Replicability and Reproducibility: the interplay between statistical science and data science

Nancy Reid
University of Toronto

SAPIENZA
UNIVERSITÀ DI ROMA

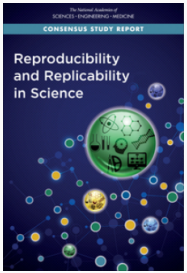
March 5 2021

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 "

"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."



… reproducibility and replicability

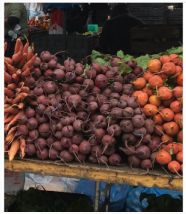
"reproducibility is a property of a study"

"replicability is a property of the result of a study"

"replicability of the stand-alone study cannot be ensured, only enhanced"

"unattended selective inference … is a silent killer of replicability in middle-size studies"

"the borders between replicability and generalization are not sharp, and need not be so"



Selective Inference: The Silent Killer of Replicability

by Yoav Benjamini

HDSR
HARVARD DATA SCIENCE REVIEW

geoscience, particle physics, economics, metrology, climate science, genomics

Statistical Science and Data Science

HDSR
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

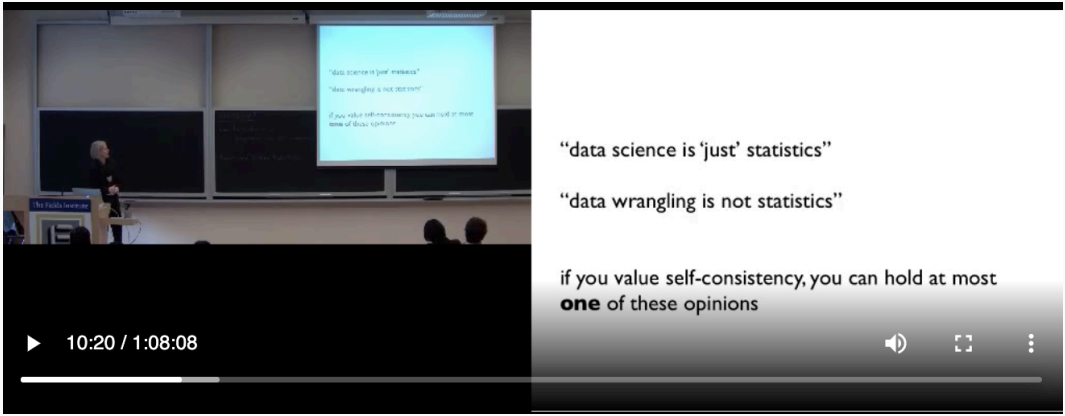
Jeannette M. Wing

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



“It’s important that we build a really big tent” FieldsLive, 2015

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

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

- modelling and analysis
- computational efficiency

- data and analysis preservation , reproducibility

- communicate
- visualization strategies, conveyance of uncertainties

- dissemination and visualization

security , privacy , ethics , policy , impact

Examples

Example 1: Vaccine efficacy

Guardian, Jan 24

Pfizer-BioNTech vaccine trial:

vaccine: 22000 subjects, 8 cases

placebo: 22000 subjects, 162 cases

$8/162 = 5\% \Longrightarrow$ 95% efficacy

Press release November 18 2020

Published December 31 2020 in NEJM

Behind the numbers: what does it mean if a Covid vaccine has '90% efficacy'?

David Spiegelhalter and Anthony Masters

Confusion surrounds the vaccines' effectiveness. The leading Cambridge professor clarifies the data behind the trials

▲ People rest in Salisbury Cathedral, England, after receiving the Pfizer/BioNTech vaccine. Photograph: Neil Hall/EPA

Pfizer publication

NEJM Dec 31

Editor's Note: This article was published on December 10, 2020, at NEJM.org.

ORIGINAL ARTICLE

Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Absalon, M.D., Alejandra Gurtman, M.D., Stephen Lockhart, D.M., John L. Perez, M.D., Gonzalo Pérez Marc, M.D., Edson D. Moreira, M.D., Cristiano Zerbini, M.D., Ruth Bailey, B.Sc., Kena A. Swanson, Ph.D., et al., for the C4591001 Clinical Trial Group*

Article Figures/Media Metrics

13 References 263 Citing Articles Letters

December 31, 2020
N Engl J Med 2020; 383:2603-2615
DOI: 10.1056/NEJMoa2034577
Chinese Translation 中文翻译

Results: A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6).

