Do we have enough power?

Consequences of small sample size experiments and possible solutions

Silvia Sorce

Journal Club, 4th June 2013

Premise

- researchers must publish in order to succeed
- publishing is a highly competitive enterprise
- certain kinds of findings are more likely to be published than others
- researchers have strong incentives to engage in research practices that make their findings publishable quickly
- practices include using flexible study designs and flexible statistical analyses and running small studies with low statistical power…

… leading to statistically significant publishable results

Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

Statistically significant

1. Null hypothesis implies that there is **no relationship** between two measured phenomena
e.g. a potential medical treatment has no effect

2. The goal of each experiment is to **reject or disprove the null hypothesis**

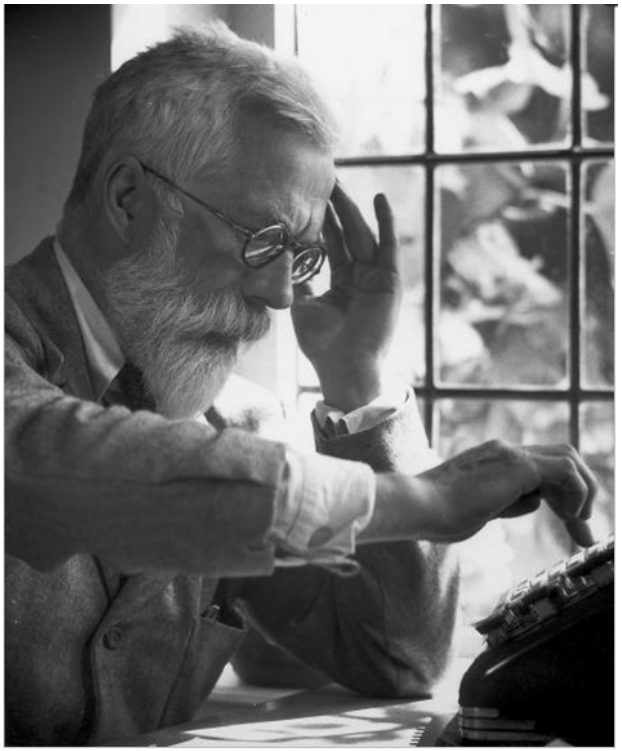
3. P values, or significance levels, measure the strength of the evidence against the null hypothesis; the smaller the P value, the stronger the evidence against the null hypothesis

«Through the 1960s, it was standard practice in many fields to report summaries with

one star attached to indicate ***p<0.05**
two stars to indicate **p<0.01.
three stars were used to indicate ***p<0.001. » http://www.jerrydallal.com/LHSP/p05.htm

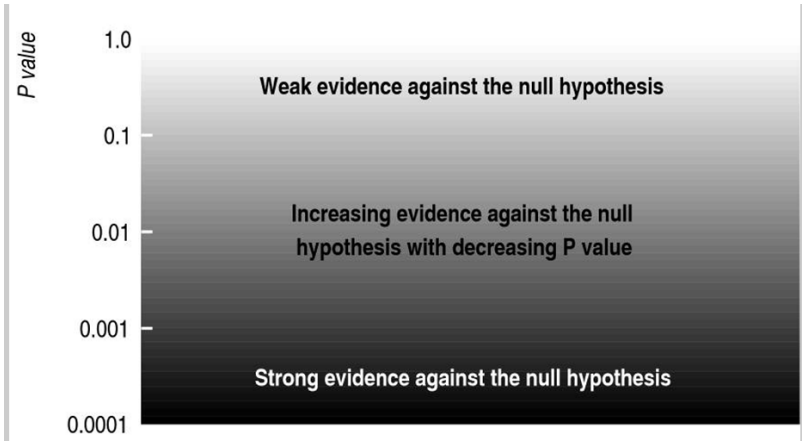
Why $p<0.05$?

concept introduced by Ronald Aylmer Fisher



Fisher published *Statistical Methods for Research Workers (SMRW) in 1925*

« If P is between 0.1 and 0.9 there is certainly no reason to suspect the hypothesis tested.
If it is below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts.
We shall not often be astray if we draw a conventional line at 0.05….it is convenient to take this point as a limit in judging whether a deviation ought to be considered significant or not»



Sterne and Smith, BMJ 2001;322:226–31

How much the result is significant?

However, Neyman and Pearson argued that it is not enough to say that a result is significant or not significant

There were 2 types of error that could be made in interpreting the results of an experiment

Type I error is the false rejection of the null hypothesis → false positive

Type II error is the false acceptance of the null hypothesis → false negative

	Diseased	**Healthy**
Positive to the test	**Diseased - positive**	Healthy - positive
Negative to the test	Diseased - negative	Healthy - negative

	True relationship	**No relationship**
Relationship	$1 - \beta$	α False positives
No relationship	β False negatives	$1 - \alpha$

power

Type I error

Type II error

Sterne and Smith, BMJ 2001;322:226–31

Statistical power → 1-β

The power of a statistical test is the probability that the test will reject the null hypothesis

The power is calculated based on :

- the α statistical threshold

- the magnitude of the effect → effect size

- the sample size used to detect the effect → n

How to calculate the power?

Behavior Research Methods
2007, 39 (2), 175-191

G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences

FRANZ FAUL
Christian-Albrechts-Universität Kiel, Kiel, Germany

EDGAR ERDFELDER
Universität Mannheim, Mannheim, Germany

AND

ALBERT-GEORG LANG AND AXEL BUCHNER
Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

- ***A priori power analysis***: sample size (n) is computed based on the required power, significance level and effect size

- ***Post Hoc power analysis***: power (1-β) is computed based on the significance level, effect size and sample size

Using G*Power 3 typically involves the following four steps:

(1) Select the statistical test appropriate for the problem,
(2) choose one of the five types of power,
(3) provide the input parameters required for the analysis,
(4) click on “Calculate” to obtain the results

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]

STANFORD
SCHOOL OF MEDICINE

John Ioannidis

Professor, Medicine - Stanford Prevention Research Center
Member, Stanford Cancer Institute
Professor, Health Research & Policy - Epidemiology
Professor (By courtesy), Natural Sciences Cluster - Statistics

