Center for Stroke Research Berlin

Non-reproducibility is an integral part of a productive and efficient scientific enterprise

- A key task for biomedical research is to scan a large number of hypotheses
- Efficient scanning of a large number of candidates requires fast and frugal methodologies, like use of small sample sizes and surrogate endpoints.
- By their very nature, however, fast and frugal methods are prone to producing false positive signals of activity.
- Drugs that look promising in preclinical studies but fail conceptual reproduction in phase 2 trials are to some degree a byproduct of rational decision-making across a sequence of investigations rather than poor study design.
- Failed conceptual reproduction maps the boundaries of a theory.
- Observation of phenomena in one experimental system, but not in another, often leads to deeper understandings of processes and of experimental systems. It defines the boundaries of a theory's explanatory power.

see also: London and Kimmelman. Why clinical translation cannot succeed without failure eLife 2015;4:e12844.

Center for Stroke Research Berlin

'Replication' - Speed dating

- Falsch Positive und Inflation von Effektgrössen
- Publikationsbias
- Exploration vs. Konfirmation
- Gründe für 'Nicht-Replizierbarkeit'
- Probleme
- Quo vadis

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- Fraud
- Niedrige prior probabilities (=low base rates)
- Bias
- p-Hacking
- Hypothesizing after the results are known
- Verwechslung von p-Wert mit positiv prädiktivem Wert
- Power niedrig
- 'Winner's curse'
- Regression to the mean

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- Betrug

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- Niedrige prior probabilities (=low base rates)

The low 'low hanging fruits' are already harvested....

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- Bias

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- p-Hacking

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2012, i*First*, 1–9

Psychology Press
Taylor & Francis Group

A peculiar prevalence of *p* values just below .05

E. J. Masicampo[1], and Daniel R. Lalande[2]

[1]Department of Psychology, Wake Forest University, Winston-Salem, NC, USA
[2]Department of Health Sciences, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- Hypothesizing after the results are known ('Hark-ing')

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- Verwechslung von p-Wert mit positiv prädiktivem Wert

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

Positive predictive value (PPV)

$$PPV = ([1 - \beta] \times R) / ([1 - \beta] \times R + \alpha)$$

PPV = (Power x Prestudy odds) / (Power x prestudy odds + alpha)

power	0,5
prestudy odds	0,2
alpha	0,05
PPV	0,66

power	0,8
prestudy odds	0,2
alpha	0,01
PPV	0,94

Center for Stroke Research Berlin

Falsch Positive und Inflation der Effektgrösse

- Power niedrig

Power failure: why small sample size undermines the reliability of neuroscience

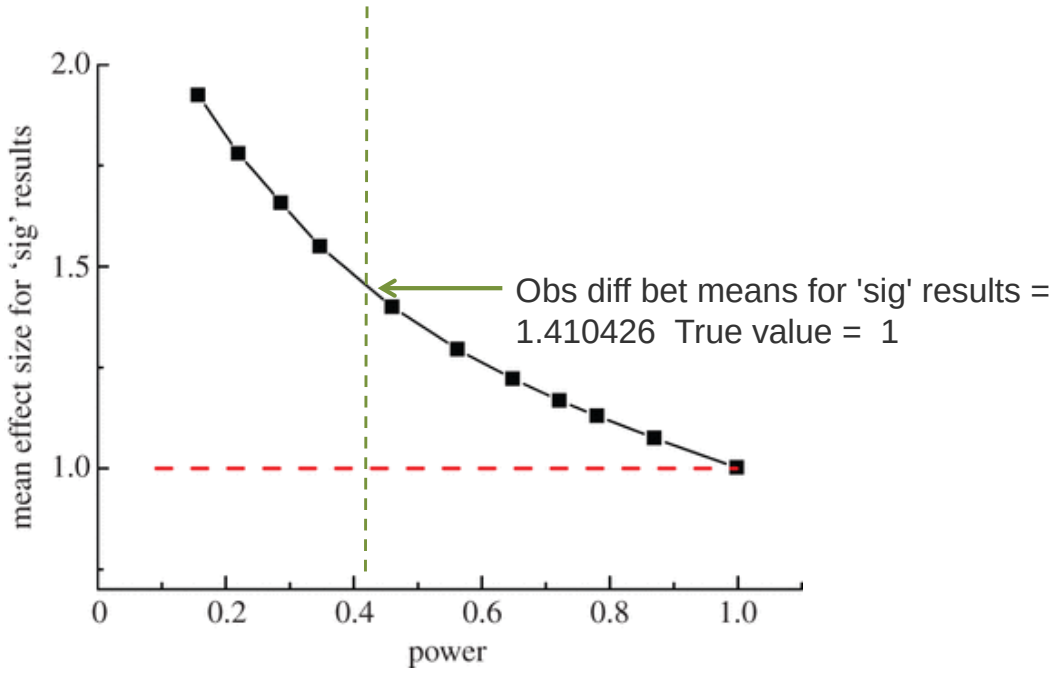
Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

