Some Statistical Issues in Reproducible Research

Donald Berry
M.D. Anderson Cancer Center

dberry@mdanderson.org

1

Outline

- Regression to the mean
- Multiplicities
  - BIG DATA
  - Silent multiplicities
- The process of learning given a positive experimental result
- Recommendations for investigators and journals

2

Outline

- **Regression to the mean**
- Multiplicities
  - BIG DATA
  - Silent multiplicities
- The process of learning given a positive experimental result
- Recommendations for investigators and journals

3

PLOS | MEDICINE

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • DOI: 10.1371/journal.pmed.0020124

Abstract

Summary

There is increas
the number of ot
framework, a res
and lesser presel
and other interes
designs and setti
simply accurate f

THE NEW YORKER

ANNALS OF SCIENCE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

BY JONAH LEHRER

DEC

O

in a
start
atyp
on t
bran

been tested on schizophrenics in several large clinical trials, all of which had demonstrated a dramatic decrease in the subjects' psychiatric symptoms. As a result, second-generation antipsychotics had become one of the fastest-growing and most profitable pharmaceutical classes. By 2001, Eli Lilly's Zyprexa was generating more revenue than Prozac. It remains the company's top-selling drug.

The Economist World politics Business & finance Economics Science & technology Cul

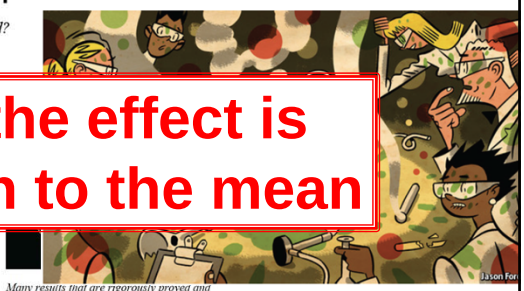
Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

Like 20k Tweet 1,984

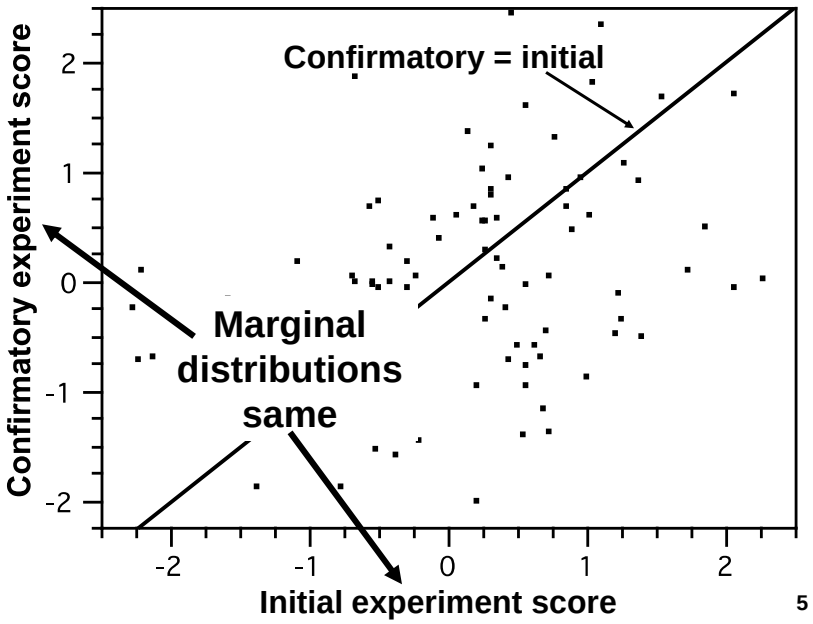


Many results that are rigorously proved and accepted start shrinking in later studies.