View Larger

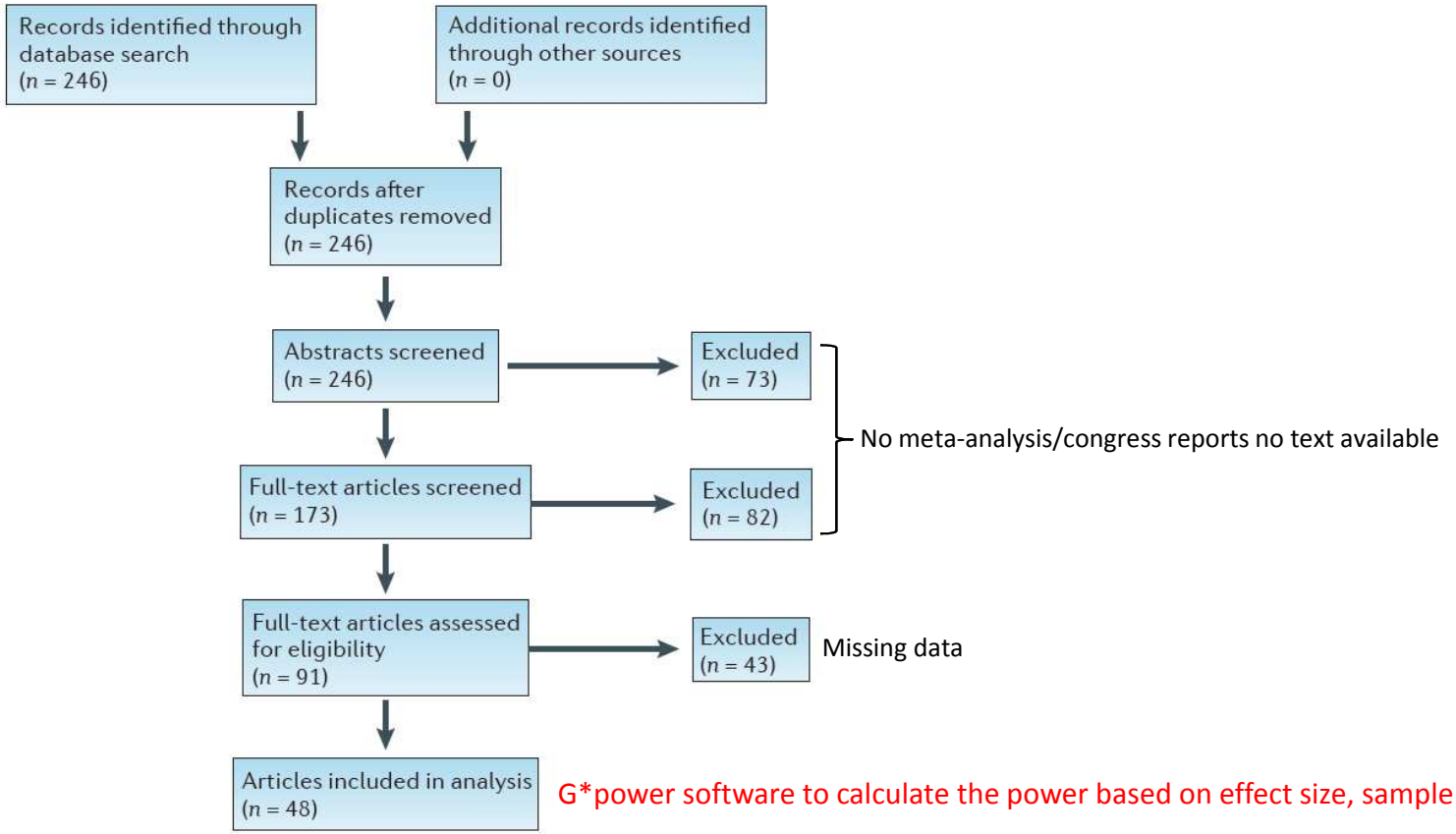
Aim: calculating the average statistical power in neuroscience

- Clinical studies
- Animal experiments

How to estimate statistical power in neuroscience → clinical studies

Methods

Search for articles published in 2011 : « neuroscience » and « meta-analysis »



No meta-analysis/congress reports no text available

Missing data

G*power software to calculate the power based on effect size, sample size, significance ($p<0.05$)

Figure 2 | **Flow diagram of articles selected for inclusion.** Computerized databases were searched on 2 February 2012 via Web of Science for papers published in 2011, using the key words 'neuroscience' and 'meta-analysis'.

Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

Median power of 49 analysed studies

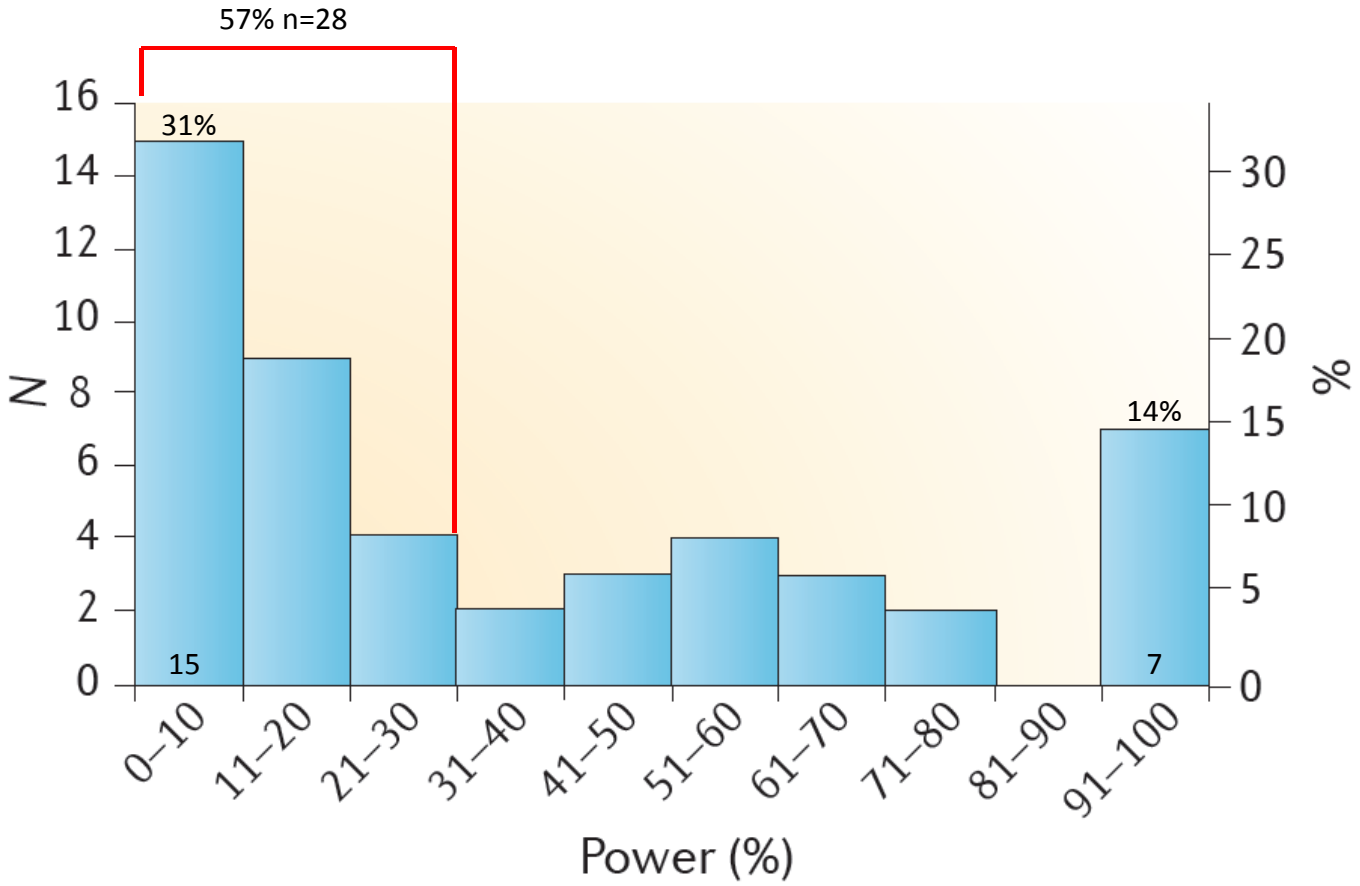


Figure 3 | **Median power of studies included in neuroscience meta-analyses.** The figure shows a

Median statistical power is 21%
Without 7 highest = 18%

Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

How to estimate statistical power in neuroscience → animal models

Methods: representative meta-analysis that combined data from studies investigating **sex differences in water maze performance and radial maze performance**

Test	N° studies	Effect size Cohen's *d*	Average sample size	Median statistical power
Water maze	19	0.49	22	18%
Radial maze	21	0.67	20	31%

From both clinical studies and animal experiments, it emerges that the statistical power in neuroscience is generally below 30%

Possibly true also for other disciplines

Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

What is the consequence of low power studies?

Dr Ioannidis had described a model to understand the impact of statistical power on the veridicity of the findings → PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

has been the most downloaded paper from *PLoS Medicine* *(the second on the list had "only" 101'096 views)*

If the power is low, there is:

1. low chance to find a genuinely true effect

i.e. if there are 100 findings to be discovered, with 20% power, only 20 will be discovered

2. overestimate of effect size

3. low chance that a statistically significant result reflects a true effect → PPV

PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

Positive predictive value (PPV)

PPV expresses the probability that an effect is true → depends from the power of the study

	Diseased	**Healthy**
Positive to the test	**Diseased - positive**	Healthy - positive
Negative to the test	Diseased - negative	Healthy - negative

	True relationship	**No relationship**
Relationship	$1 - \beta$	α False positives
No relationship	β False negatives	$1 - \alpha$
	p	1-p

$(1 - \beta)$ → power α → type I error

$$\textbf{PPV} = (1 - \beta) / (1 - \beta) + \alpha$$

considering the pre-study odds →

R → pre-study odds= p/1-p

$$\textbf{PPV} = [(1 - \beta) \times p] / [(1 - \beta) \times p + \alpha (1-p)]$$

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

PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

Positive predictive value (PPV) → examples

Example: 1 out 5 effects is true

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

R = p/1-p= (1/5)/(1-(1/5)) = 0.2/1-0.2= 0.2/0.8 = **0.25**

1. Study with power = 20%

$(1 - \beta) = 20\% = 0.2$ $\alpha = 0.05$ **PPV**= {0.2 x 0.25} / {0.2 x 0.25 + 0.05}= **0.5**

50% possibility that the result is true

2. Study with power = 80%

$(1 - \beta) = 80\% = 0.8$ $\alpha = 0.05$ **PPV**= {0.8 x 0.25} / {0.8 x 0.25 + 0.05}= **0.8**

80% possibility that the result is true

Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

PPV depends from the power

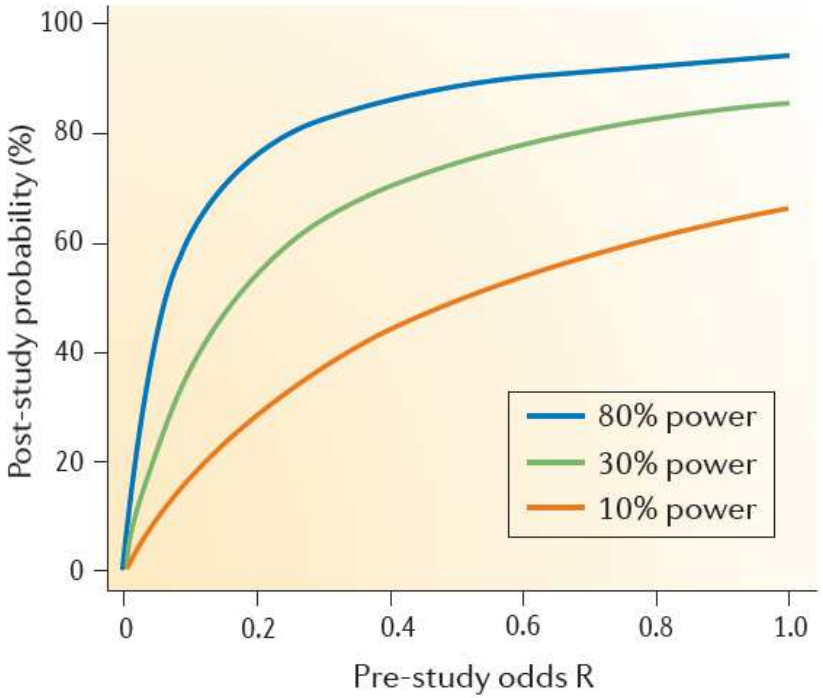


Figure 4 | **Positive predictive value as a function of the pre-study odds of association for different levels of statistical power.**