Table 2. Vaccine Efficacy against Covid-19 at Least 7 days after the Second Dose.*

Efficacy End Point	BNT162b2		Placebo		Vaccine Efficacy, % (95% Credible Interval)‡	Posterior Probability (Vaccine Efficacy >30%)§
	No. of Cases	Surveillance Time (n)†	No. of Cases	Surveillance Time (n)†		
		(N=18,198)		(N=18,325)		
Covid-19 occurrence at least 7 days after the second dose in participants without evidence of infection	8	2.214 (17,411)	162	2.222 (17,511)	95.0 (90.3–97.6)	>0.9999
		(N=19,965)		(N=20,172)		
Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection	9	2.332 (18,559)	169	2.345 (18,708)	94.6 (89.9–97.3)	>0.9999

* The total population without baseline infection was 36,523; total population including those with and those without prior evidence of infection was 40,137.
† The surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.
‡ The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.
§ Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19

Efficacy Endpoint Subgroup	n1[b]	BNT162b2 (30 µg) (N[a]=18198) Surveillance Time[c] (n2[d])	n1[b]	Placebo (N[a]=18325) Surveillance Time[c] (n2[d])	VE (%)	(95% CI[e])
Overall	8	2.214 (17411)	162	2.222 (17511)	95.0	(90.0, 97.9)
At risk[f]						
Yes	4	1.025 (8030)	86	1.025 (8029)	95.3	(87.7, 98.8)
No	4	1.189 (9381)	76	1.197 (9482)	94.7	(85.9, 98.6)
Age group (years) and at risk						

- main paper efficacy estimate 95.0% (90.3 – 97.6) 0.95 credible interval
- supplementary material estimate 95.0% (90.0 – 97.9) Clopper-Pearson adjusted for surveillance time
- Binomial estimate 95.0% (90.9 – 97.9) back of the envelope

$$X_{placebo} \sim \text{Poisson}(\lambda); \quad X_{vacc} \sim \text{Poisson}(\psi\lambda);$$

$$X_{vacc} \mid (X_{vacc} + X_{placebo}) \sim \text{Binom}(s, \frac{\psi}{1+\psi})$$

$$s = x_{vacc} + x_{placebo} = 8 + 162; \quad \widehat{\psi} = 8/162$$

safety

More good news

Moderna says its coronavirus vaccine appears to be 94.5% effective

Rival Pfizer announced last week its own vaccine appeared similarly effective

The Associated Press · Posted: Nov 16, 2020 7:23 AM ET | Last Updated: November 16, 2020

