

Ethics in science: Experimental design and data presentation

Frédéric Schütz
Frederic.Schutz@unil.ch

Biostatistics platform, FBM



**There are infinitely many ways to
introduce bias into an experiment**

To avoid biases: Some rules of experimental design

Replication

Sample-size determination

Randomisation

Blocking

Blind, double-blind

Comparison, placebo

Factorial design

COMMENT



Six red flags for suspect work

C. Glenn Begley explains how to recognize the preclinical papers in which the data won't stand up.

A few months ago, I received a desperate e-mail from a postdoctoral scientist. Researchers — including me and my colleagues — had just reported that the majority of preclinical cancer papers in top-tier journals could not be reproduced, even by the investigators themselves^{1,2}. The postdoc pleaded with me to identify those papers, saying, "I could be

wasting my time working on that project." This was true, but we had signed confidentiality agreements that prevented us from revealing the specific papers. Furthermore, identifying them would not address the broader, systemic issues in research and publishing that create a plethora of papers that don't stand up to scrutiny.

There were some glaring differences ▶

**"Were experiments
performed blindly ?"**

**A class of students were asked
to count the number
of brown nuclei
in two images.**

Image 1

This image was labelled as coming from a **healthy patient**.
We expect to observe a **low** number of brown nuclei.

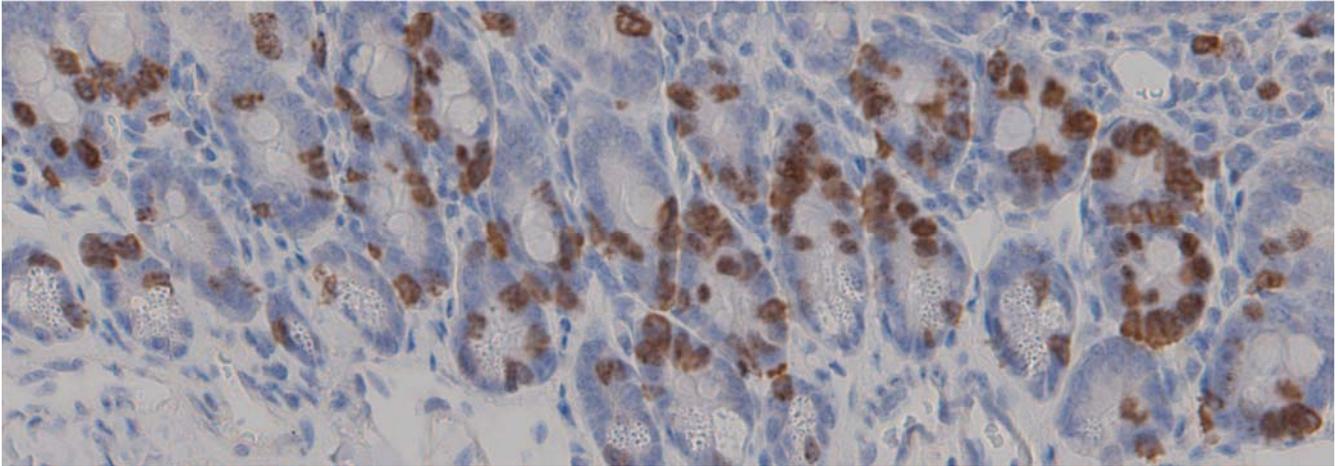
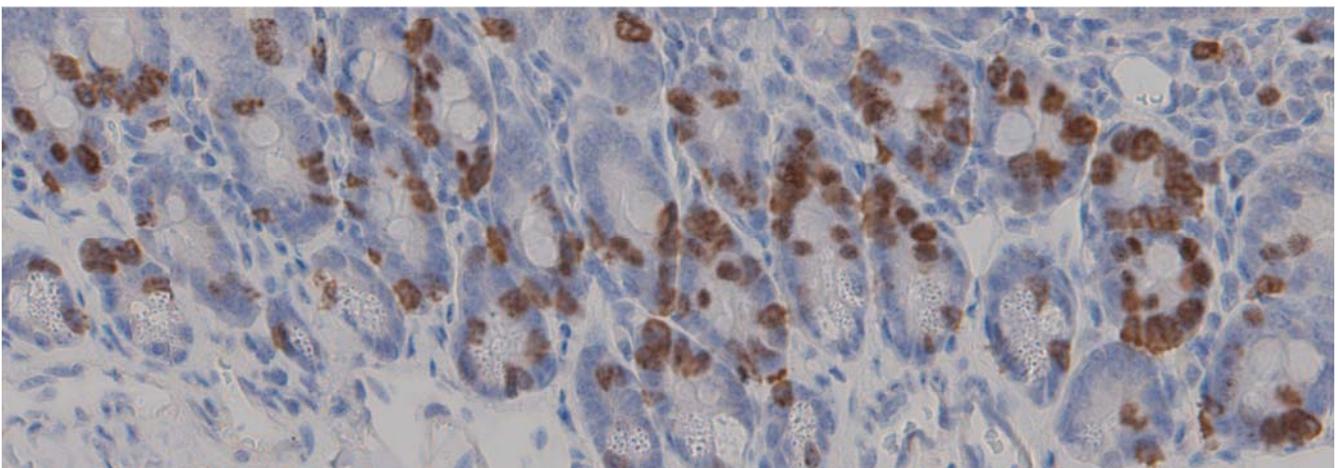