If the possibility to find an effect (R) is low
with low power, there is very low possibility that the discovered effect is true

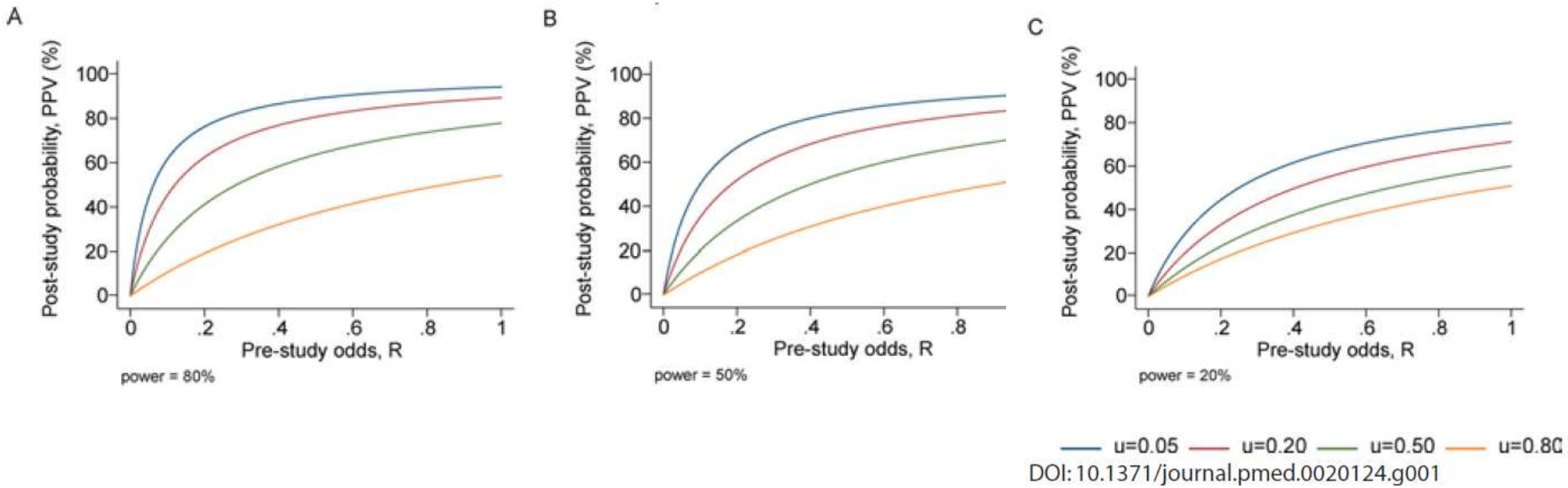
Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

PPV is lower in the presence of bias

Bias= combination of various design, data, analysis, and presentation factors that tend to produce research findings when they should not be produced

e.g. manipulation in the analysis or reporting of findings

A proportion *u* of the effects is reported as positive → this affects the PPV



DOI: 10.1371/journal.pmed.0020124.g001

PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

Ioannidis corollaries

Corollary 1: The smaller the studies conducted in a scientific field, the less likely the research findings are to be true. small sample size → small power

Corollary 2: The smaller the effect sizes in a scientific field, the less likely the research findings are to be true. pre-study odd is low →PPV will be low, with low power

Corollary 3: The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true.

results from large phase III randomized controlled trials *vs* hypothesis-generating experiments (*e.g.* high-throughput research)

Corollary 4: The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true.

true findings may be more common when outcomes are unequivocal and universally agreed (e.g. death) or when stereotyped analytical methods instead of new experimental analytical method are used

Corollary 5: The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true.

Conflicts of interest and prejudice may increase bias, u.

Corollary 6: The hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true.

With many teams involved, timing is of the essence in beating competition.
Team may prioritize on pursuing and disseminating its most impressive "positive" results.
Negative result become only attractive to refute a positive claim made in some prestigious journal → *Proteus effect ("recycled" results)*

PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

Which are the consequences?

Impact and examples

- Low reproducibility **- Discrepancies between animal and patient studies**

1

ORIGINAL CONTRIBUTION

Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

JAMA, July 13, 2005—Vol 294, No. 2

2

Comparison of treatment effects between animal experiments and clinical trials: systematic review

Pablo Perel, Ian Roberts, Emily Sena, Philipa Wheble, Catherine Briscoe, Peter Sandercock, Malcolm Macleod, Luciano E Mignini, Pradeep Jayaram, Khalid S Khan

BMJ, doi:10.1136/bmj.39048.407928.BE (published 15 December 2006)

3

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

NATURE REVIEWS | DRUG DISCOVERY VOLUME 10 | MAY 2011 | 1

1

ORIGINAL CONTRIBUTION

Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

JAMA, July 13, 2005—Vol 294, No. 2

Objective: evaluate the replication outcomes of highly cited clinical studies

Study selection:
- Publications with more than 1000 citations
- Published between 1990 and 2003 in the 3 medical journals with highest impact factor (*New England Journal of Medicine, JAMA, Lancet*) + medical specialty journals with impact factor > 7
- Addressed the efficacy of therapeutic or preventive interventions

…compared with

- Other concurrently or subsequently published clinical research addressing the same question
- Similar or larger sample size/better controlled design (*e.g.* randomized)

Control group:
- Less cited articles (median 157 citations)
- Matched 1:1 for journal, year of publication and design

Result summary

highly cited articles
49 were eligible:

