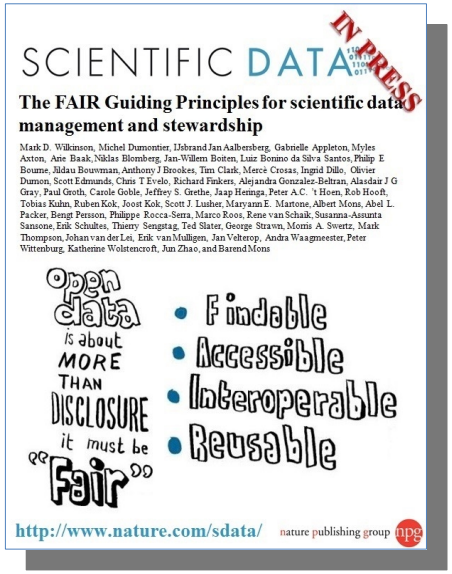
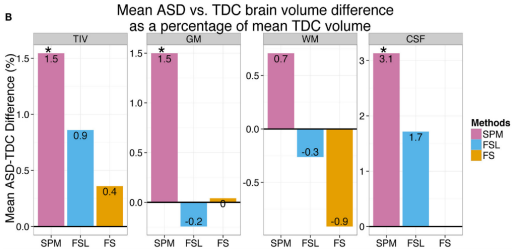
Reproducible, efficient and reusable neuroimaging research

Jean-Baptiste Poline

Henry Wheeler Jr. Brain Imaging Center
UC Berkeley

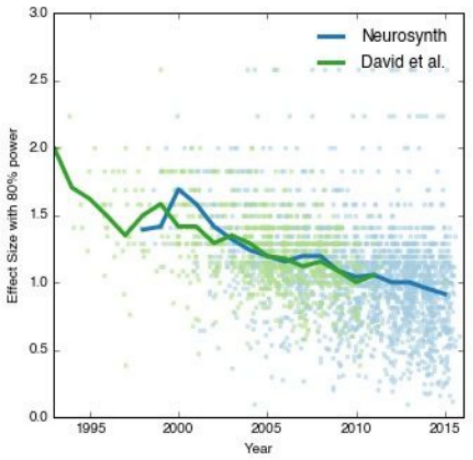


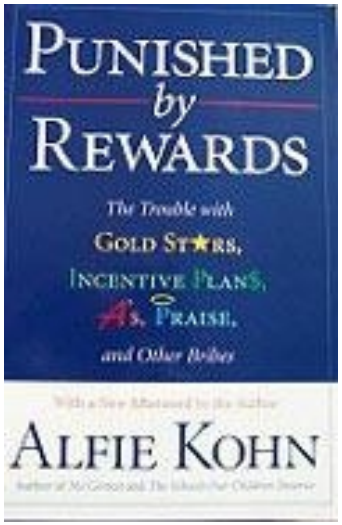


COMMENTARY

Toward standard practices for sharin computer code and programs in neuroscience

Stephen J Eglen[1], Ben Marwick[2], Yaroslav O Halchenko[3], Michael Hanke[4,5], Shoaib Sufi[6], Padraig Gleeson[7], R Angus Silver[7], Andrew P Davison[8], Linda Lanyon[9], Mathew Abrams[9], Thomas Wachtler[10], David J Willshaw[11], Christophe Pouzat[12] & Jean-Baptiste Poline[13]





Data **Code** **Statistics** **Culture**

Outline

- Reproducibility issues: examples in bio-sciences, psychology and in neuroimaging
- Causes:
  - Data, Informatics, Statistics, Cultural
- Solutions:
  - Some directions
- Conclusions

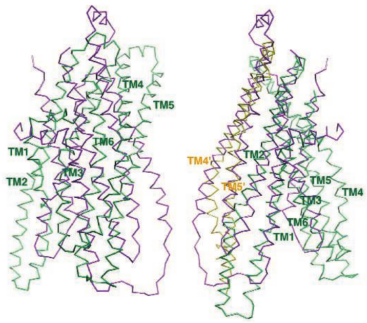
Some remarkable examples

- **In Biology**
  - Amgen replication
  - Protein structure flip
  - Forensic analysis: data row shift
  - HeLa contamination of cell lines
- In Neuroimaging:
  - Imaging genetics
  - Flexibility in analysis
  - Replication of neuroimaging study
- (Psychology: Nosek et al)

Amgen replication

- **53** papers examined at Amgen in preclinical cancer research, Begley and Ellis, Nature, 2012 **findings were confirmed in only 6 (11%)**

- G. Chang: 3 Science, 1 PNAS, 1 J Mol Biol retracted



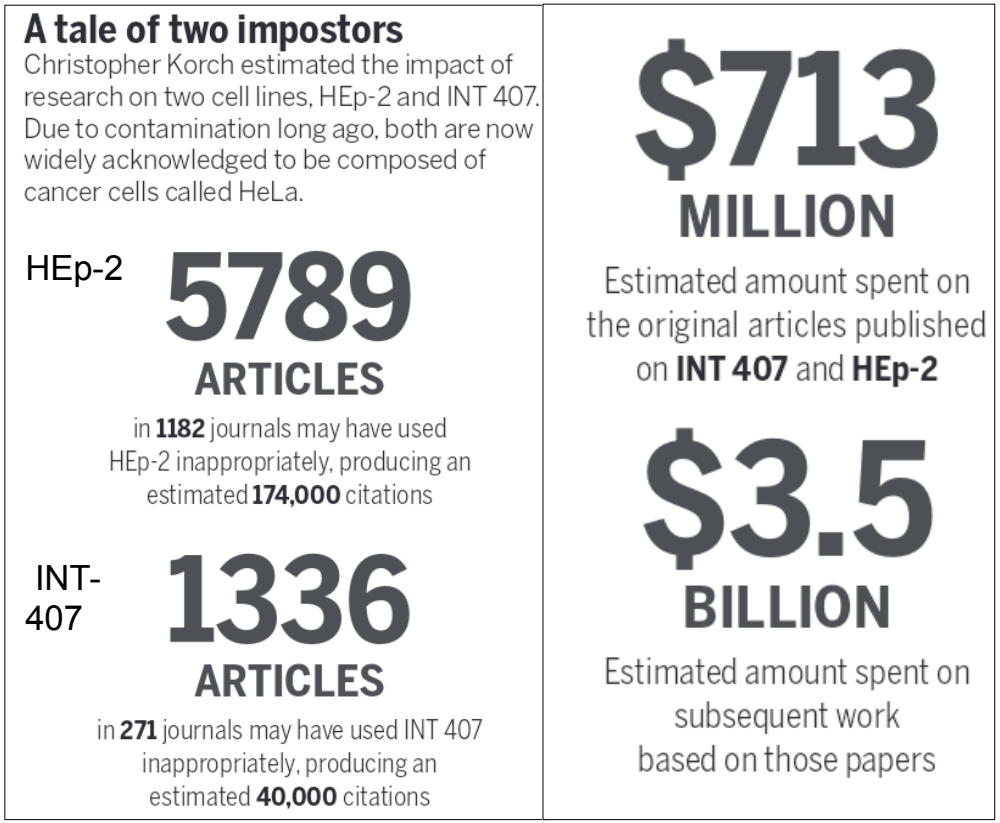
Flipping fiasco. The structures of MsbA (purple) and Sav1866 (green) overlap little (*left*) until MsbA is inverted (*right*).