Image 2

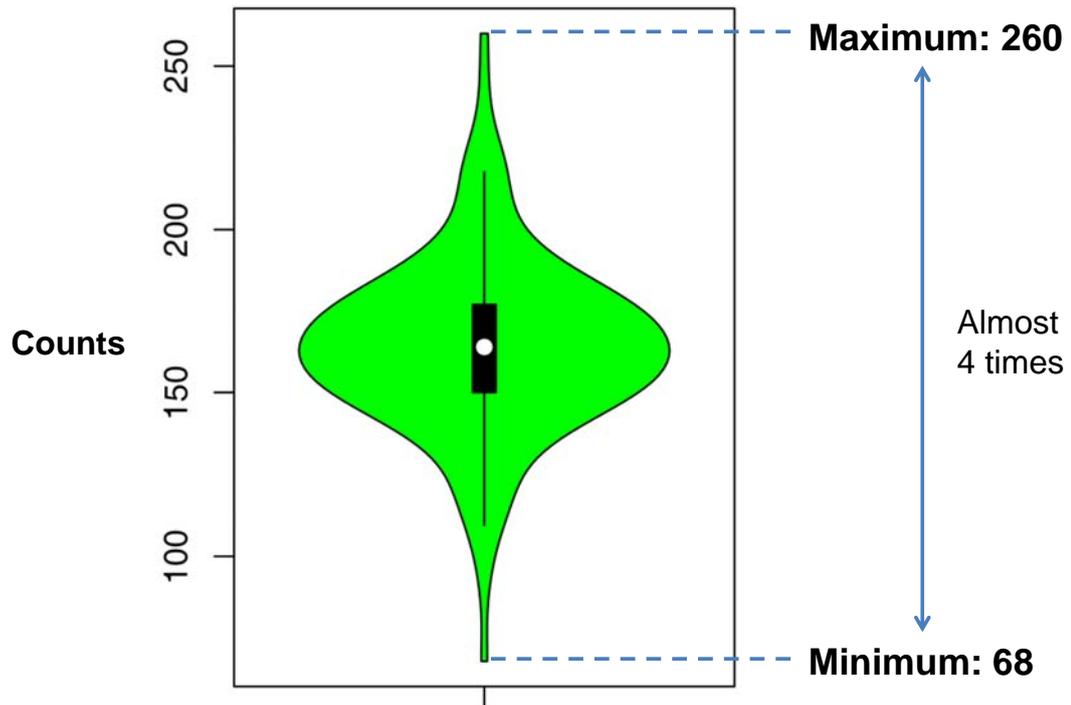
This image was labelled as coming from a **cancer patient**.
We expect to observe a **large** number of brown nuclei.