- **7** (16%) were subsequently contradicted
- **7** (16%) showed initial stronger effect

- **20** (44%) were replicated
- **11** (24 %) remained unchallenged

- **4** contained «negative» results

less cited articles (control group):
49:

- **2** were subsequently contradicted
- **8** showed initial stronger effect

- **20** were replicated
- **8** remained unchallenged

- **11** contained «negative» results

Considerations:

→ Among highly cited articles: a large number of nonrandomized studies (5/6 vs 9/39) were not reproduced
→ Subsequent studies were either larger or better controlled

→ A trend for more contradicted studies in the highly cited group
→ Striking positive findings are quickly challenged
→ More negative results in the less cited studies

→ there is no proof that the subsequent studies were necessarily correct

→ Discrepancies can be interesting: careful scrutiny of the data and reappraisal of our beliefs

→ Uncertainity for clinical practice

Research

BMJ

Comparison of treatment effects between animal experiments and clinical trials: systematic review

Pablo Perel, Ian Roberts, Emily Sena, Philipa Wheble, Catherine Briscoe, Peter Sandercock, Malcolm Macleod, Luciano E Mignini, Pradeep Jayaram, Khalid S Khan

BMJ, doi:10.1136/bmj.39048.407928.BE (published 15 December 2006)

Objective To examine concordance between treatment effects in animal experiments and clinical trials.

Study selection
Animal studies for interventions with unambiguous evidence of a treatment effect (benefit or harm) in clinical trials:

1. corticosteroid for head injury
2. antifibrinolytics in haemorrhage
3. thrombolysis in acute ischaemic stroke
4. tirilazad in acute ischaemic stroke
5. antenatal corticosteroids to prevent neonatal respiratory distress syndrome
6. bisphosphonates to treat osteoporosis.

Result summary

intervention	patient study result	animal study results
corticosteroid for head injury	No benefit, increased mortality Alderson *et al., Cochrane Database Syst Rev* 2005	17 reports beneficial
antifibrinolytics in haemorrhage	Reduce blood loss during surgery Henry et al., *Cochrane Database Syst Rev* 1999	8 reports inconsistent results
thrombolysis in acute ischaemic stroke	Reduce death or dependency after stroke Mielke et al., *Cochrane Database Syst Rev* 2004	113 reports beneficial
tirilazad in acute ischaemic stroke	Increase death or dependency after stroke Trilazad commitee, *Cochrane Database Syst Rev* 2001	18 reports beneficial
Antenatal corticosteroids to prevent neonatal respiratory distress syndrome	Reduce respiratory distress and mortality Roberts et al., *Cochrane Database Syst Rev* 2006	56 reports beneficial
bisphosphonates to treat osteoporosis	Increase bone mineral density in post-menopausal women Cranney et al., *Endocr Rev* 2002	17 reports beneficial

Possible reasons for discrepancy:

- **Poor methodological quality of the studies, e.g. no randomization and blinding**
- **Low power + publication bias**
- **Failure of animal model to represent human disease e.g comorbidity effect: stroke, hypertension, diabetes**

BMJ, doi:10.1136/bmj.39048.407928.BE (published 15 December 2006)

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

Bayer HealthCare

→ the validity of published data on potential targets is crucial for companies when deciding to start novel projects

→ Pharmaceutical companies run in-house target validation programmes

→ Validation programmes at Bayer revealed that **exciting published results could not be reproduced**

→ Talking to scientists, both in academia and in industry, there seems to be a general impression that many results that are published are hard to reproduce

Objective: To substantiate incidental observations that published reports are frequently not reproducible with quantitative data

Nature Reviews Drug Discovery 10, 712 (September 2011) | doi:10.1038/nrd3439-c1

Methods

Methods:

an analysis of early in-house projects (target identification and validation) that were performed over 4 years

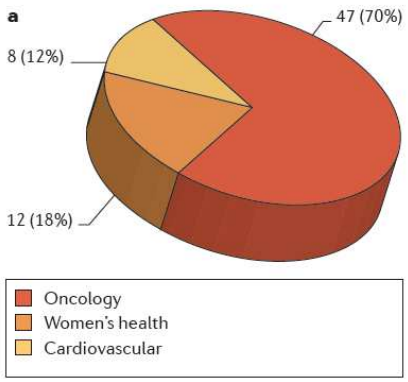
Questionnaire to scientists fro target discovery department:

- Names
- main relevant published data
- in-house data
- relationship with published data
- Impact of the results obtained for the outcome of the projects
- Models used in the experiments and in the publications

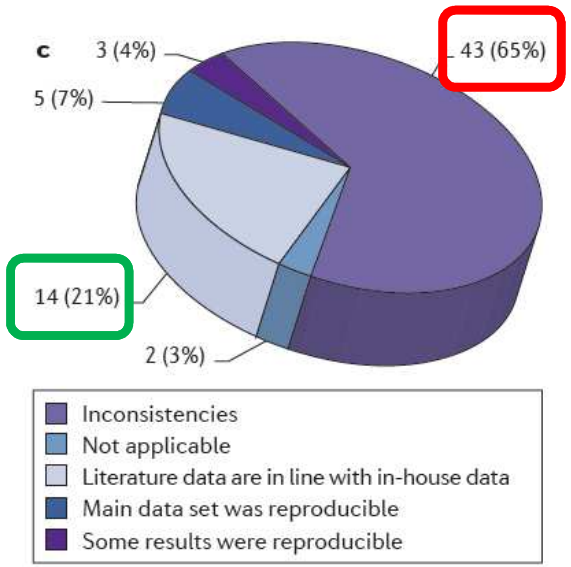
Nature Reviews Drug Discovery 10, 712 (September 2011) | doi:10.1038/nrd3439-c1

Results

results from 67 projects



not reproduced in the majority of cases



…is it due to different experimental conditions?