Most of the effect is regression to the mean

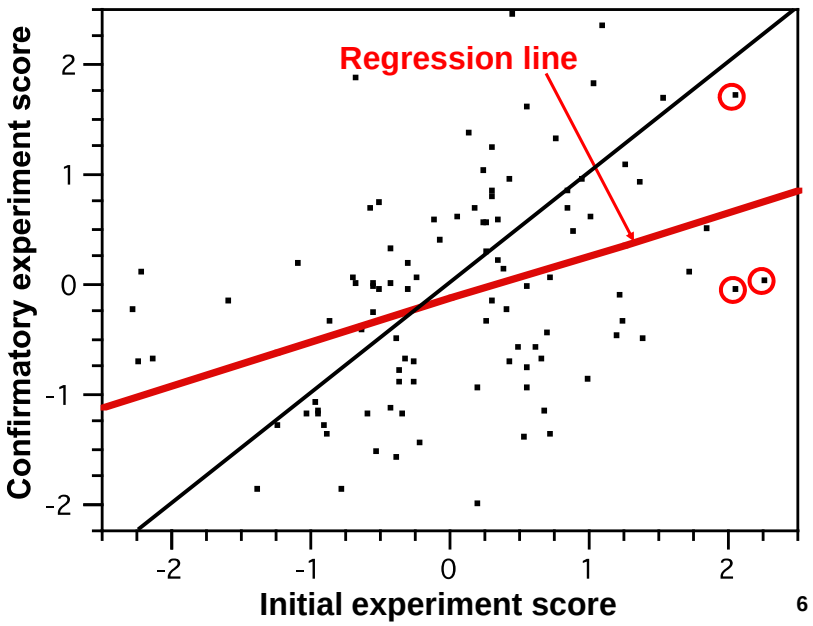
4

Variability in Experiments, but Correlation



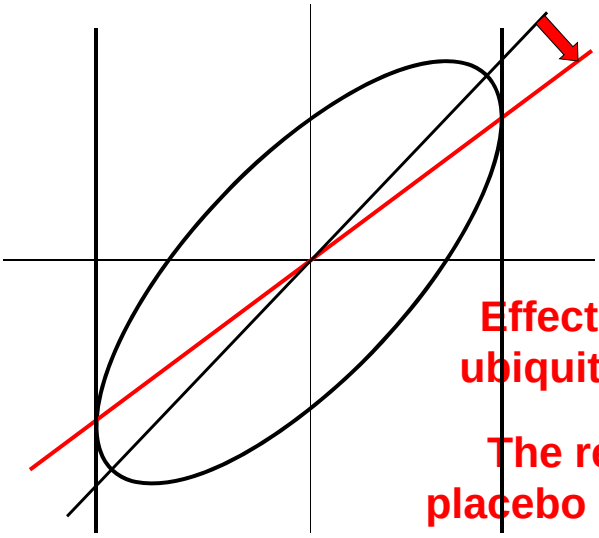
5

Suppose correlation coefficient = 0.50



6

Regression to the mean



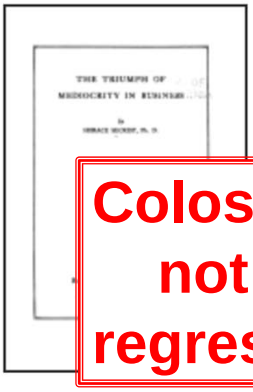
Effect is ubiquitous

The real placebo effect

7

Google Books

The Triumph of Mediocrity in Business

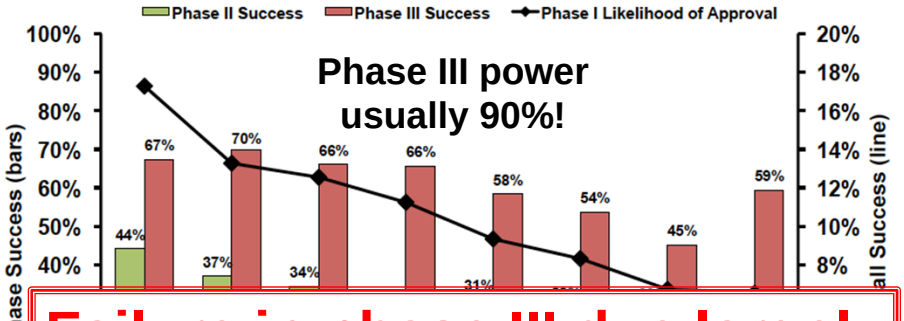


Horace Secrist

Bureau of Business Research, Northwestern University, 1933 - Business - 468 pages

Colossal failure due to not understanding regression to the mean

SUCCESS AT PHASE II AND III



Failure in phase III due largely to extrapolating from a biased selection of phase II trials: Regression to the Mean

biomed tracker

Jan 2003 through Aug 2012

BIO BIOTECHNOLOGY INDUSTRY ORGANIZATION

Sometimes easy to tell when not reproducible

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

NXY-059 for Acute Ischemic Stroke

Kennedy R. Lees, M.D., Justin A. Zivin, M.D., Tim Ashwood, Ph.D., Antonio Davalos, M.D., Stephen M. Davis, M.D., Hans-Christoph Diener, M.D., James Grotta, M.D., Patrick Lyden, M.D., Ashfaq Shuaib, M.D., Hans-Göran Hårdemark, M.D., and Warren W. Wasiewski, M.D., for the Stroke–Acute Ischemic NXY Treatment (SAINT I) Trial Investigators*