ESTABLISHED IN 1812 FEBRUARY 4, 2021 VOL. 384 NO. 5

Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine

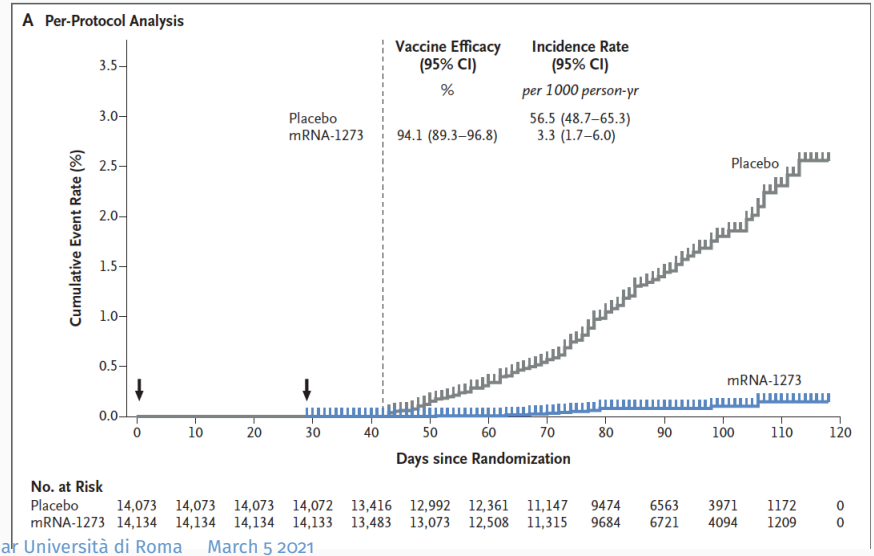
L.R. Baden, H.M. El Sahly, B. Essink, K. Kotloff, S. Frey, R. Novak, D. Diemert, S.A. Spector, N. Rouphael, C.B. Creech, J. McGettigan, S. Khetan, N. Segall, J. Solis, A. Brosz, C. Fierro, H. Schwartz, K. Neuzil, L. Corey, P. Gilbert, H. Janes, D. Follmann, M. Marovich, J. Mascola, L. Polakowski, J. Ledgerwood, B.S. Graham, H. Bennett, R. Pajon, C. Knightly, B. Leav, W. Deng, H. Zhou, S. Han, M. Ivarsson, J. Miller, and T. Zaks, for the COVE Study Group*

ABSTRACT

Results: Symptomatic Covid-19 illness was confirmed in 185 participants in the placebo group and in 11 participants in the mRNA-1273 group

vaccine efficacy was 94.1% (95% CI, 89.3 to 96.8%; $P < 0.001$)

$H_0 : efficacy < 30\%$


A Per-Protocol Analysis
Vaccine Efficacy (95% CI)
%
Incidence Rate (95% CI)
per 1000 person-yr
Placebo
mRNA-1273
94.1 (89.3–96.8)
56.5 (48.7–65.3)
3.3 (1.7–6.0)
Cumulative Event Rate (%)
Placebo
mRNA-1273
Days since Randomization
No. at Risk
Placebo 14,073 14,073 14,073 14,072 13,416 12,992 12,361 11,147 9474 6563 3971 1172 0
mRNA-1273 14,134 14,134 14,134 14,133 13,483 13,073 12,508 11,315 9684 6721 4094 1209 0

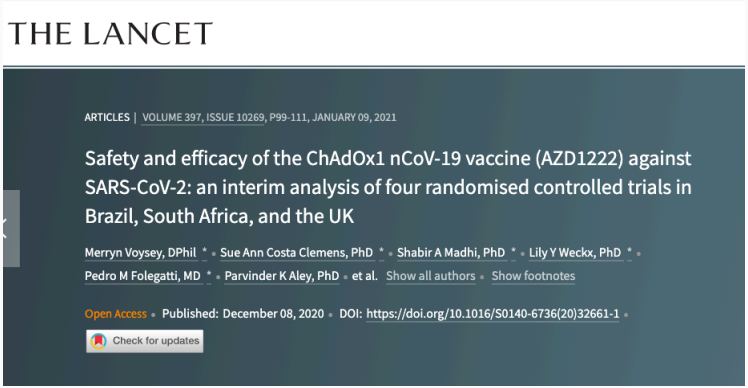
"a stratified Cox proportional hazard (PH) model with Efron's method of tie handling and with vaccine groups (mRNA-1273 or Placebo) as covariate is used to assess the vaccine efficacy (i.e. 1-HR) between mRNA-1273 vs. placebo. The model is adjusted for the same stratification factors used for randomization. The estimator of VE and its 95% CI is provided from the stratified Cox proportional model. "

stratified on age and risk groups

Estimate of efficacy: Cox PH	94.1% (89.3 – 96.8)
Binomial calculation:	94.1% (89.6 – 97.0)

Not quite as exciting

AZ press release Nov 23



"Overall vaccine efficacy across both groups was 70.4% (95.8% CI 54.8 – 80.6)"

back of envelope 70.3% (68.6 – 84.3)

9.2.1 Primary Efficacy Analyses

9.2.1.1 Pooled Analysis of Primary Efficacy Endpoint

A Poisson regression model with robust variance (Zou 2004) will be used as the primary efficacy analysis model to estimate the relative risk (RR) of the incidence of SARS-CoV-2 virologically-confirmed primary symptomatic COVID-19 between the AZD1222 and control groups. The model contains the term of study code, treatment group, and age group at randomization (ie, 18-55 years, 56-69 years, and ≥ 70 years). The logarithm of the period at risk for primary endpoint for pooled analysis will be used as an offset variable in the model to adjust for volunteers having different follow up times during which the events occur.