- Baggerly and Coombes Forensic of Potti analyses:

  "with poor documentation and irreproducibility even well meaning investigator may argue for drug that are contraindicated to some patients"

HeLa and cell line contamination

- Cell lines used to explore basic questions, e.g. how cancers and normal tissues respond to drug - assumed to retain the properties of the original tissue
- Cell lines can be contaminated by other cells that outgrow the original cells
- A problem shown by Nelson-Rees (1970s), R Nardone (2007), C. Korch (2015)

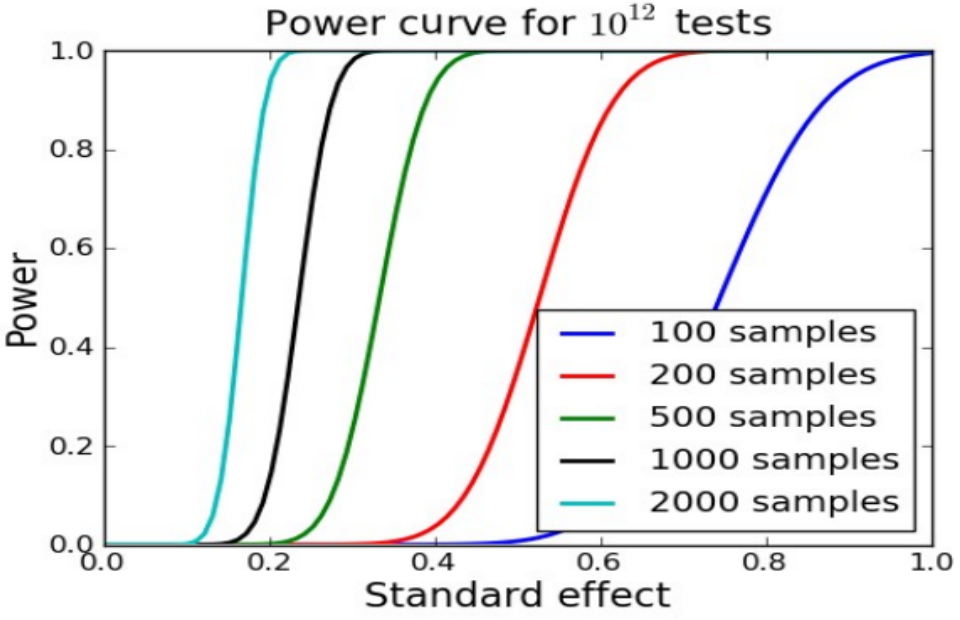


J. Neimark, Science, 2015, Korch 2015

Imaging Genetics GWAS

Stein et al., 2012, Nature Genetics, study of the hippocampal volume in more than 7+10k subjects

Previously identified candidate polymorphisms associated with hippocampal volume in general showed little association within our meta-analysis :(