10

Table 2. Efficacy of the Study Drug at Day 90 or at the Last Rating.*

Outcome Variable	Placebo Group	NXY-059 Group	Difference between NXY-059 and Placebo†	P Value
			% or score (95% CI)	
Modified Rankin scale score (primary end point)				
No. of patients	849	850		
Score — no. (%)				
0	93 (11.0)	131 (15.4)	4.4	
1	170 (20.0)	153 (18.0)	−2.0	
2	99 (11.7)	97 (11.4)	−0.3	
3	108 (12.7)	121 (14.2)	1.5	
4	175 (20.6)	144 (16.9)	−3.7	
5 (or death)	204 (24.0)	204 (24.0)	0	0.038
Change from baseline in total NIHSS score (coprimary outcome)				
No. of patients	851	851		
Score — LSM ±SE	−1.7±0.5	−1.8±0.5	−0.1 (−1.4 to 1.1)	0.86
Barthel index (dichotomized analysis)				
No. of patients	848	850		
Score, ≥95 — no. (%)	346 (40.8)	368 (43.3)	2.5	0.14
Stroke Impact Scale				
No. of patients	676	669		
Score — LSM ±SE	63.4±1.1	66.2±1.1	2.8 (−0.3 to 5.9)	0.08
EuroQoL EQ-5D (weighted index)				
No. of patients	816	819		
Score — LSM ±SE	0.43±0.013	0.47±0.013	0.04 (−0 to 0.07)	0.06
EuroQoL EQ-5D (VAS)				
No. of patients	671	670		
Score — LSM ±SE	62.0±0.9	64.5±0.9	2.5 (−0.1 to 5.0)	0.05

Odds ratio 1.20
p = 0.038

11

SAINT II:

- N = 3200 (up from 1700)
- Power 80% for Odds Ratio 1.20
- My probability that SAINT II would be positive: 10%

12

Press Release

"SAINT II did not meet its primary outcome (p=0.33, odds ratio 0.94) compared to placebo."

"The company plans no further development of NXY-059 in acute ischemic stroke."

13

Multiplicities

- BIG DATA:
  - False-positives galore
  - GWAS standard: $p < 5 \times 10^{-8}$
  - False-negative rate is huge; too high a price to pay?
- Statistical analyses are important, but not the main problem
  - Most important analyses sometimes happen before I see the data

14

Outline

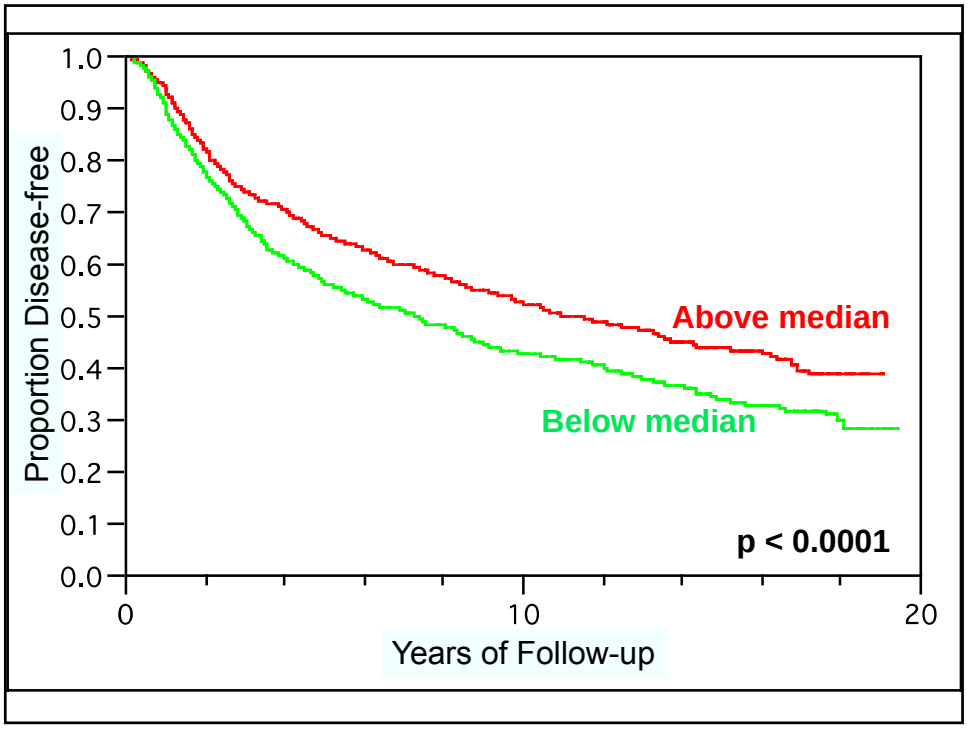
- Regression to the mean
- Multiplicities
  - BIG DATA
  - Silent multiplicities
- The process of learning given a positive experimental result
- Recommendations for investigators and journals

