la Biostatistica: Miseria e Nobiltà

Massimo Borelli

7 luglio 2016

Three to five years later: long-term effects on ovarian function of prophylactic bilateral salpingectomy.

Authors: Roberta Venturella[a], MD, Daniela Lico[a], MD, Massimo Borelli[b], Phd, Maria G. Imbrogno[a], MD, Gabriele Cevenini[c], Phd, Errico Zupi[d], Phd, Fulvio Zullo[a], Phd, and Michele Morelli[a], Phd.

This is an important question ... Although the study has small numbers (N = 79) it can certainly help us ... The study thus contributes to the literature and should be considered for publication ... Overall, excellent research question and great addition to the literature.

DOSE AND DIAGNOSTIC PERFORMANCE COMPARISON BETWEEN PHASE-CONTRAST MAMMOGRAPHY WITH SYNCHROTRON RADIATION AND DIGITAL MAMMOGRAPHY.

The clinical study was based on 71 female patients (age range 41-82 years) that underwent to DM at hospital and, after having identify questionable or suspicious breast abnormalities confirmed also at UltraSonography (US), underwent to SRM.

Editorial note: This is a manuscript that has heavily utilized complex statistical methods that the authors do not do a good job explaining nor applying. I am also not convinced that they understand the purpose of a pilot study. Their analyses indicate that they do not. If this paper is to be considered seriously for publication the authors should bring a more experienced statistician to help frame the study design and the corresponding analyses appropriately.

```
modello = glm( Age ~ OvAge + 0,
                    family = quasigaussian)
```

```
modello = glm(Outcome ~ Score * Machine,
                    family = binomial)
```

Sergio ▸ **ROARS Return on Academic ReSearch**

3 h ·

che volete che sia, tanto i papers si contano, mica si leggono o si usano.

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[a,c]

Significance

Functional MRI (fMRI) is 25 years old, yet surprisingly its most common statistical methods have not been validated using real data. Here, we used resting-state fMRI data from 499 healthy controls to conduct 3 million task group analyses. Using this null data with different experimental designs, we estimate the incidence of significant results. In theory, we should find 5% false positives (for a significance threshold of 5%), but instead we found that the most common software packages for fMRI analysis (SPM, FSL, AFNI) can result in false-positive rates of up to 70%. These results question the validity of some 40,000 fMRI studies and may have a large impact on the interpretation of neuroimaging results.

2005, John Ioannidis

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

It can be proven that most claimed research findings are false.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2×2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2×2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Copyright: © 2005 John P. A. Ioannidis. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