Vaccine efficacy (VE), which is the incidence of infection in the vaccine group relative to the incidence of infection in the control group expressed as a percentage, will be calculated as VE = 1- relative risk. The VE, and its corresponding 2-sided (1-α) % confidence interval (CI), will be estimated from the model. In addition, the 2-sided p value testing null hypothesis that the incidence of SARS-CoV-2 virologically-confirmed primary symptomatic COVID-19 between

ORIGINAL ARTICLE

BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting

Noa Dagan, M.D., Noam Barda, M.D., Eldad Kepten, Ph.D., Oren Miron, M.A., Shay Perchik, M.A., Mark A. Katz, M.D., Miguel A. Hernán, M.D., Marc Lipsitch, D.Phil., Ben Reis, Ph.D., and Ran D. Balicer, M.D.

ABSTRACT

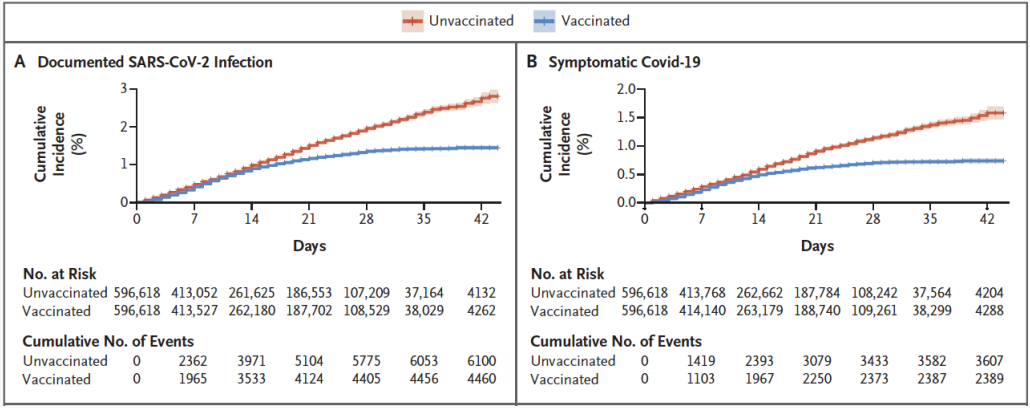
Pfizer observational study

NEJM Feb 24

- “We designed this observational study to emulate a target trial of the causal effect of the vaccine on Covid-19 outcomes”
- “We matched vaccine recipients and controls on variables associated with the probability of both vaccination and infection or severity of Covid-19”
- “Survival curves for the vaccinated and unvaccinated groups were estimated with the Kaplan-Meier estimator”
- “We calculated 95% confidence intervals using the percentile bootstrap method with 500 repetitions”

… Pfizer observational study

NEJM Feb 24



WORLD VIEW · 03 MARCH 2021

COVID vaccination studies: plan now to pool data, or be bogged down in confusion

Incompatible research designs will obscure essential answers about vaccine effectiveness. It's time to plan together.

Natalie Dean

Data Science/Statistical Science

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

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

WHY COVID VACCINES ARE SO DIFFICULT TO COMPARE

Despite the widespread roll-out of several vaccines, it could be months before they can be ranked.

By Heidi Ledford

Yusuff Adebayo Adebisi knows that a vaccine that offers 70% protection against COVID-19 could be a valuable tool against the coronavirus pandemic in Nigeria — especially if that vaccine is cheap and doesn't have to be stored at extremely cold temperatures. But what if another vaccine — one that is more expensive to buy and to store — was 95% effective?