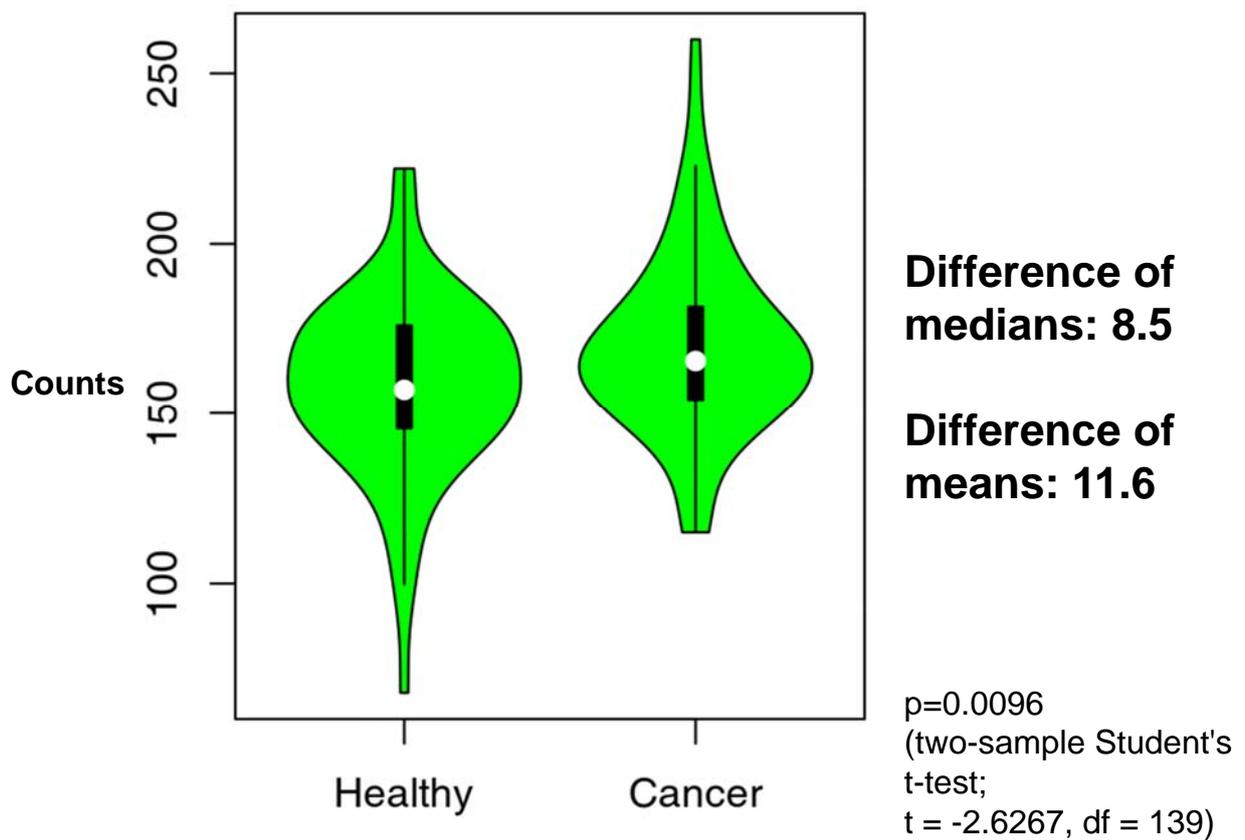
Goal:
**Assess whether the students were
(unconsciously) influenced by the
stated expectations**