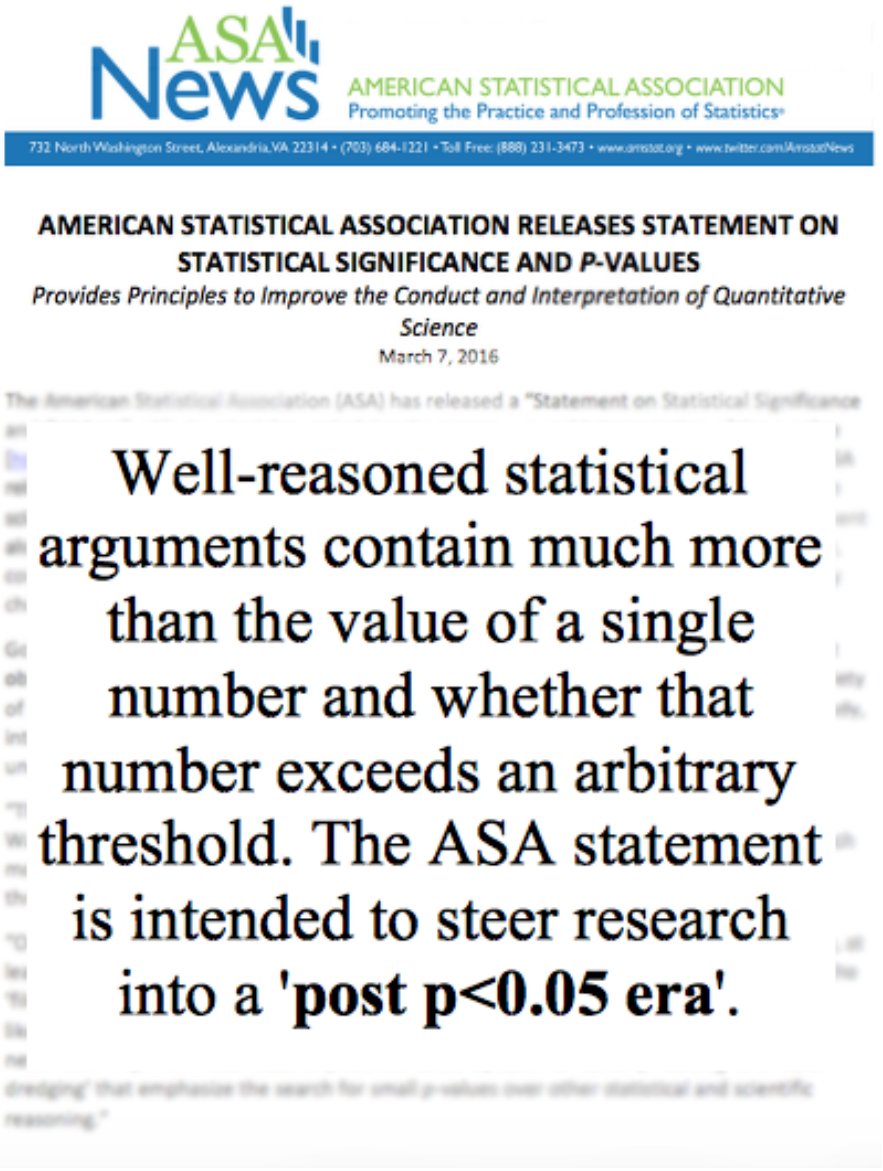
2009, Marcia Angell

It is simply no longer possible to believe much of the clinical research that is published, or to rely on the judgment of trusted physicians or authoritative medical guidelines. I take no pleasure in this conclusion, which I reached slowly and reluctantly over my two decades as an editor of The New England Journal of Medicine.

2015, Richard Horton

The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness.

2016, Ron Wasserstein



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

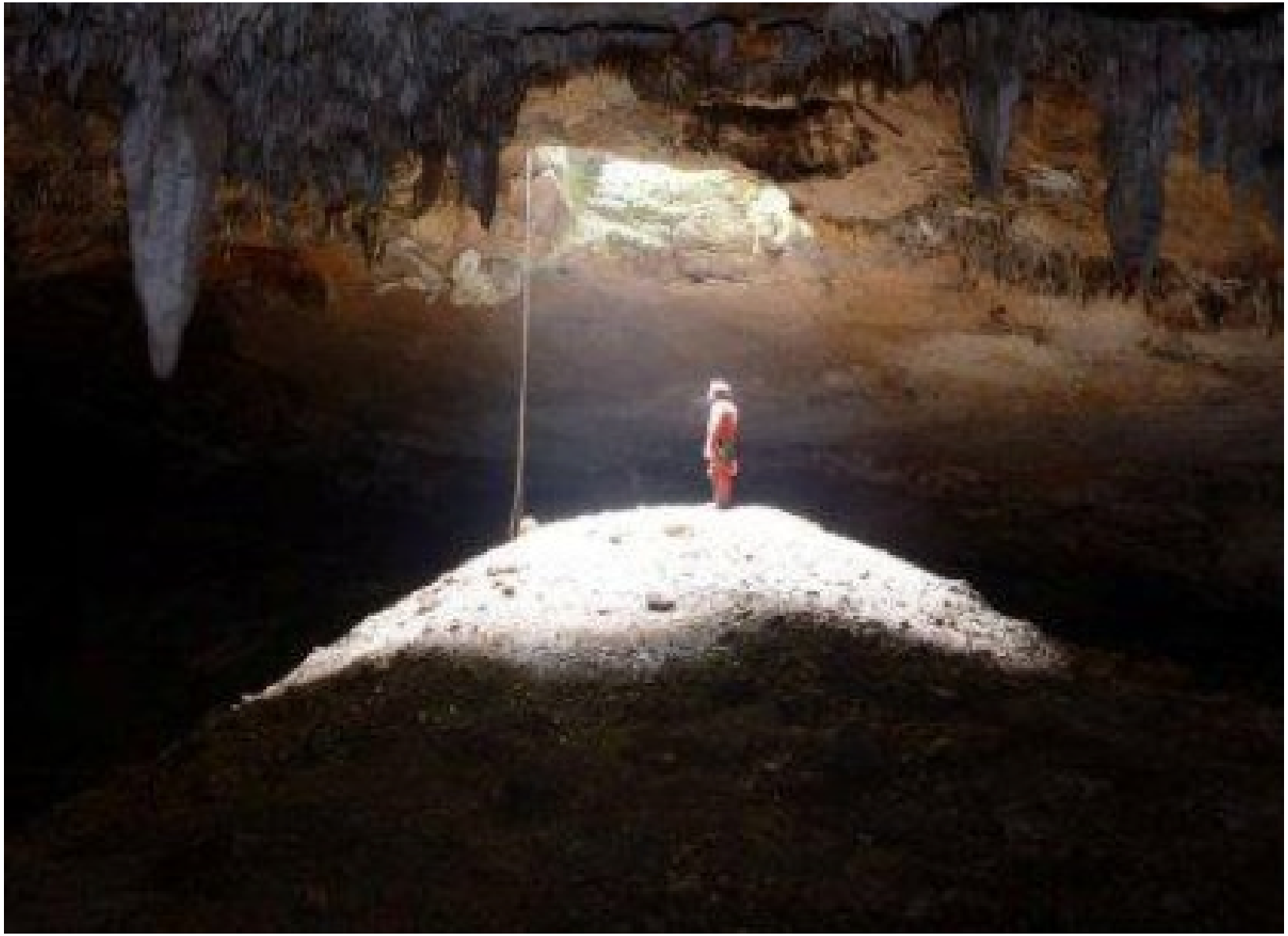
Provides Principles to Improve the Conduct and Interpretation of Quantitative Science