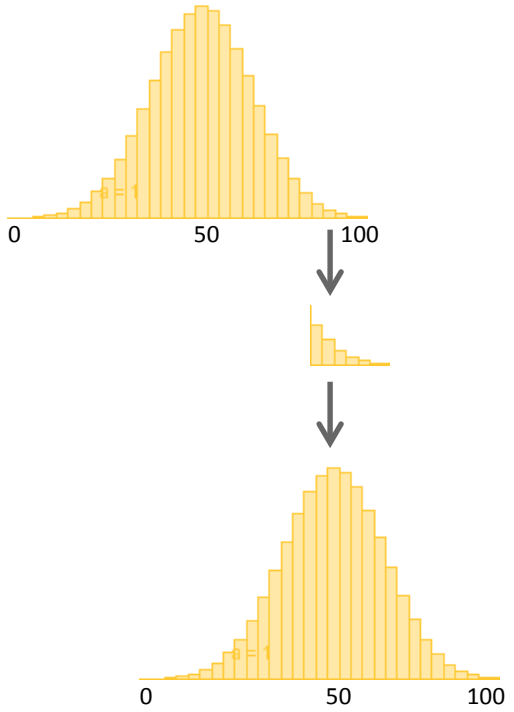
Nature Reviews Neuroscience | AOP, published online 10 April 2013; doi:10.1038/nrn3475

Overall median power of 730 primary neuroscience studies: 21 %

Falsch Positive und Inflation der Effektgrösse

- 'Winner's curse' , Regression to the mean





Center for Stroke Research Berlin

Publication bias

"Publication bias is highly prevalent (present in the literature describing the efficacy of at least 16 of 18 interventions) and accounts for around 30% of the reported efficacy of candidate neuroprotective interventions."

Publication bias in reports of animal stroke studies leads to major overstatement of efficacy

Emily S Sena, H. Bart van der Worp, Philip M.W. Bath, David W Howells, and Malcolm R Macleod (PLoS Biol. 2010 Mar 30;8(3):e1000344)

Center for Stroke Research Berlin

Exploratory vs confirmatory research

OPEN ACCESS Freely available online

PLOS | BIOLOGY

Perspective

Distinguishing between Exploratory and Confirmatory Preclinical Research Will Improve Translation

Jonathan Kimmelman[1]*, Jeffrey S. Mogil[2], Ulrich Dirnagl[3,4,5]

Center for Stroke Research Berlin

Exploratory vs confirmatory research

In **exploratory investigation**, researchers should aim at generating robust pathophysiological theories of disease.

In **confirmatory investigation**, researchers should aim at demonstrating strong and reproducible treatment effects in relevant animal models.

Protocols and publications should pre-specify whether they are ‘‘exploratory’’ or ‘‘confirmatory’’ studies, with the latter category reserved for studies that aim at demonstrating promise of clinical utility for an intervention.

Center for Stroke Research Berlin

Exploratory vs confirmatory research

	Exploratory 'Discovery'	Confirmatory
Hypothesis	(+)	+++
Establish pathophysiology	+++	(+)
Sequence and details of experiments established at onset	(+)	+++
Primary endpoint	-	++
Sample size calculation	(+)	+++
Blinding	+++	+++
Randomization	+++	+++
External validity (aging, comorbidities, etc.)	-	++
In/Exclusion criteria	++	+++
Test statistics	+	+++
Preregistration	(-)	++
Find what might work: Sensitivity (accept Type I errors)	+++	+
Weed out false positives: Specificity (accept Type II errors)	+	+++

Center for Stroke Research Berlin

Ursachen für Nichtreplizierbarkeit

- Fraud
- False positive (Replication = true Null result)
- Methodical (Incompetence, biased study design, low power, etc.)
- Poor reporting (incl. 'tacit knowledge')
- Random variation

Center for Stroke Research Berlin

Replicating a statistically significant finding

$p = 0.049$ ($p < \alpha = 0.05$)

Assume that the experimental result is correct, i.e. measured difference equals (unknown) treatment effect.