1000 subjects, 80% power effect size needs to be > .26

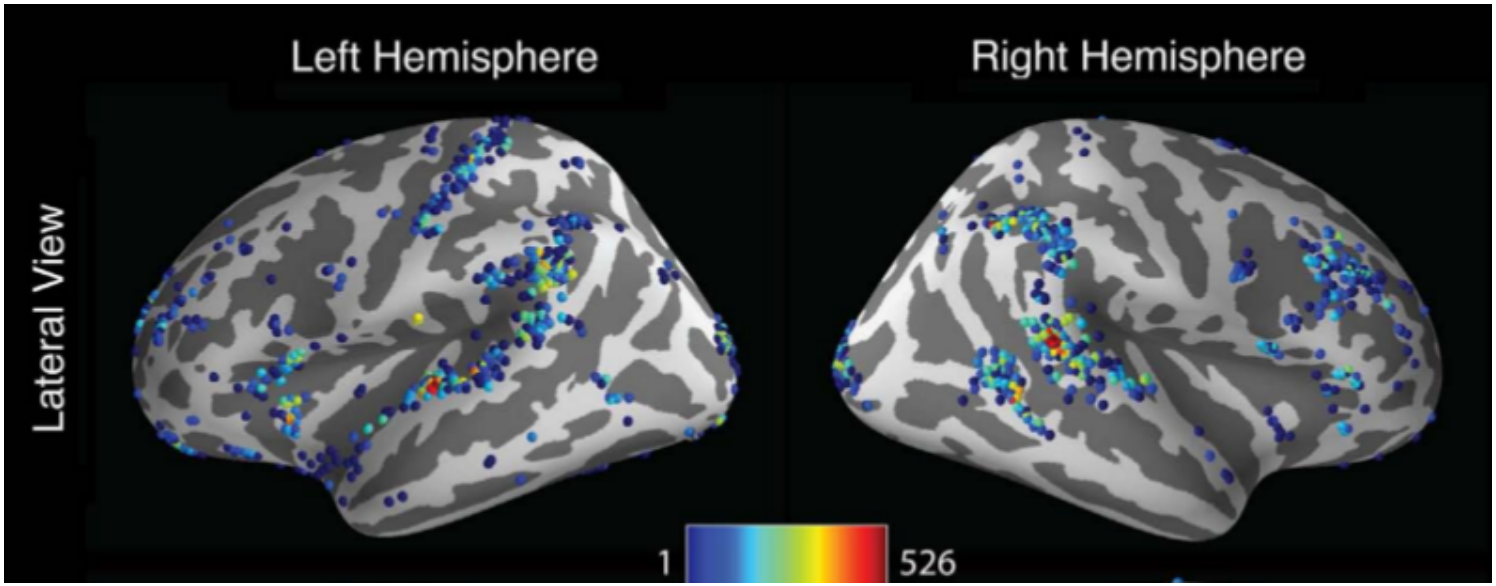
APOE : .22

Estimating analytic flexibility of fMRI

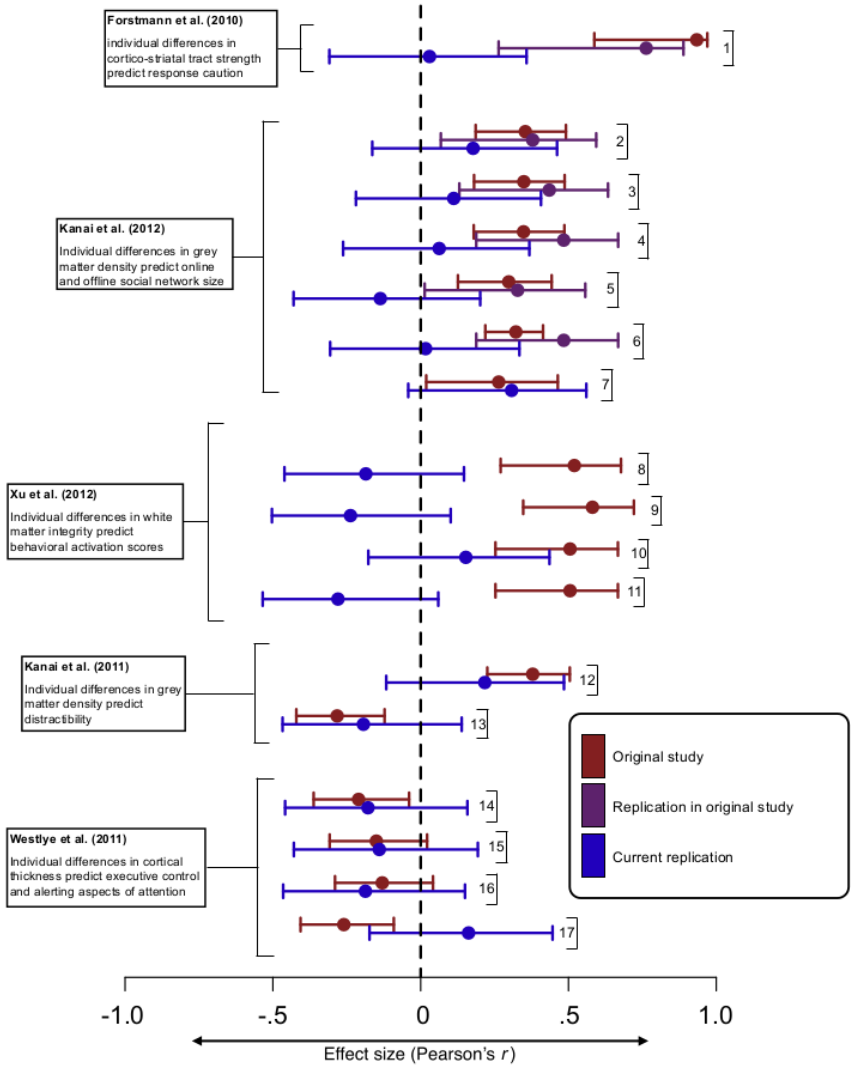
J. Carp, frontiers Neuroscience, 2012

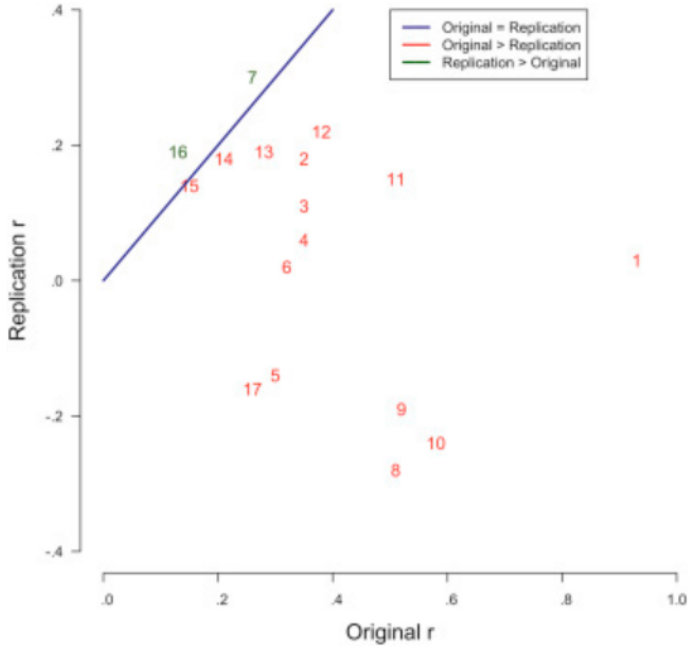
- A **single** event-related fMRI experiment to a large number of unique analysis procedures
- Ten analysis steps for which multiple strategies appear in the literature : **6,912 pipelines**

Maximum peak



Brain Behaviour Correlation





Boekel et al, Cortex, 2015

8 out of a total of 17 hypothesized effects were contradicted with moderate or strong levels of evidence