March 7, 2016

The American Statistical Association (ASA) has released a "Statement on Statistical Significance

Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold. The ASA statement is intended to steer research into a '**post p<0.05 era**'.

dredging' that emphasize the search for small p-values over other statistical and scientific reasoning."



the bayesian approach

a very petalous example

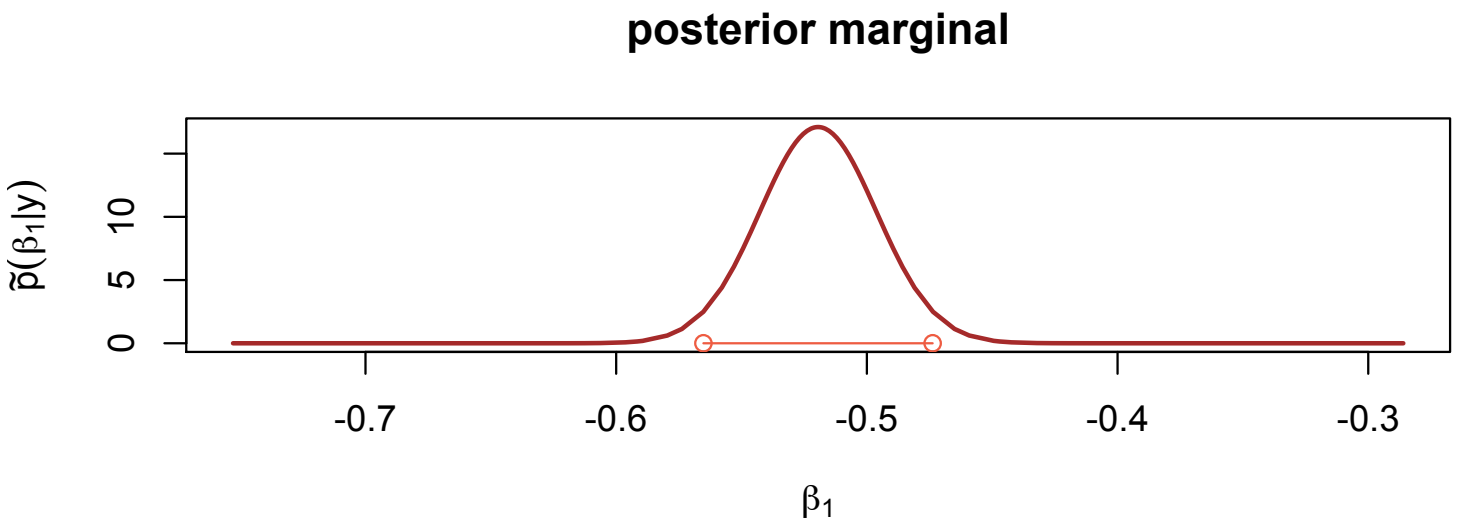
```
relazione = Petal.Length ~ Petal.Width
mB = inla(formula, family = gaussian)
summary(mB)
```

```
Fixed effects:
               mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)  4.0879 0.1414     3.8102   4.0879     4.3656  4.0878   0
Time        -0.5194 0.0233    -0.5652  -0.5194    -0.4736 -0.5195   0

Random effects:
Name    Model
 Subject   IID model

Model hyperparameters:
                                          mean     sd 0.025quant 0.5qu
Precision for the Gaussian observations 0.5377 0.0194     0.5004   0.5
Precision for Subject                   0.2063 0.0187     0.1717   0.2
```