	Model reproduced 1:1	Model adapted to internal needs (cell line, assays)	Literature data transferred to another indication	Not applicable
In-house data in line with published results	1 (7%)	12 (86%)	0	1 (7%)
Inconsistencies that led to project termination	11 (26%)	26 (60%)	2 (5%)	4 (9%)

"Surprisingly, even publications in prestigious journals or from several independent groups did not ensure reproducibility"

Nature Reviews Drug Discovery 10, 712 (September 2011) | doi:10.1038/nrd3439-c1

SciBX
Science-Business eXchange

Hedging against academic risk

By Lev Osherovich, Senior Writer

ATLAS VENTURE

Bruce Booth

D.Phil.,
Oxford University

BS,
summa cum laude,
Penn State

HOMETOWN
West Chester, PA

COMPANIES
Bicycle
Miragen
Nimbus
ProtAffin
Zafgen

... according to Atlas Venture partner Bruce Booth:

… the "unspoken rule" among early stage VCs is that **at least 50% of published studies, even those in top-tier academic journals, can't be repeated with the same conclusions by an industrial lab**. "

As a result, Atlas now insists on **external validation studies of a new company's basic science as a precondition to further investment.**

*SciBX 4**(15)**; doi:10.1038/scibx.2011.416* ***Published online April 14 2011***

Can the situation be improved?

Increasing sample size?

Table 2 | Sample size required to detect sex differences in water maze and radial maze performance

	Total animals used	Required *N* per study		Typical *N* per study		Detectable effect for typical *N*	
		80% power	95% power	Mean	Median	80% power	95% power
Water maze	420	134	220	22	20	$d = 1.26$	$d = 1.62$
Radial maze	514	68	112	24	20	$d = 1.20$	$d = 1.54$

Meta-analysis indicated an effect size of Cohen's $d = 0.49$ for water maze studies and $d = 0.69$ for radial maze studies.

…will the veterinarians allow this?

«We argue that it is important to appreciate the waste associated with an underpowered study — even a study that achieves only 80% power still presents a 20% possibility that the animals have been sacrificed without the study detecting the underlying true effect»

Nat Rev Neurosci. 2013 May;14(5):365-76. doi: 10.1038/nrn3475. Epub 2013 Apr 10.

How much would it cost?

In U201= CHF- mouse/day 0.15 + cage/day (T 2L-IVC) 0.60

in A219= CHF-mouse/day 0.15 + cage/day (T 3) 1.20

Test	power	Males	Females	Total	n° cages (5 mice/cage)
Water maze	18%	22	22	44	9
Water maze	80%	134	134	268	54
Water maze	95%	220	220	440	88

18% power

U201	price	n°	n° days	Total CHF
mice	0.15	44	60	396
cages	0.6	9	60	324
				720

A219	price	n°	n° days	Total CHF
mice	0.15	44	60	396
cages	1.2	9	60	648
				1044

80% power

U201	price	n°	n° days	Total CHF
mice	0.15	268	60	2412
cages	0.6	54	60	1944
				4356

A219	price	n° mice	n° days	Total CHF
mouse/day	0.15	268	60	2412
cage/day	1.2	54	60	3888
				6300

95% power

U201	price	n°	n° days	Total CHF
mice	0.15	440	60	3960
cages	0.6	88	60	3168
				7128

A219	price	n°	n° days	Total CHF
mice	0.15	440	60	3960
cages	1.2	88	60	6336
				10296

!!!! Not considering genotyping, time/person cost for regular checks

…and time for running the actual experiment and analysing samples

…other suggestions?

Improve and standardize experimental design/statistics/reporting systems

Box 2 | **Recommendations for researchers**

Perform an a priori power calculation

Use the existing literature to estimate the size of effect you are looking for and design your study accordingly. If time or financial constraints mean your study is underpowered, make this clear and acknowledge this limitation (or limitations) in the interpretation of your results.

Disclose methods and findings transparently

If the intended analyses produce null findings and you move on to explore your data in other ways, say so. Null findings locked in file drawers bias the literature, whereas exploratory analyses are only useful and valid if you acknowledge the caveats and limitations.

Pre-register your study protocol and analysis plan

Pre-registration clarifies whether analyses are confirmatory or exploratory, encourages well-powered studies and reduces opportunities for non-transparent data mining and selective reporting. Various mechanisms for this exist (for example, the Open Science Framework).

Make study materials and data available

Making research materials available will improve the quality of studies aimed at replicating and extending research findings. Making raw data available will enhance opportunities for data aggregation and meta-analysis, and allow external checking of analyses and results.

Work collaboratively to increase power and replicate findings

Combining data increases the total sample size (and therefore power) while minimizing the labour and resource impact on any one contributor. Large-scale collaborative consortia in fields such as human genetic epidemiology have transformed the reliability of findings in these fields. → Journals should offer submission options for registered replications

ARRIVE guidelines= **A**nimals in **R**esearch: **R**eporting **I**n **V**ivo **E**xperiments

ARRIVE guidelines

Table 2. Animal Research: Reporting *In Vivo* experiments: The ARRIVE guidelines.

	ITEM	RECOMMENDATION
TITLE	1	Provide as accurate and concise a description of the content of the article as possible.
ABSTRACT	2	Provide an accurate summary of the background, research objectives (including details of the species or strain of animal used), key methods, principal findings, and conclusions of the study.
INTRODUCTION		
Background	3	a. Include sufficient scientific background (including relevant references to previous work) to understand the motivation and context for the study, and explain the experimental approach and rationale. b. Explain how and why the animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology.
Objectives	4	Clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested.

Kilkenny et al., PLoS Biol. **8**, e1000412 (2010).

