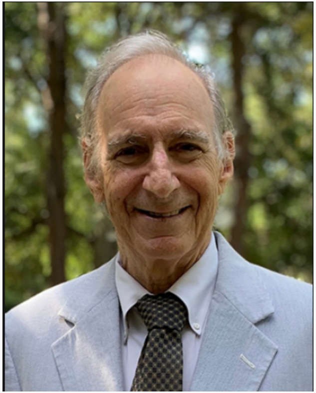
1 Department of Statistics, Machine Learning Department, and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **2** Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States of America, **3** Department of Statistics, North Carolina State University, Raleigh, North Carolina, United States of America, **4** Department of Statistics, Harvard University, Cambridge, Massachusetts, United States of America, **5** Department of Statistics and Department of Electrical Engineering and Computer Science, University of California Berkeley, Berkeley, California, United States of America, **6** Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

* reid@utstat.utoronto.ca

Thank you!



Thank you!




FLORIDA STATE UNIVERSITY
VIRES ARTES MORES
1851
Nancy Reid, Ph.D.
Hollander Distinguished
Lecturer
2020