Repeat experiment under identical conditions.

What is the probability to replicate the significant findings?

50 %!

Center for Stroke Research Berlin

Randomisierte multicenter präklinische Studien

RESEARCH ARTICLE

STROKE

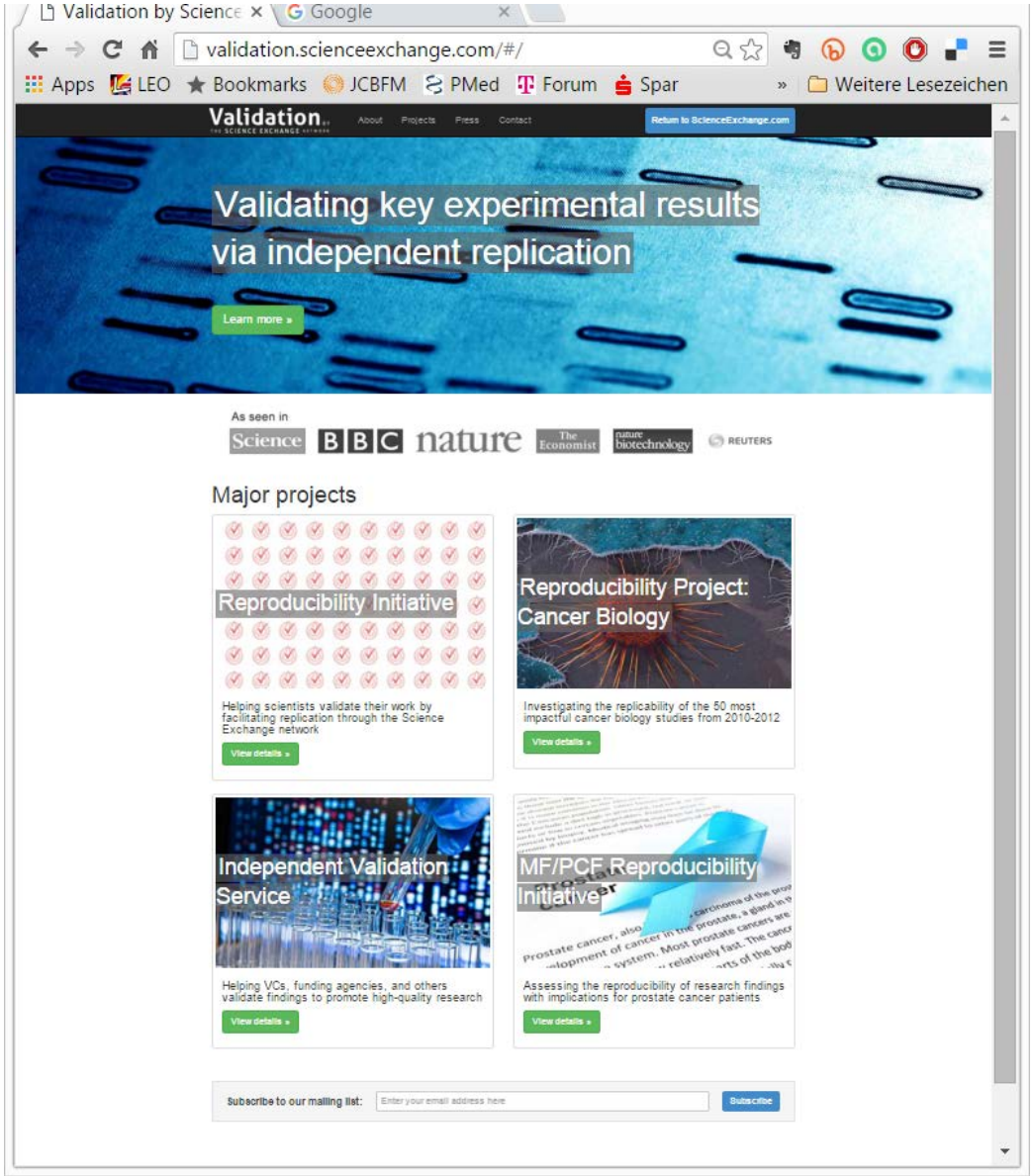
Results of a preclinical randomized controlled multicenter trial (pRCT): Anti-CD49d treatment for acute brain ischemia