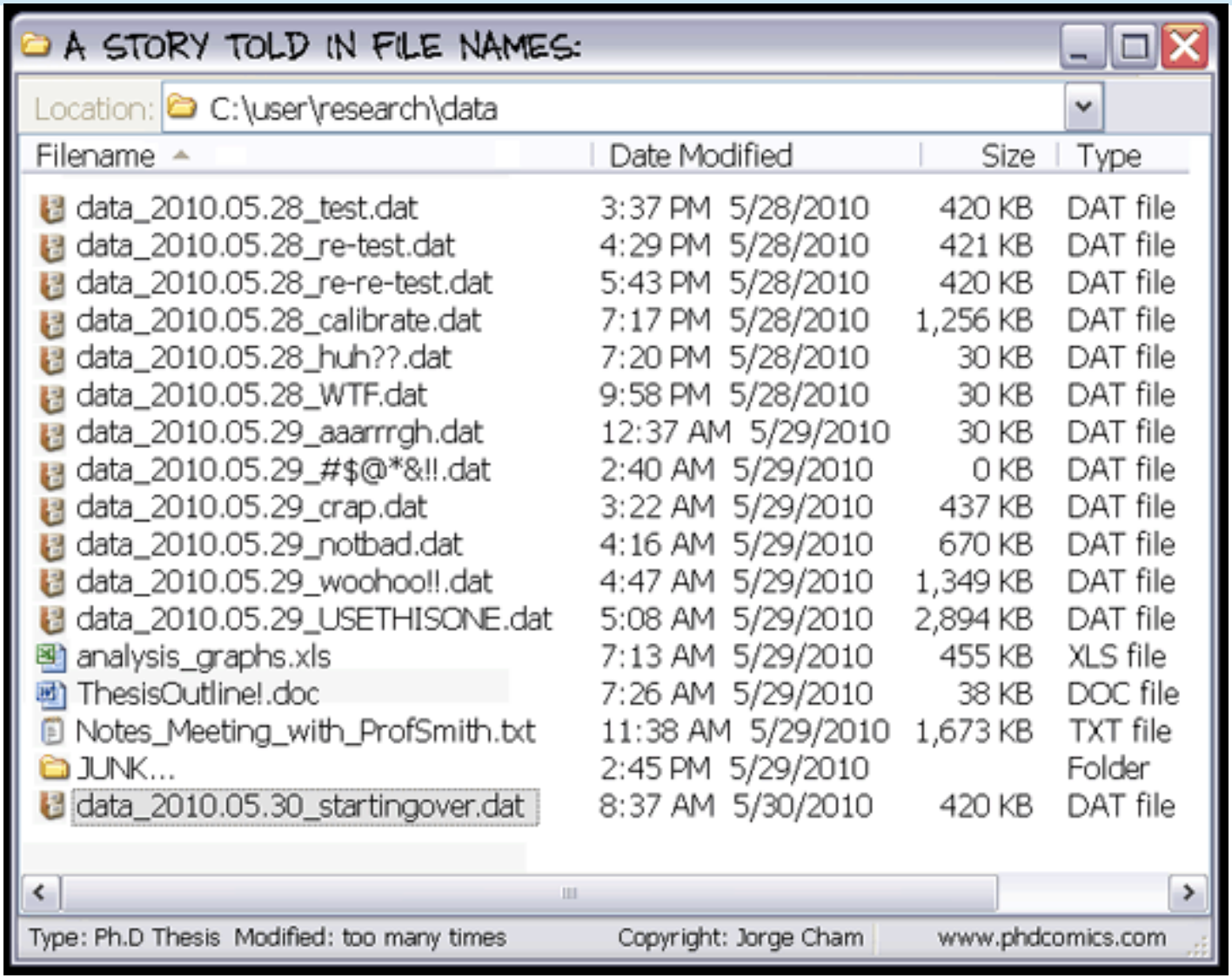
Reproducibility issues: Causes

- **Informatics: software and data**
- Statistics
  - P-value interpretation, P-hacking, File drawer
  - Power
- Social / cultural

Informatics Causes: software and data

- Data
  - Still massively not available
  - Data errors (eg excel spread sheet)
  - Data versioning
- Software
  - Most of our mainstream software package do not have unit testing / CI
  - Recent bugs (RFT limitation? AFNI ?)
  - Home made scripts

Data

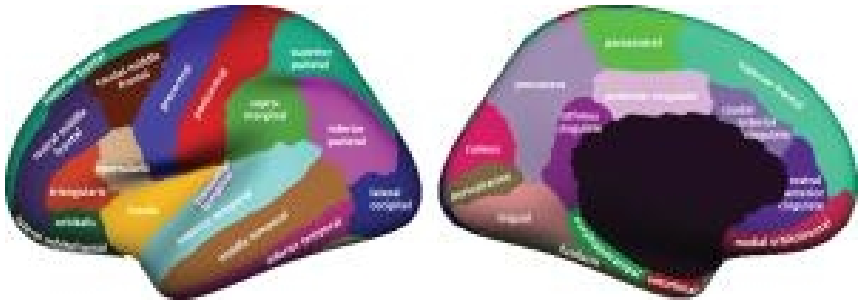

A STORY TOLD IN FILE NAMES:
Location: C:\user\research\data
Filename
Date Modified
Size
Type
data_2010.05.28_test.dat 3:37 PM 5/28/2010 420 KB DAT file
data_2010.05.28_re-test.dat 4:29 PM 5/28/2010 421 KB DAT file
data_2010.05.28_re-re-test.dat 5:43 PM 5/28/2010 420 KB DAT file
data_2010.05.28_calibrate.dat 7:17 PM 5/28/2010 1,256 KB DAT file
data_2010.05.28_huh??.dat 7:20 PM 5/28/2010 30 KB DAT file
data_2010.05.28_WTF.dat 9:58 PM 5/28/2010 30 KB DAT file
data_2010.05.29_aaarrrgh.dat 12:37 AM 5/29/2010 30 KB DAT file
data_2010.05.29_#$@*&!!.dat 2:40 AM 5/29/2010 0 KB DAT file
data_2010.05.29_crap.dat 3:22 AM 5/29/2010 437 KB DAT file
data_2010.05.29_notbad.dat 4:16 AM 5/29/2010 670 KB DAT file
data_2010.05.29_woohoo!!.dat 4:47 AM 5/29/2010 1,349 KB DAT file
data_2010.05.29_USETHISONE.dat 5:08 AM 5/29/2010 2,894 KB DAT file
analysis_graphs.xls 7:13 AM 5/29/2010 455 KB XLS file
ThesisOutline!.doc 7:26 AM 5/29/2010 38 KB DOC file
Notes_Meeting_with_ProfSmith.txt 11:38 AM 5/29/2010 1,673 KB TXT file
JUNK... 2:45 PM 5/29/2010 Folder
data_2010.05.30_startingover.dat 8:37 AM 5/30/2010 420 KB DAT file
Type: Ph.D Thesis Modified: too many times
Copyright: Jorge Cham
www.phdcomics.com

Software issues – misuse

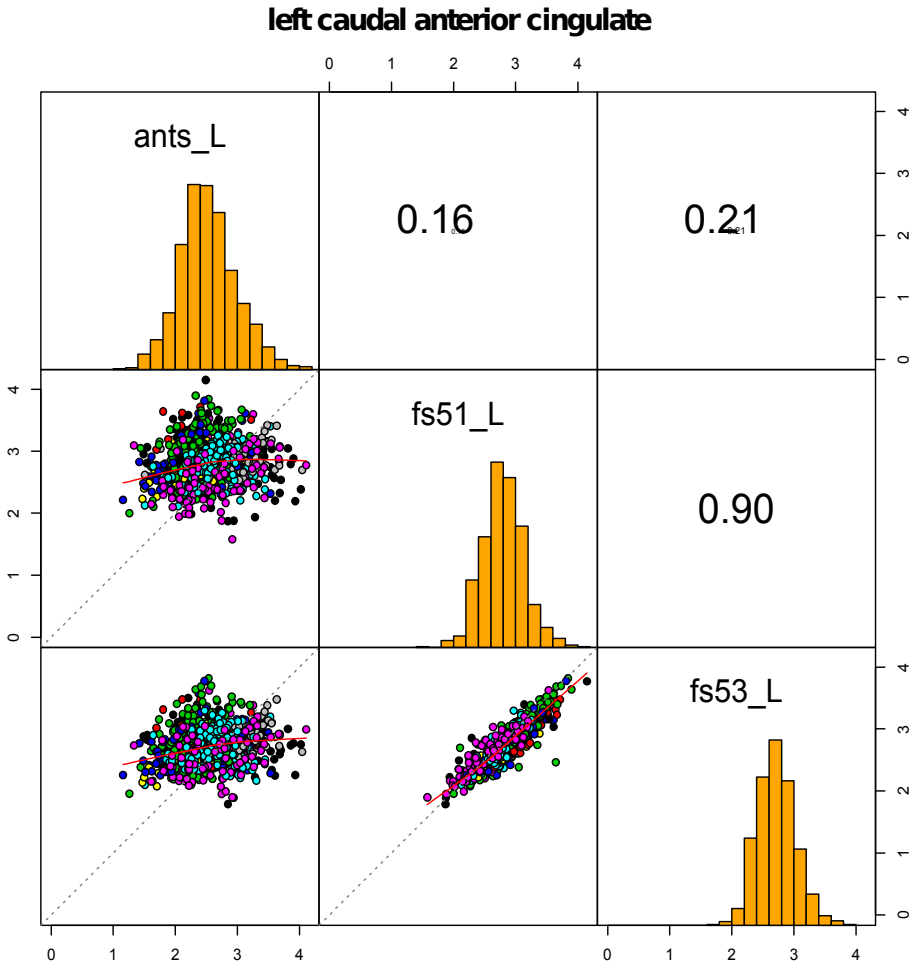
- 1990's: software industry realizes that:

  "**untested code is broken code**"
- The unit and integration testing framework started to be developed, coverage introduced
- Neuroimaging **software have bugs** – many **unknown?**
- How do you test the script that you inherit from the previous PhD/Postdoc ?

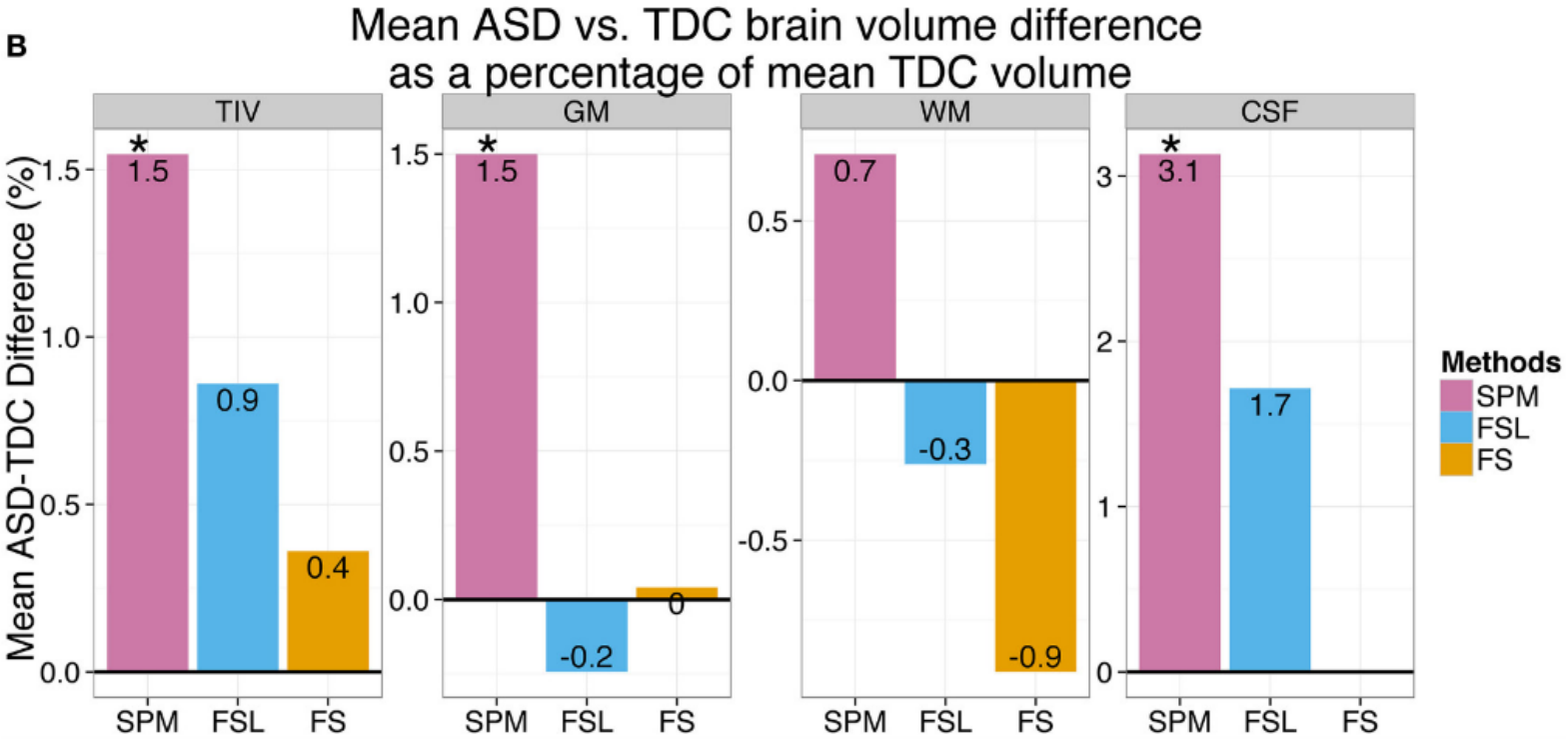
Tools matter: Ants, FS5.1, FS5.3



D.N. Kennedy, E. Dickie, S.M. Hodge, R.C. Craddock, J-B. Poline



Software, version, OS



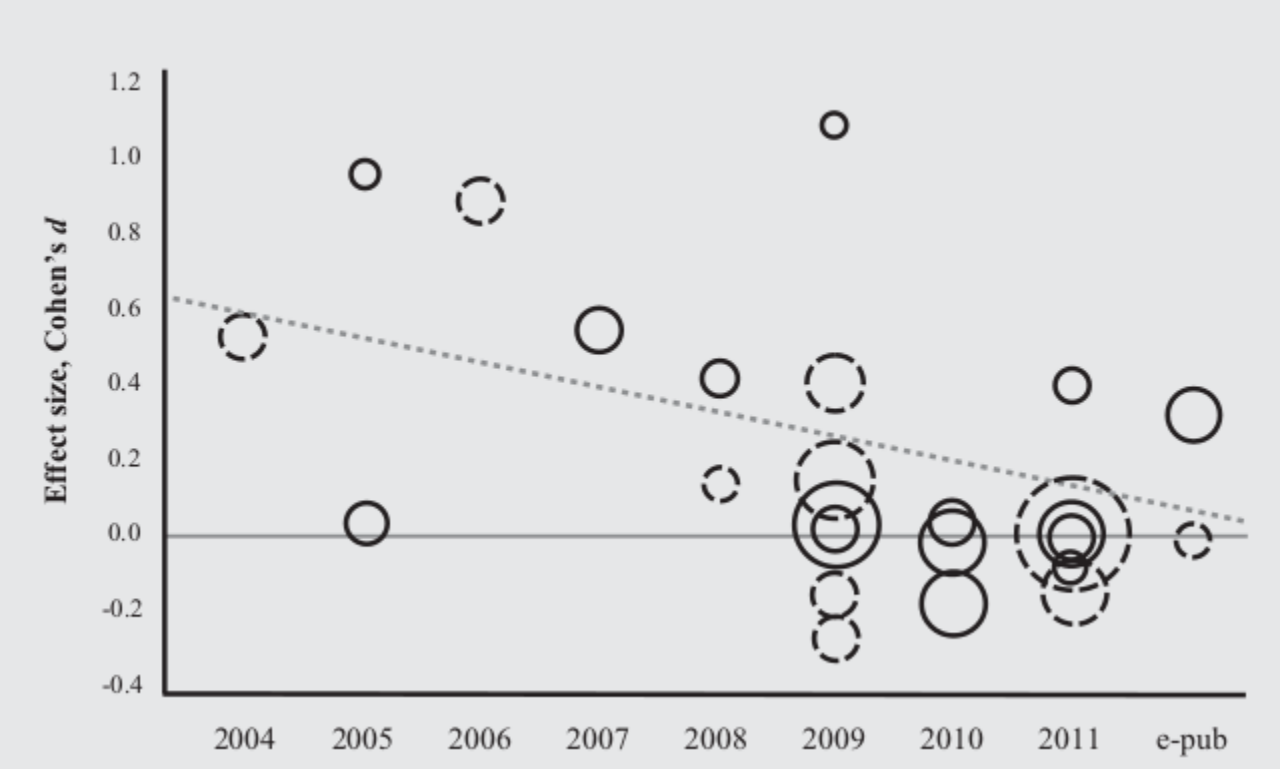
G. Katuwal, frontiers Brain Imaging Methods, 2016

- Change from FSL to SPM?
- Change from v.1.12 to v.2.1 ?
- Change from cluster A to cluster B? Cf Glatard et. al., frontiers, 2015

D. Donoho on software

"The scientific method's central motivation is the ubiquity of error - the awareness that mistakes and self-delusion can creep in absolutely anywhere and that the scientist's effort is primarily expended in recognizing and rooting out error."
David Donoho et al. (2009)

Statistical causes: evil p-value?