Results

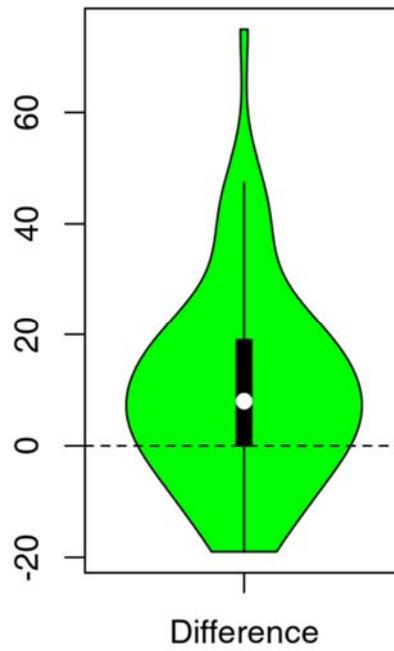
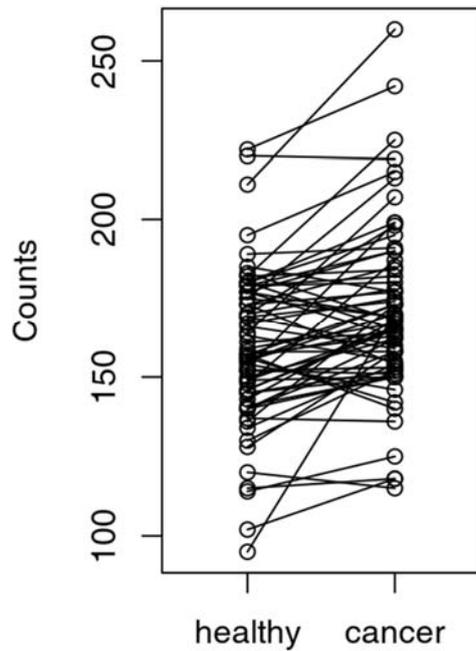
Distribution of all counts



Healthy vs cancer



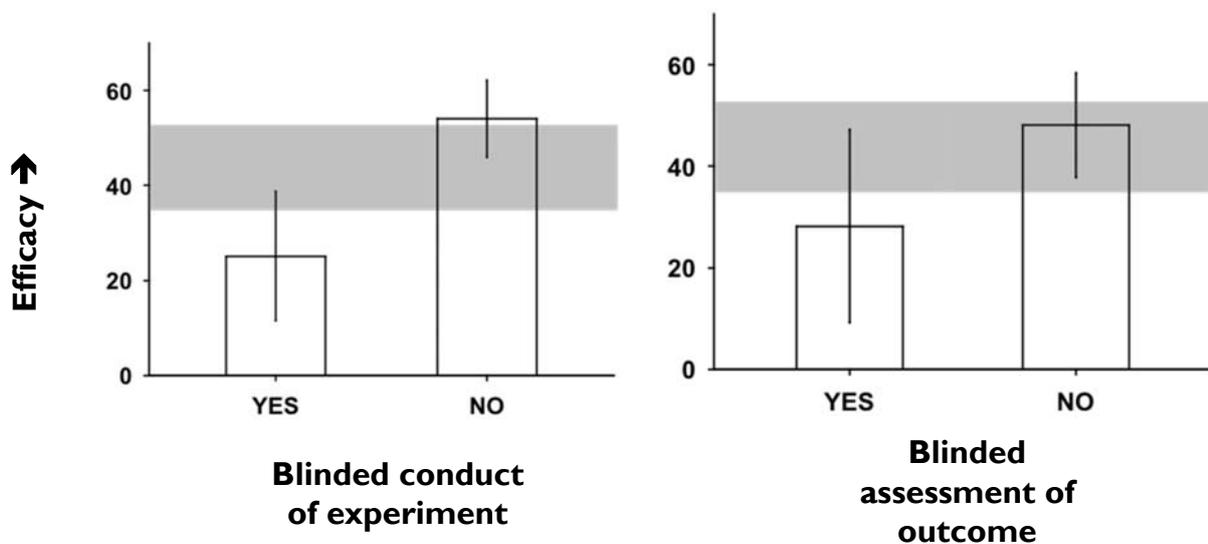
Paired data



75% of counts were higher for "cancer" compared to "healthy"