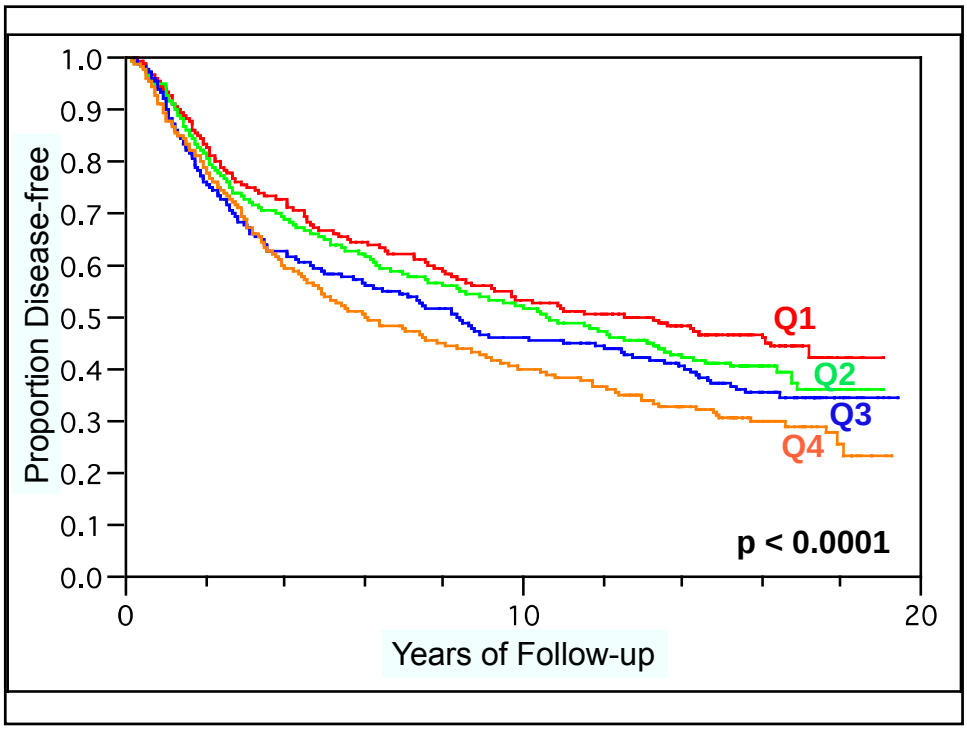
15

Multiplicities and building a prognostic -omic index

- 1550 node-positive BC patients
- 20 markers
- Select markers with p < 0.10
- Build Recurrence Score using multivariate Cox regression

16


Proportion Disease-free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
Years of Follow-up
Above median
Below median
p < 0.0001


Proportion Disease-free
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
Years of Follow-up
Q1
Q2
Q3
Q4
p < 0.0001

The punchline:

All 20 markers were white noise

19

The good news:

I had a protocol!