"Should we send the less-effective vaccine to Africa? Or should we look for a way to strengthen the cold storage?" asks Adebisi, by limited supplies and hampered by limited data, says Cristina Possas, a public-health researcher at the Oswaldo Cruz Foundation in Rio de Janeiro, Brazil. "It is not possible to compare these vaccines at this point," she says.

In Bangladesh, health economist Shafiun Shimul at the University of Dhaka worries about the risks if governments delay vaccinations for months to build cold-chain infrastructure. "If you want to control infection, you have to rely on something that is contextually doable for you — it's not only about effectiveness," he says. "If they wait for perfection, I think it will be a long wait."

Statistics in the spotlight

- 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

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

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

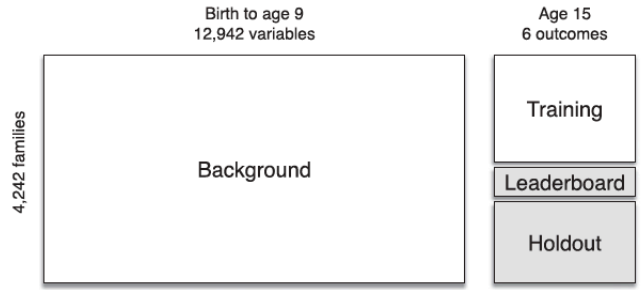


Fig. 2. Datasets in the Fragile Families Challenge. During the Fragile Families Challenge, participants used the background data (measured from child's birth to age 9 y) and the training data (measured at child age 15 y) to predict the holdout data as accurately as possible. While the Fragile

- predict any or all of 6 outcome variables
- evaluated on relative mean-squared-error on leaderboard data
- final evaluations on holdout data at the end of the challenge

3 continuous, 3 binary

160 teams

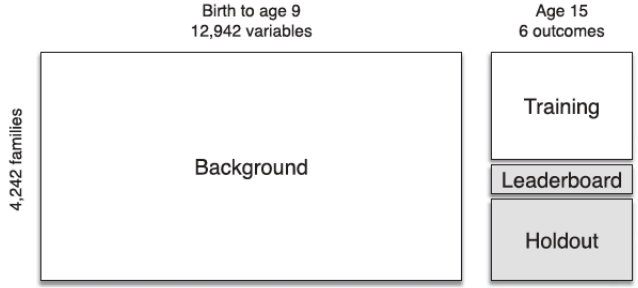
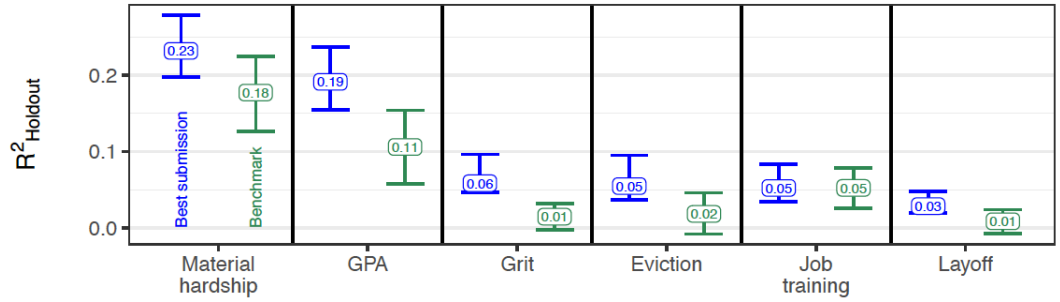


Fig. 2. Datasets in the Fragile Families Challenge. During the Fragile Families Challenge, participants used the background data (measured from child's birth to age 9 y) and the training data (measured at child age 15 y) to predict the holdout data as accurately as possible. While the Fragile