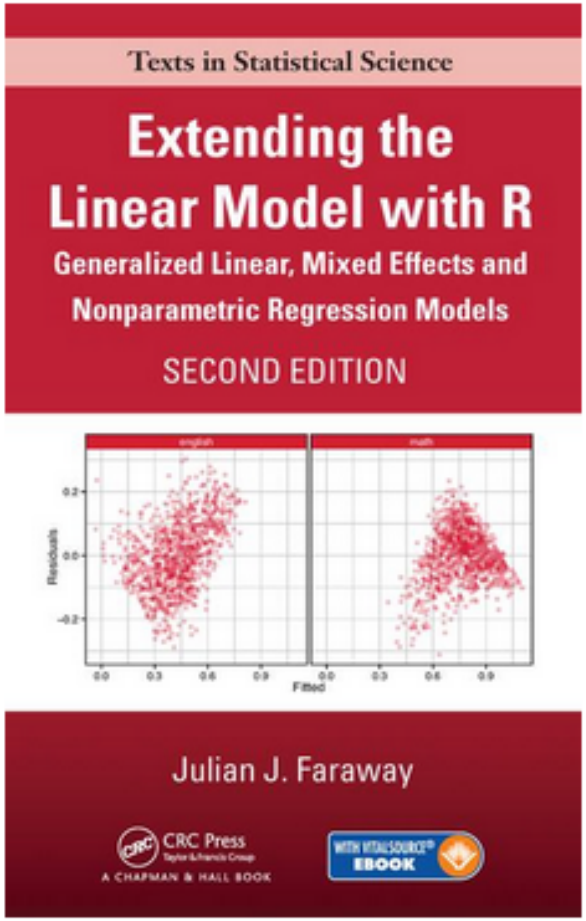
great advantage

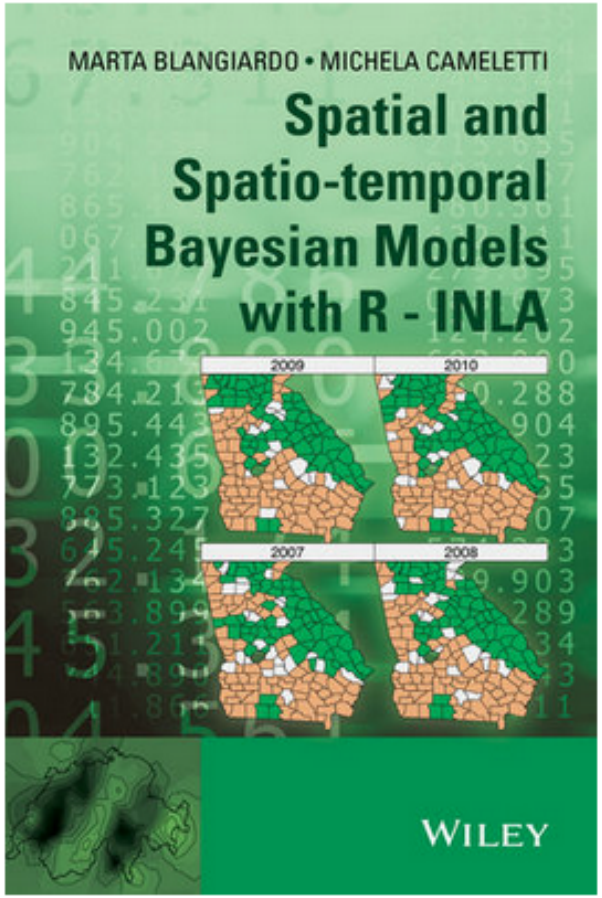



```
Fixed effects:
              mean     sd 0.025quant 0.5quant 0.975quant
(Intercept)  4.0879 0.1414     3.8102   4.0879     4.3656  4
Time        -0.5194 0.0233    -0.5652  -0.5194    -0.4736 -0

Random effects:
Name    Model
```

some textbooks





ringraziamenti

Massimo Borelli
borelli@units.it

www.dmi.units.it/borelli/