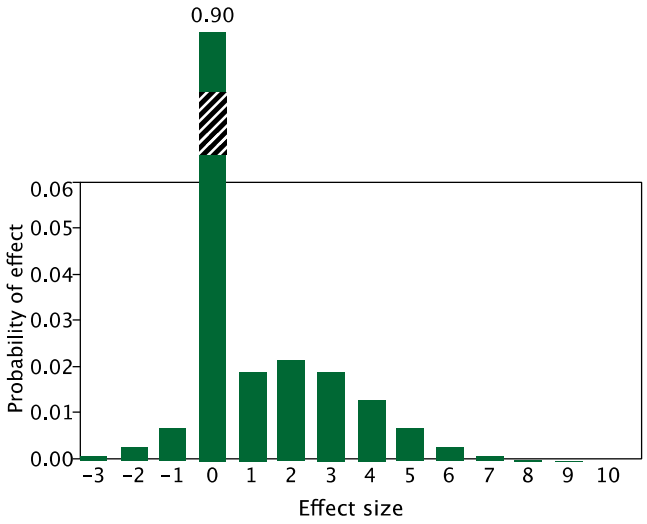
20

Outline

- Regression to the mean
- Multiplicities
  - BIG DATA
  - Silent multiplicities
- **The process of learning given a positive experimental result**
- Recommendations for investigators and journals

21

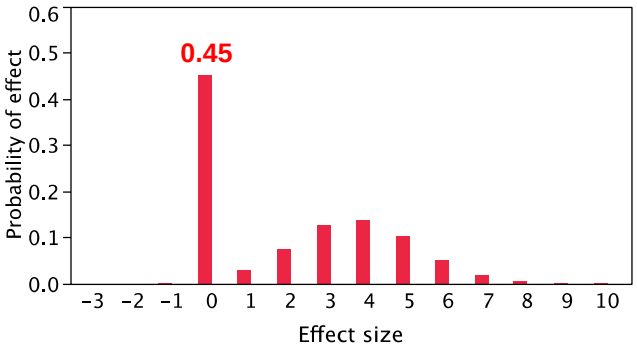
Prior probabilities of effect size



22

Probability that modestly powered experiment is positive: 5%

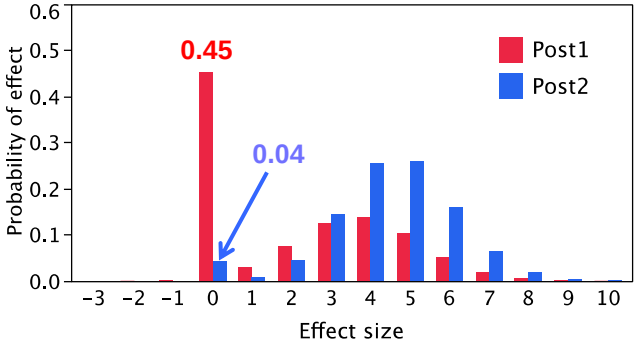
If positive, new probs of effect size:



23

Probability that confirmatory experiment is positive: 27%

If 2nd positive, new probs of effect size:



24

Multiplicities in Cancer Research: Ubiquitous and Necessary Evils

JNCI 2012

Donald Berry

Manuscript received October 4, 2011; revised January 27, 2012; accepted May 31, 2012.

Recommendations for Investigators (and Journals)

The following are r
regarding publishing
principles. One is to
thing you planned t
need to address sile
made everything kn

1. *Need for a proto*
study. The prot
ods. It should s
possible results
etc. It should
handled. These
analyses, the va
cal calculations
2. *Analyses not do*
endpoints or s
not included in
3. *Log of actual an*
not specified in
whether or not
ficiently detail
The cover lette
this log and that it is available to reviewers upon request.
4. *Unspecified analyses.* Discoveries from analyses not specified in

7. *Piecemeal publication.* Some studies are published piecemeal. In each publication, the full study and previous publications and
uld be described.
e onus for adjust-
ion multiplicities.
viously, it is best
for any empirical
curred before the
f retrospectively
cluding when the
biological mech-
of independent
roups.
s are acceptable.
e differences in
cess, but the basic
unusual observa-
udes to confirm-
ies.
ies, follow the
ove.
Researchers must
us consequences,
including understanding that multiplicities are sometimes silent. The rub is that progress in personalized medicine requires exposing

1. Need for a protocol.
2. Analyses not done.
3. Log of actual analyses.
4. Unspecified analyses.
5. To adjust for multiplicities?
6. Confirmation.
7. Piecemeal publication.
8. Biological rationale.
9. Frequentist versus Bayesian.
10. REMARK (for biomarkers).

Outline

- Regression to the mean
- Multiplicities
  - BIG DATA
  - Silent multiplicities
- The process of learning given a positive experimental result
- Recommendations for investigators and journals

26