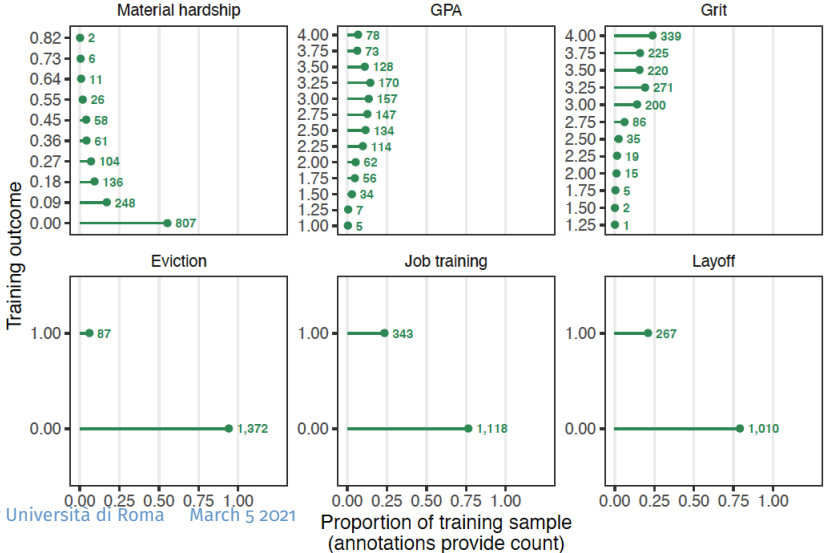
- "even the best predictions were not very accurate
- "the best submissions were only somewhat better than ... a simple benchmark model that used linear ... or logistic regression with four predictor variables selected by a domain expert and a measure of the outcome [from wave 5]
- "teams used a variety 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.



- 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
  - ...

Salganik et al., PNAS Supp. Fig S2



SOCIUS

Special Collection: Fragile Families Challenge

ASA AMERICAN SOCIOLOGICAL ASSOCIATION

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

DOI: 10.1177/2378023119871580
srd.sagepub.com

SAGE

Introduction to the Special Collection on the Fragile Families Challenge

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

Statistical Theory

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.

Role of Probability

- probability to describe physical haphazard variability — frequentist
  - probabilities represent features of the “real” world in somewhat idealized form
  - subject to empirical test and improvement
- probability to describe the uncertainty of knowledge — Bayesian
  - measures rational or “impersonal” degree of belief, — Jeffreys, 1939,1961

  or
  - measures a particular person’s degree of belief — F.P. Ramsey, 1926
  - linked to personal decision making

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

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.

HDSR
HARVARD DATA SCIENCE REVIEW

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

by Deborah Mayo

- report actual p-value, not "*", $p < 0.05$, etc. — to sensible number of decimal points
- supplement p-value with sample size, estimated power, etc.
- clarify 'exploratory' and 'confirmatory' p-values — Spiegelhalter 2017

- report effect sizes and estimated standard errors
- report confidence intervals

- pre-register trials, specifying primary and secondary outcomes
- pre-specify data analysis — NEJM

- provide a p-value function — significance function
- or some analogous distribution — Bayes posterior

… 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

Statistical theory

- causality
- data on networks
- multivariate extremes
- quantile regression
- high-dimensional inference
- model selection
- sparsity
- inference after model select
- multivariate responses
- nonparametric, robust meth
- foundations
- ...

Biometrika Advance Access published June 29, 2015
Biometrika (2015), *pp.* 1–5
Printed in Great Britain
doi: 10.1093/biomet/asv033

Big data and precision

BY D. R. COX
Nuffield College, Oxford OX1 1NF, U.K.
david.cox@nuffield.ox.ox.uk

SUMMARY

So-called big data are likely to have complex structure, in particular implying that estimates of precision obtained by applying standard statistical procedures are likely to be misleading, even if the point estimates of parameters themselves may be reasonably satisfactory. While this possibility is best explored in the context of each special case, here we outline a fairly general representation of the accretion of error in large systems and explore the possible implications for the estimation of regression coefficients. The discussion raises issues broadly parallel to the distinction between short-range and long-range dependence in time series theory.

Some key words: Components of variance; Large data; Long-range dependence; Multilevel model; Time series.