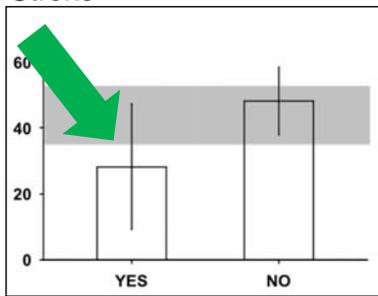
Mean difference: **10.5 units**; p-value: **5×10^{-6}**
(paired Student's t-test; $t = 4.9585$, $df = 68$)

Effect of bias in an animal experiment: the example of stroke



Evidence from other domains

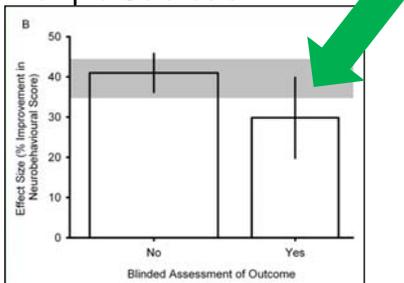
Stroke



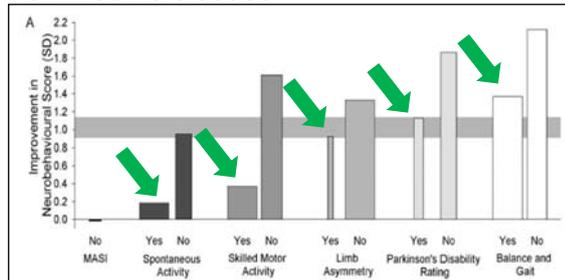
Alzheimer's disease



Multiple Sclerosis



Parkinson's disease



From Malcolm Macleod

Scientific process, in theory...

" A scientist is someone who loves to be proven wrong "

" Don't fall in love with your hypothesis "

(Adriano Aguzzi)

In practice, however...

***We all suffer from
confirmation bias***

**"Were there
positive and negative
controls ?"**

**Controls
are at the heart
of experimental design**

**Example
Can you find the rule ?**

I have chosen a rule that some sequences of three numbers obey and some do not.

Your job is to find out what the rule is using example sequences.

I start by giving you a sequence which obeys the rule:

2

4

8

Obeys the rule

Now it's your turn. Write a number sequence, and I'll tell you whether it satisfies the rule or not. You can test (more or less) as many sequences as you want.

?

?

?

?

When you think you know the rule, describe it in words:

A typical session:

2	4	8	Obeys the rule
1	2	4	Obeys the rule
3	6	12	Obeys the rule

...

Each number is twice the previous one.

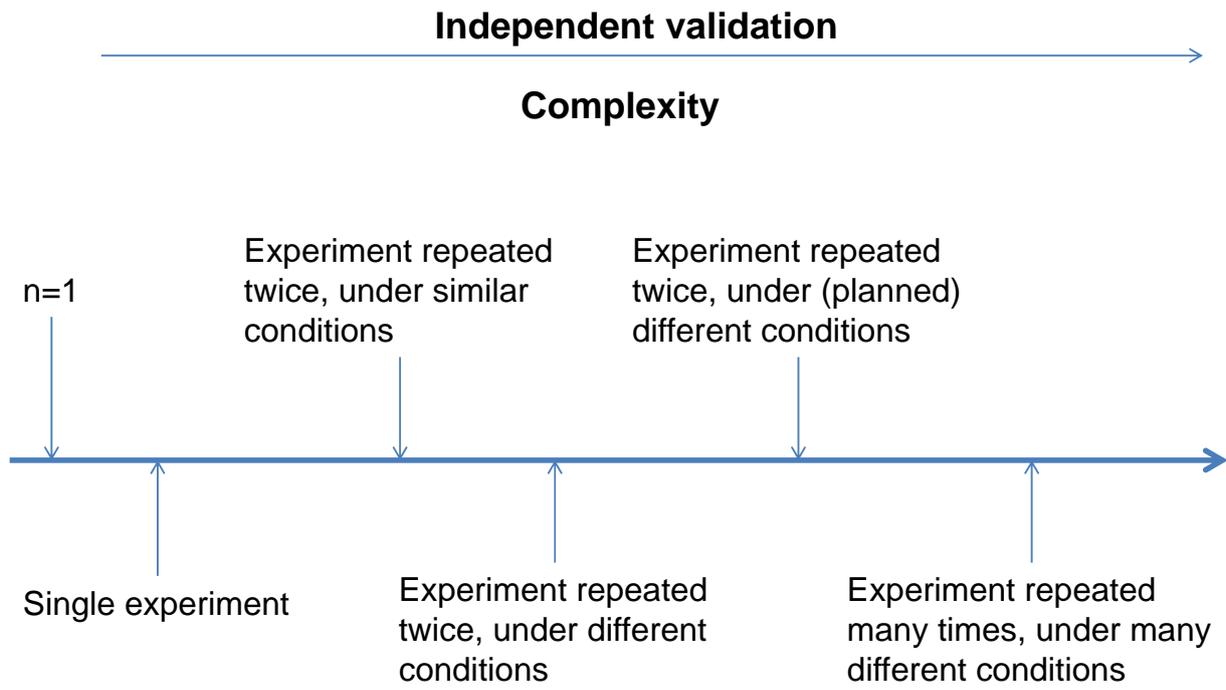
The original rule for the three numbers was that each of them must be larger than the one before.