Molendijk, 2012: BDNF and hippocampal volume

See also : Mier, 2009: COMT and DLPFC

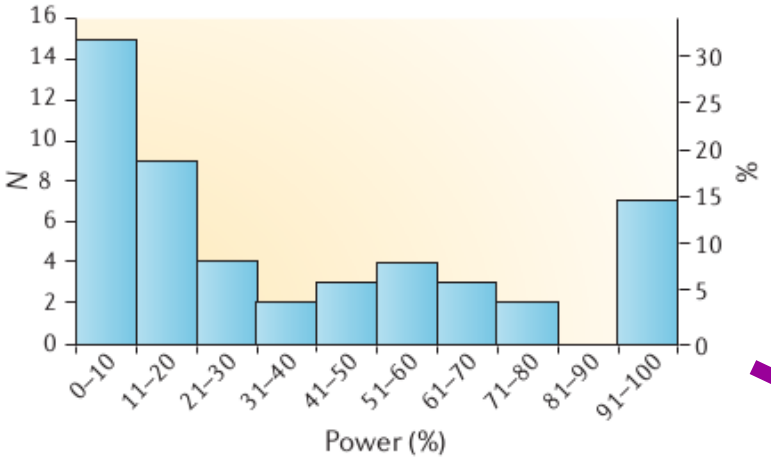
Power issues

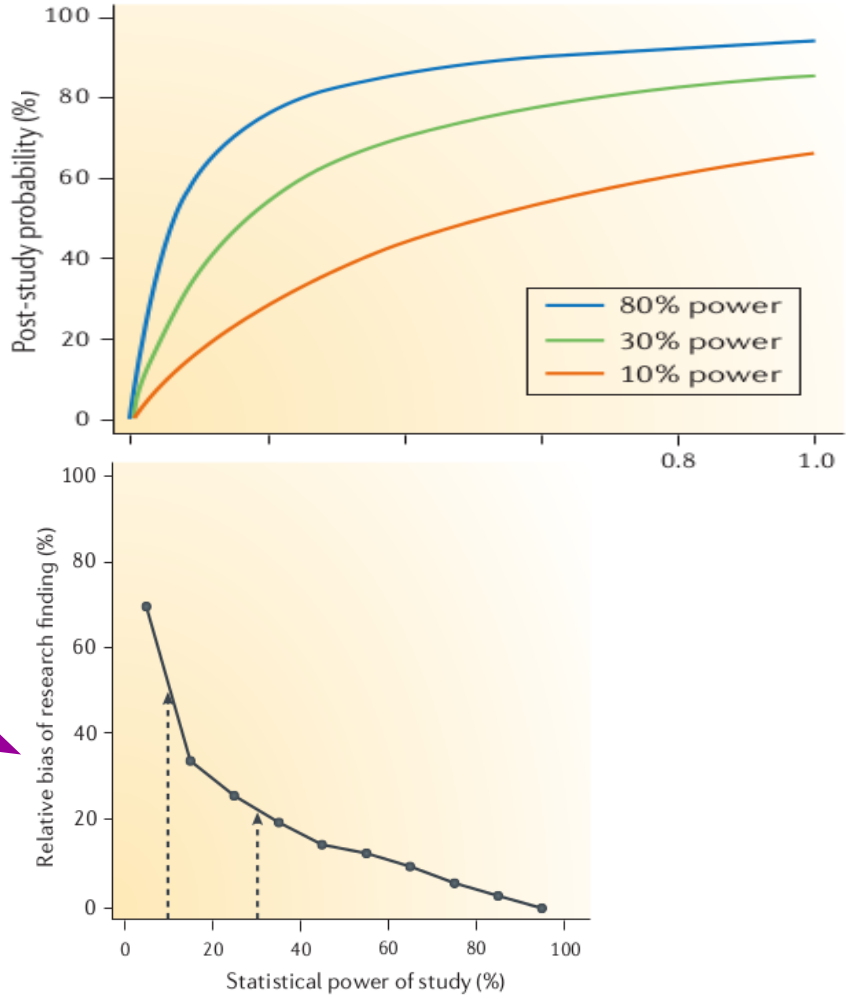
Open access, freely available online

Essay

Why Most Published Research Findings Are False

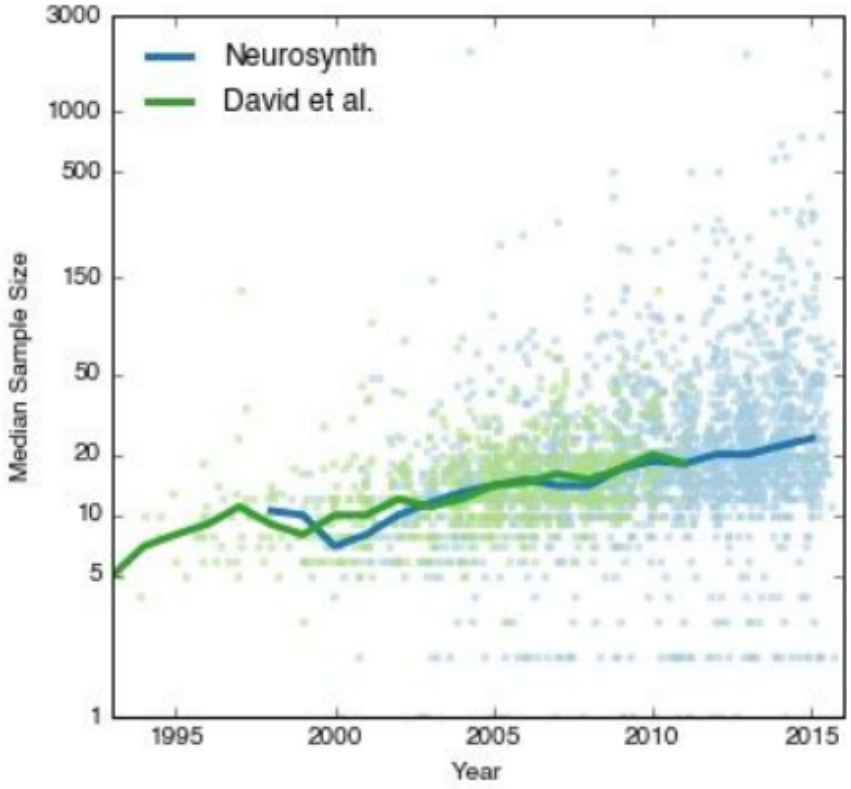
John P. A. Ioannidis

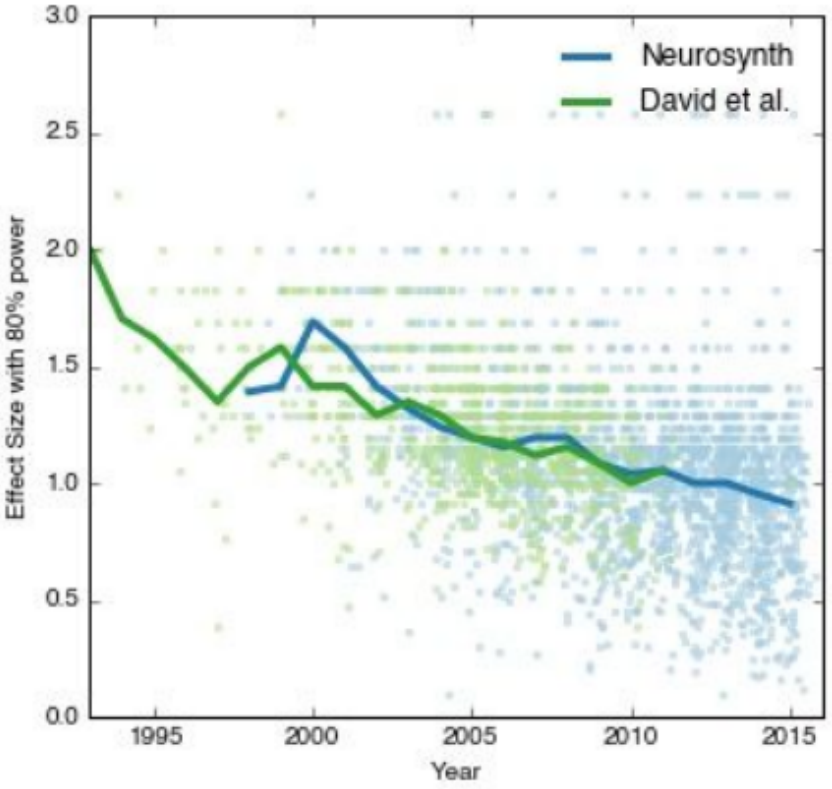




Button et al., NNR, 2013

Power - Effect size issues



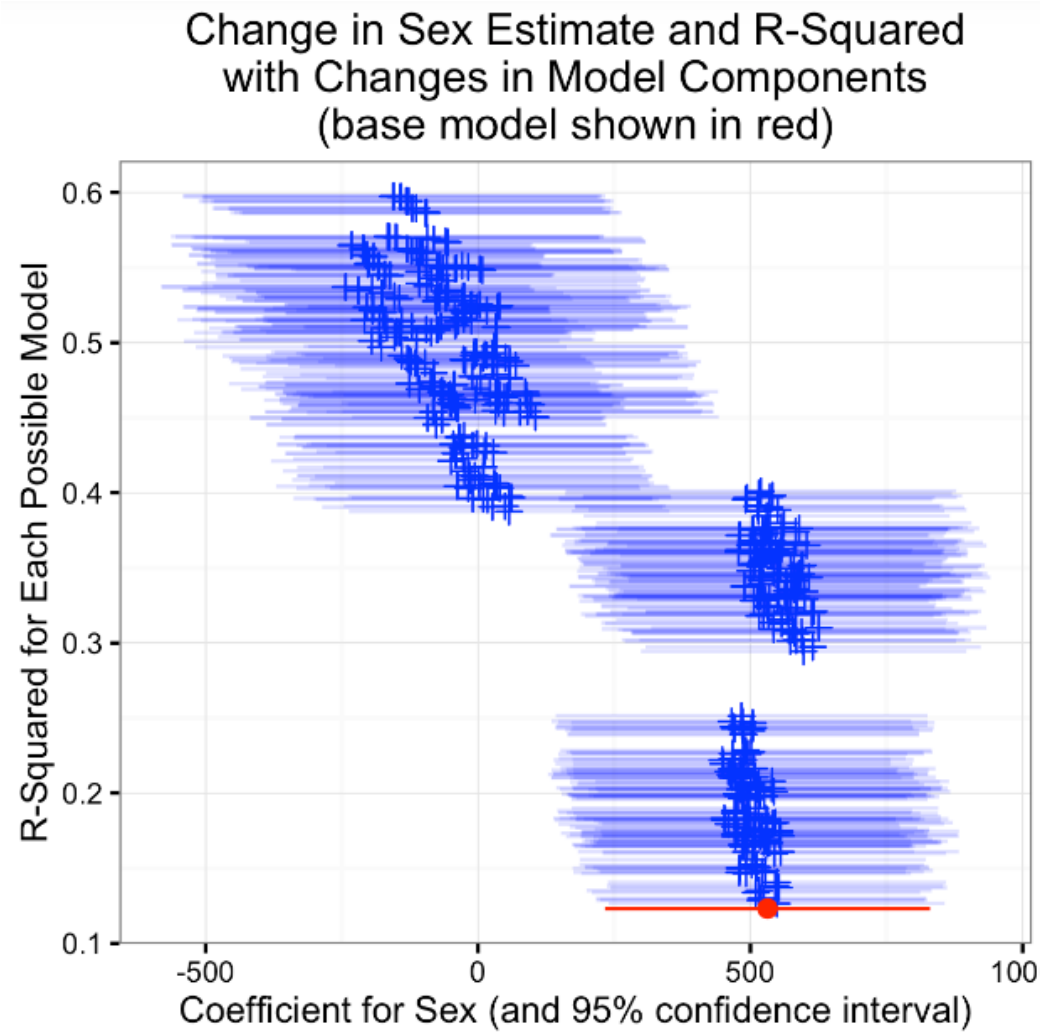


Poldrack et al., Nature Neuro. Rev., 2016

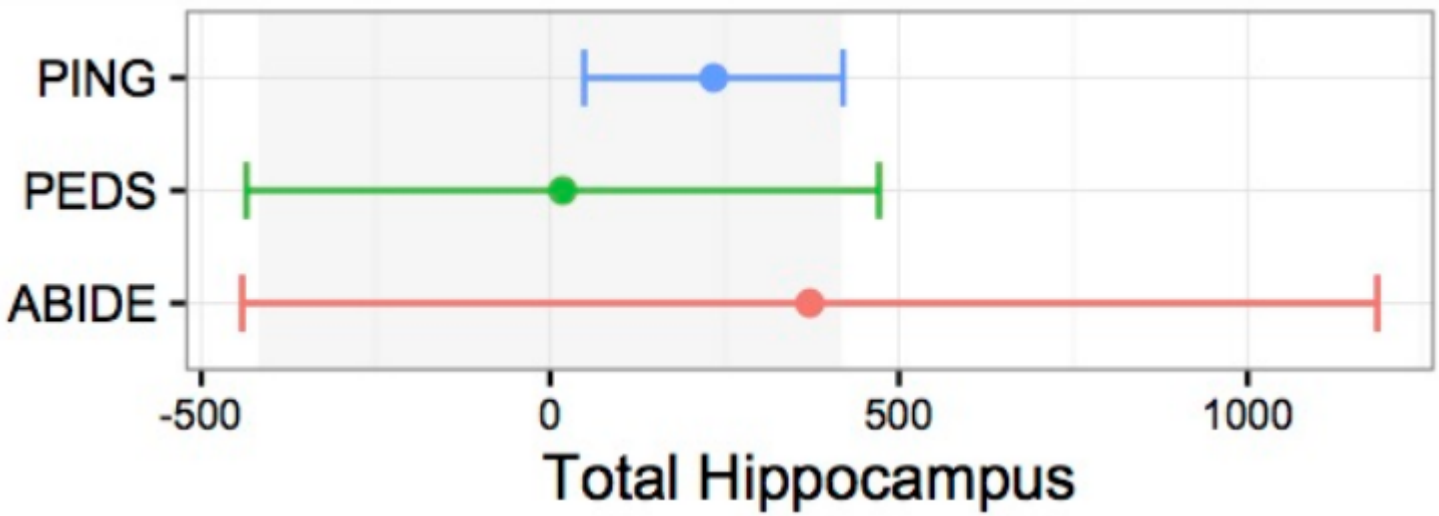
Effect and sample sizes

Paradigm	Intersection mask	mask size (vox)	Cohen D			BOLD		
			P10	median	P90	P10	median	P90
MOTOR	Bilateral Precentral Gyrus	12894	0.158	0.628	1.070	0.505	2.707	8.582
	Bilateral Supplementary motor cortex	3418	0.211	0.716	1.197	0.911	4.033	12.510
	Left putamen	1532	0.114	0.513	0.864	0.586	2.388	4.318
	Right putamen	1437	-0.008	0.369	0.749	-0.045	1.696	3.609
WM	Bilateral Middle frontal gyrus	7116	0.101	0.474	0.837	0.130	0.986	2.504
EMOTION	Left amygdala	1133	0.265	0.534	1.065	0.516	1.198	3.379
	Right amygdala	1082	0.308	0.645	1.140	0.581	1.350	3.557
GAMBLING	Left accumbens	455	0.138	0.310	0.461	0.369	0.849	1.440