1. INTRODUCTION

Big data raise several essentially statistical issues. There may be concern over data quality and the standardization of definitions and with the rationale for inclusion in the data base. Importantly also, there is a distinction between investigations in which the research questions are at least broadly defined from

Research

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock

The ANDROMEDA-SHOCK Randomized Clinical Trial

Glenn Hernández, MD, PhD; Gustavo A. Ospina-Tascón, MD, PhD; Lucas Petri Damiani, MSc; Elisa Estenssoro, MD; Arnaldo Dubin, MD, PhD; Javier Hurtado, MD; Gilberto Friedman, MD, PhD; Ricardo Castro, MD, MPH; Leyla Alegría, RN, MSc; Jean-Louis Teboul, MD, PhD; Maurizio Cecconi, MD, FFICM; Giorgio Ferri, MD; Manuel Jibaja, MD; Ronald Pairumani, MD; Paula Fernández, MD; Diego Barahona, MD; Vladimir Granda-Luna, MD, PhD; Alexandre Biasi Cavalcanti, MD, PhD; Jan Bakker, MD, PhD; for the ANDROMEDA-SHOCK Investigators and the Latin America Intensive Care Network (LIVEN)

Aside: Andromeda Trial

Hernández et al. 2019

- comparing two treatments for septic shock
- randomized clinical trial

- estimated hazard ratio 0.75 [0.55, 1.02] — after adjusting for confounders
- 2-sided p-value 0.06 — 34.9% vs 43.4% unadjusted

- Discussion: " a peripheral perfusion-targeted resuscitation strategy did not result in a significantly lower 28-day mortality when compared with a lactate level-targeted strategy"
- Abstract: "Among patients with septic shock, a resuscitation strategy targeting normalization of capillary refill time, compared with a strategy targeting serum lactate levels, did not reduce all-cause 28-day mortality."

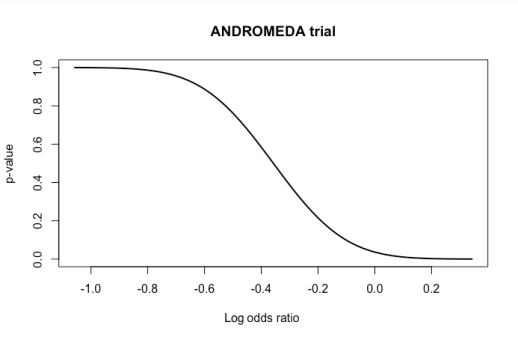
A *p*-value function

ANDROMEDA trial

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided *p*-value = 0.07

likelihood ratio test
no adjustment for covariates



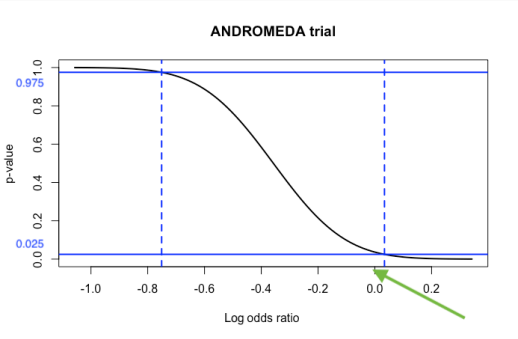
90% confidence interval: [−0.688, −0.030]
95% confidence interval: [−0.751, 0.034]
99% confidence interval: [−0.825, 0.107]

ANDROMEDA trial

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided *p*-value = 0.07

likelihood ratio test
no adjustment for covariates



90% confidence interval: [−0.688, −0.030]
95% confidence interval: [−0.751, 0.034]
99% confidence interval: [−0.825, 0.107]

Thank you!

DSS Statistics Seminar

March 5, 2021, 15:00

https://uniroma1.zoom.us/j/86881977368?pwd=SWRFcVFjMDZTa0lXZk05TE1zNm5adz09

Passcode: 432940

Replicability and Reproducibility: the interplay between statistical science and data science

N. Reid

University of Toronto

The current pandemic has brought into sharp relief the essential role of data in nearly all aspects of science, government and public health. But data is useless without explanation and interpretation, and statistical science has a long history and rich traditions for providing explanation and interpretation. In this talk I describe how data science and statistical science together can provide a robust framework for extracting insights from data reliably, and thus contribute to both replicability and reproducibility. This is illustrated with a selection of examples from recent news articles, along with some discussion on the role of the theory of inference in this framework.

SAPIENZA
UNIVERSITÀ DI ROMA