This game was proposed by the New-York Times (<http://www.nytimes.com/interactive/2015/07/03/upshot/a-quick-puzzle-to-test-your-problem-solving.html>) and is an example of "confirmation bias".

On the NYT website, almost 80% of respondents answered with a rule without trying a "negative control" – a sequence that should not pass their proposed rule. This is an example of confirmation bias, where a person just tries to confirm his/her hypothesis without considering alternative explanations – something that should be avoided, especially in science.

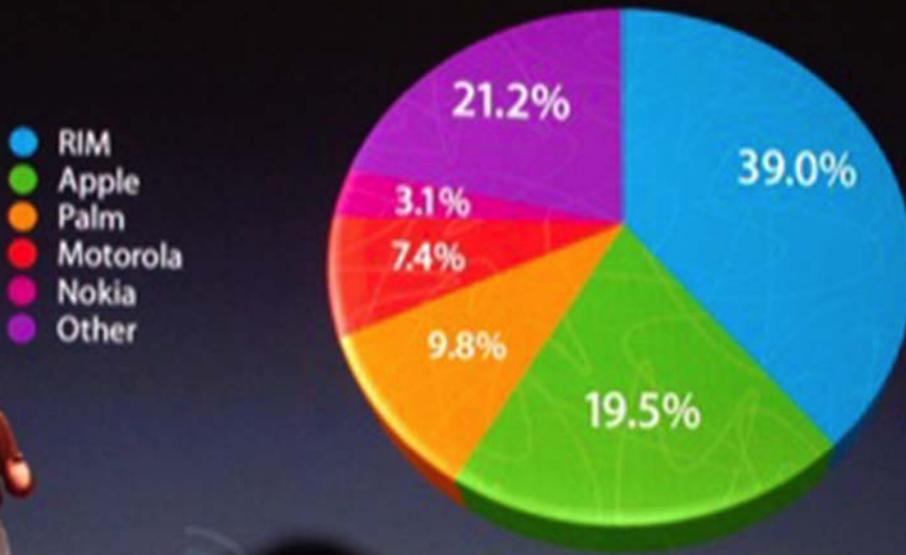
"Were all the results presented ?"
"Were statistical tests appropriate ?"

"Were basic experiments repeated ?"