Gemma Llovera,[1,2] Kerstin Hofmann,[1,2] Stefan Roth,[1,2] Angelica Salas-Pérdomo,[3,4] Maura Ferrer-Ferrer,[3,4] Carlo Perego,[5] Elisa R. Zanier,[5] Uta Mamrak,[1,2] Andre Rex,[6] Hélène Party,[7] Véronique Agin,[7] Claudine Fauchon,[8] Cyrille Orset,[7,8] Benoît Haelewyn,[7,8] Maria-Grazia De Simoni,[5] Ulrich Dirnagl,[6] Ulrike Grittner,[9] Anna M. Planas,[3,4] Nikolaus Plesnila,[1,2] Denis Vivien,[7,8] Arthur Liesz[1,2]*

Science Translational Medicine
AAAS

Center for Stroke Research Berlin

Strukturierte Initiativen



Center for Stroke Research Berlin

Probleme

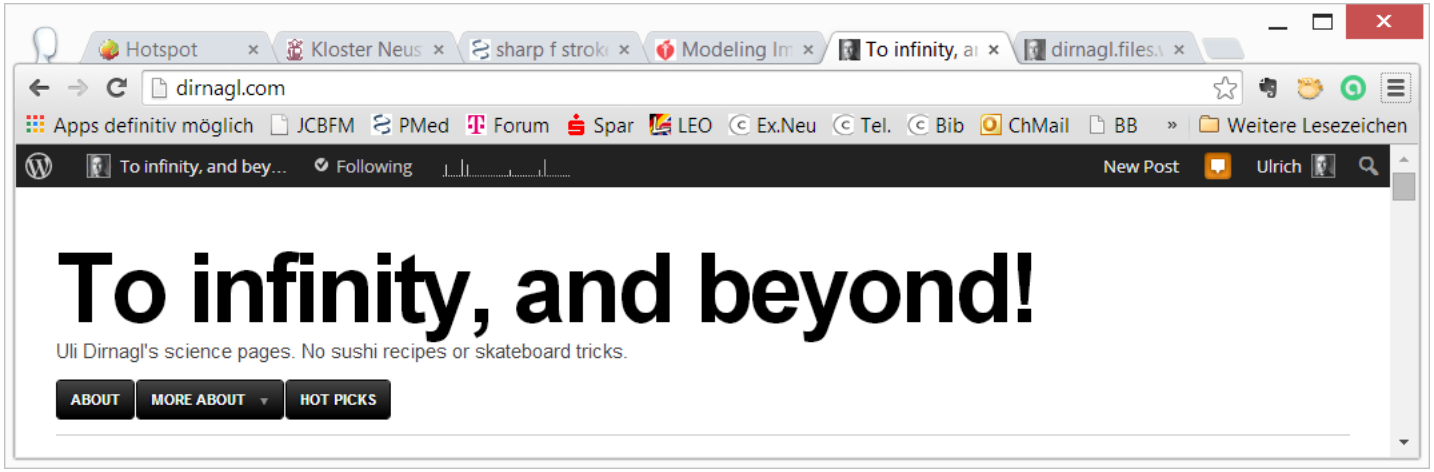
- Finanzierung
- Fehlende Incentivierung; Kultur (Replikation nicht gleich Innovation)
- Regulatorisch (Erkenntnisgewinn?)

Center for Stroke Research Berlin

Quo vadis?

- Methodischer Rigor (Ausbildung, auch von PIs!)
- Präregistrierung
- Unterscheidung Exploration / Konfirmation
- Replikation 'kritischer' Befunde
- Internationale Initiativen
- Funding
- Wertschätzung ('Incentivierung')

http://bit.ly/dirnaglreplication

Button Power failure Nat Neuros...	Adobe Acrobat-Dokument	515 KB
Colquhoun False dicsovery rate ...	Adobe Acrobat-Dokument	383 KB
Dirnagl Sind die meisten publizi...	Adobe Acrobat-Dokument	1.387 KB
Dirrnagl DFG Replication Slides....	Adobe Acrobat-Dokument	1.840 KB
Good Institutional Practice Natu...	Adobe Acrobat-Dokument	584 KB
Ioannidis Why most published r...	Adobe Acrobat-Dokument	316 KB
Kimmelman Mogil Dirnagl Expl...	Adobe Acrobat-Dokument	89 KB
Open Science Collaboration Psy...	Adobe Acrobat-Dokument	559 KB

http://dirnagl.com