METHODS		
Ethical statement	5	Indicate the nature of the ethical review permissions, relevant licences (e.g. Animal [Scientific Procedures] Act 1986), and national or institutional guidelines for the care and use of animals, that cover the research.
Study design	6	For each experiment, give brief details of the study design, including: a. The number of experimental and control groups. b. Any steps taken to minimise the effects of subjective bias when allocating animals to treatment (e.g., randomisation procedure) and when assessing results (e.g., if done, describe who was blinded and when). c. The experimental unit (e.g. a single animal, group, or cage of animals). A time-line diagram or flow chart can be useful to illustrate how complex study designs were carried out.
Experimental procedures	7	For each experiment and each experimental group, including controls, provide precise details of all procedures carried out. For example: a. How (e.g., drug formulation and dose, site and route of administration, anaesthesia and analgesia used [including monitoring], surgical procedure, method of euthanasia). Provide details of any specialist equipment used, including supplier(s). b. When (e.g., time of day). c. Where (e.g., home cage, laboratory, water maze). d. Why (e.g., rationale for choice of specific anaesthetic, route of administration, drug dose used).
Experimental animals	8	a. Provide details of the animals used, including species, strain, sex, developmental stage (e.g., mean or median age plus age range), and weight (e.g., mean or median weight plus weight range). b. Provide further relevant information such as the source of animals, international strain nomenclature, genetic modification status (e.g. knock-out or transgenic), genotype, health/immune status, drug- or test-naïve, previous procedures, etc.
Housing and husbandry	9	Provide details of: a. Housing (e.g., type of facility, e.g., specific pathogen free (SPF); type of cage or housing; bedding material; number of cage companions; tank shape and material etc. for fish). b. Husbandry conditions (e.g., breeding programme, light/dark cycle, temperature, quality of water etc. for fish, type of food, access to food and water, environmental enrichment). c. Welfare-related assessments and interventions that were carried out before, during, or after the experiment.
Sample size	10	a. Specify the total number of animals used in each experiment and the number of animals in each experimental group. b. Explain how the number of animals was decided. Provide details of any sample size calculation used. c. Indicate the number of independent replications of each experiment, if relevant.
Allocating animals to experimental groups	11	a. Give full details of how animals were allocated to experimental groups, including randomisation or matching if done. b. Describe the order in which the animals in the different experimental groups were treated and assessed.
Experimental outcomes	12	Clearly define the primary and secondary experimental outcomes assessed (e.g., cell death, molecular markers, behavioural changes).
Statistical methods	13	a. Provide details of the statistical methods used for each analysis. b. Specify the unit of analysis for each dataset (e.g. single animal, group of animals, single neuron). c. Describe any methods used to assess whether the data met the assumptions of the statistical approach.

RESULTS		
Baseline data	14	For each experimental group, report relevant characteristics and health status of animals (e.g., weight, microbiological status, and drug- or test-naïve) before treatment or testing (this information can often be tabulated).
Numbers analysed	15	a. Report the number of animals in each group included in each analysis. Report absolute numbers (e.g. 10/20, not 50%[a]). b. If any animals or data were not included in the analysis, explain why.
Outcomes and estimation	16	Report the results for each analysis carried out, with a measure of precision (e.g., standard error or confidence interval).
Adverse events	17	a. Give details of all important adverse events in each experimental group. b. Describe any modifications to the experimental protocols made to reduce adverse events.
DISCUSSION		
Interpretation/scientific implications	18	a. Interpret the results, taking into account the study objectives and hypotheses, current theory, and other relevant studies in the literature. b. Comment on the study limitations including any potential sources of bias, any limitations of the animal model, and the imprecision associated with the results[a]. c. Describe any implications of your experimental methods or findings for the replacement, refinement, or reduction (the 3Rs) of the use of animals in research.
Generalisability/translation	19	Comment on whether, and how, the findings of this study are likely to translate to other species or systems, including any relevance to human biology.
Funding	20	List all funding sources (including grant number) and the role of the funder(s) in the study.

[a]Schulz, et al. (2010) [24].

doi:10.1371/journal.pbio.1000412.t002

EDITORIAL

nature neuroscience

Raising standards

Nature journals' updated editorial policies aim to improve transparency and reproducibility.

"…there is no justification for not reporting, with full transparency, how a study is designed, conducted and analyzed so that reviewers and readers can adequately interpret and build on the results. For studies using biological samples, we will require authors to state whether statistical methods were used (or not) to predetermine sample size, and what criteria they used to identify and deal with outliers while running the experiment.

Reporting Checklist for Nature Neuroscience

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results.

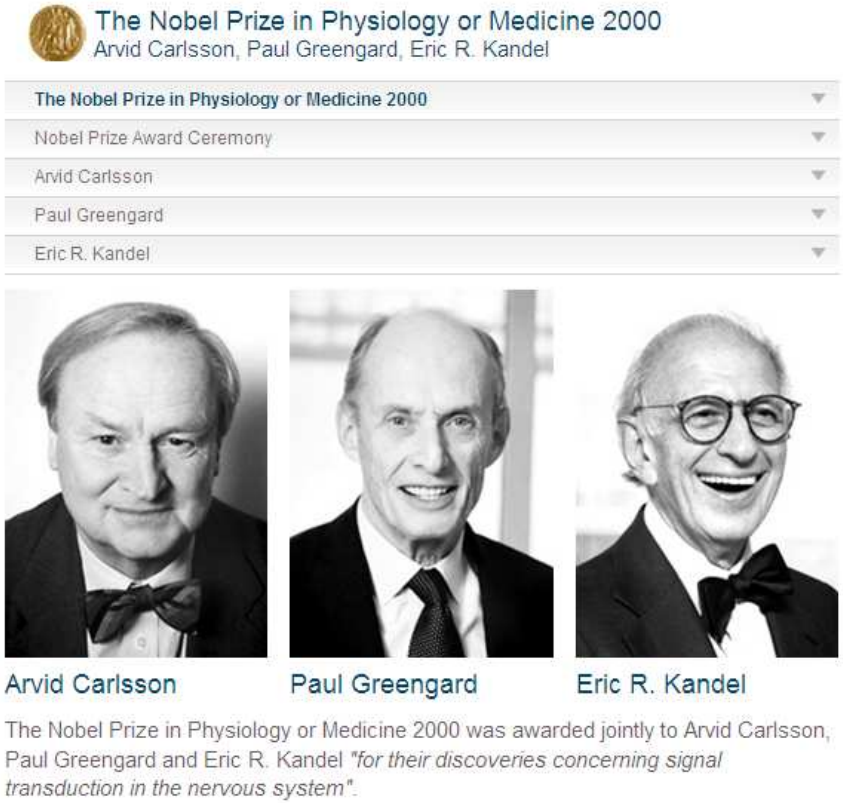
Please note that in the event of publication, it is mandatory that authors include all relevant methodological and statistical information in the manuscript.

How many mice did they use?????