	Right accumbens	417	0.141	0.332	0.488	0.373	0.981	1.618

Effect size: more issues (PING)



Classic power issues



Reproducibility issues: Causes

- Informatics: software and data
- Statistics
- **Social / cultural**
  - Lessons from issues
  - P. Smaldino: the evolution of bad science
  - R. Horton in The Lancet
  - Conclusion

Lessons learned: Social aspects

- Begley:
  - For results that could **not be reproduced**, data were **not** routinely analysed by investigators **blinded** to the experimental versus control groups
- Locher, Chang's colleague, reports:

  *"I think he was on **immense pressure to get the first structure**, and that's what made him push the limits of his data"*
- Baggerly:
  - Forensic possible because: **data, time, expertise were available**
  - No "life science" journal would take the paper

Lessons learned: Social aspects

- HeLa
  - ICLAC Initiative from scientists, **not by funding agencies, not by journals**
  - C. Korch: even after scientists are convinced their favorite cell line is contaminated, they may keep studying it.
- Donoho:

  "Publication is the advertisement, the scholarship is in the code"
- Eklund:
  - Train users to advance methods, Develop tests and checks for assumptions, prefer non parametric
- Button / Ioannidis / Poldrack:
  - We need more accessible and well documented data

The natural selection of bad science

- "... argues that some of the most powerful **incentives** in contemporary science actively encourage, reward, and **propagate poor research methods** and abuse of statistical procedures."

- ...between 1974 and 2014, the frequency of the words "innovative," "groundbreaking," and "novel" in PubMed abstracts increased by 2500% or more (Vinkers, Tijdink & Otte, 2015).

P. Smaldino and R. McElreath, Royal Society Open Science 2016

The natural selection of bad science

- The more any **quantitative social indicator** is used for social decision-making, the more subject it will be to corruption pressures and **the more apt it will be to distort and corrupt the social** processes it is intended to monitor.

  – Donald T. Campbell (1976, p. 49)

- "I've been on a number of search committees. I don't remember anybody looking at anybody's papers. **Number and IF [impact factor] of pubs are what counts**.

  – Terry McGlynn (realscientists) (21 October 2015, 4:12 p.m. Tweet.)