Nobel Prizes

Since the beginning of the 20th Century, these prizes have charted the world's greatest medical advances. Of the 98 Nobel Prizes awarded for Physiology or Medicine, 75 were directly dependant on research from animals. In a further four instances no animal experiments were performed by the prize winner, but the discovery relied on crucial data obtained from animal studies by other research groups.

http://www.animalresearch.info/en/medical-advances/nobel-prizes/



Adenosine 3′,5′-Monophosphate in Nervous Tissue: Increase Associated with Synaptic Transmission

Abstract. *Brief periods of stimulation of the preganglionic nerve fibers produced a severalfold increase in the content of adenosine 3′,5′-monophosphate in superior cervical sympathetic ganglia, whereas postganglionic stimulation did not. These and other experiments indicated that the increased concentrations of adenosine 3′5′-monophosphate were closely associated with the process of synaptic transmission. This increase occurred primarily in postsynaptic cells.*

McAfee DA, Schorderet M, Greengard P. *Science*. 1971 Mar 19; 171(3976): 1156-8

the first demonstration that the cyclic AMP content of nervous tissue can change in response to synaptic activity

Table 1. Effect of preganglionic stimulation on the content of cyclic AMP in rabbit superior cervical sympathetic ganglia. One ganglion from each rabbit was stimulated at a frequency of 10 per second. The contralateral ganglion served as an unstimulated control. The data have been calculated as the mean ± standard error for the number (N) of rabbits indicated in the second column. In the last column, the concentration of cyclic AMP in the stimulated ganglia is expressed as the percentage of that in the unstimulated control ganglia. Temperature, 33°C.

Duration of stimulation (min)	N	Cyclic AMP (picomoles per milligram of protein)			Percentage of control
		Unstimulated ganglion	Stimulated ganglion	Absolute increase	
0.5	5	23.2 ± 5.5	43.4 ± 11.2	20.3 ± 13.0	249 ± 113
1.0	5	13.1 ± 1.3	61.8 ± 6.1	48.7 ± 5.2	479 ± 39
2.0	11	18.0 ± 1.7	70.2 ± 9.1	52.1 ± 8.2	399 ± 40
4.0	4	14.2 ± 2.9	57.0 ± 3.0	43.5 ± 4.3	483 ± 136
8.0	4	16.4 ± 1.0	72.0 ± 4.5	55.7 ± 4.6	445 ± 41

Impaired Long-Term Potentiation, Spatial Learning, and Hippocampal Development in *fyn* Mutant Mice

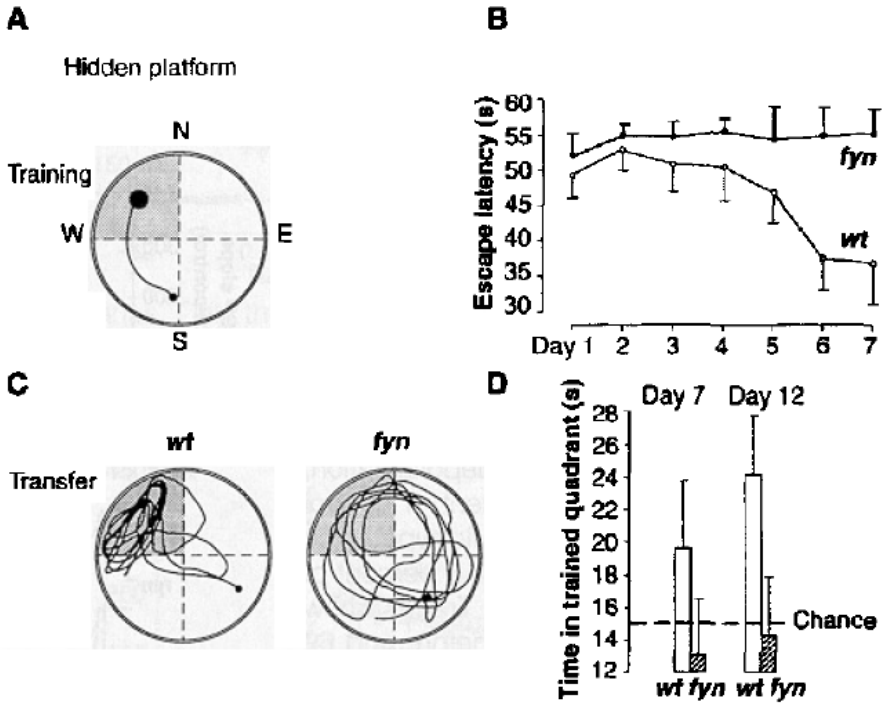
Seth G. N. Grant, Thomas J. O'Dell, Kevin A. Karl, Paul L. Stein, Philippe Soriano, Eric R. Kandel

Mice with mutations in four nonreceptor tyrosine kinase genes, *fyn*, *src*, *yes*, and *abl*, were used to study the role of these kinases in long-term potentiation (LTP) and in the relation of LTP to spatial learning and memory. All four kinases were expressed in the hippocampus. Mutations in *src*, *yes*, and *abl* did not interfere with either the induction or the maintenance of LTP. However, in *fyn* mutants, LTP was blunted even though synaptic transmission and two short-term forms of synaptic plasticity, paired-pulse facilitation and post-tetanic potentiation, were normal. In parallel with the blunting of LTP, *fyn* mutants showed impaired spatial learning, consistent with a functional link between LTP and learning. Although *fyn* is expressed at mature synapses, its lack of expression during development resulted in an increased number of granule cells in the dentate gyrus and of pyramidal cells in the CA3 region. Thus, a common tyrosine kinase pathway may regulate the growth of neurons in the developing hippocampus and the strength of synaptic plasticity in the mature hippocampus.

SCIENCE • VOL. 258 • 18 DECEMBER 1992

→ water maze experiments

n= 7 for WT and n=8 for *fyn* KO



Conclusions

- **Dramatic advances in flexibility of research design and analysis, but…**
- **… stability of sample size and research of smaller/more subtle effects → low power and low PPV**

- **Increasing the power can be «practically» difficult, but…**
- **existing scientific practices can be improved by raising the standards for study designs and reporting systems**

Thank you for your attention!

Questions/comments?