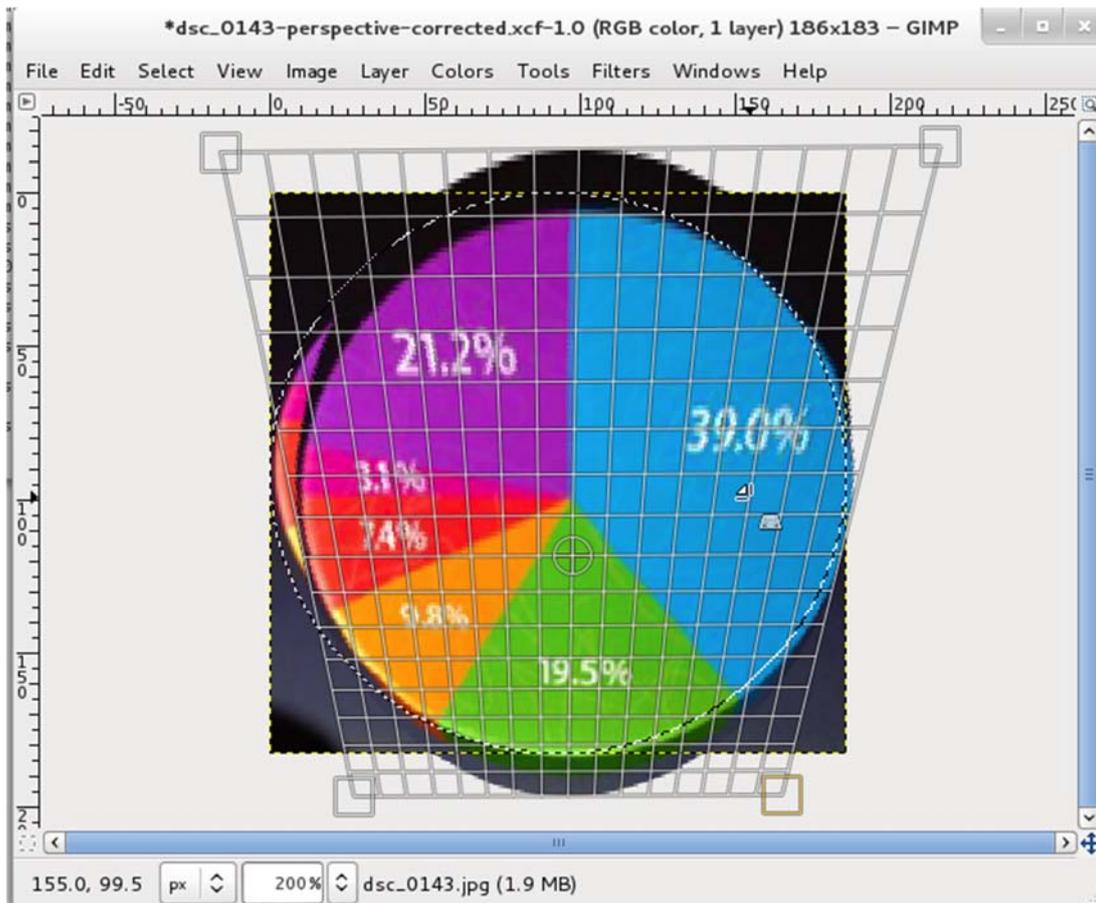
Presenting data

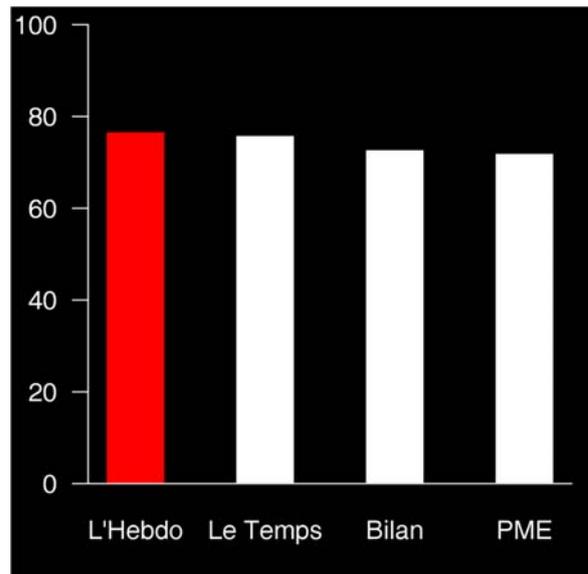
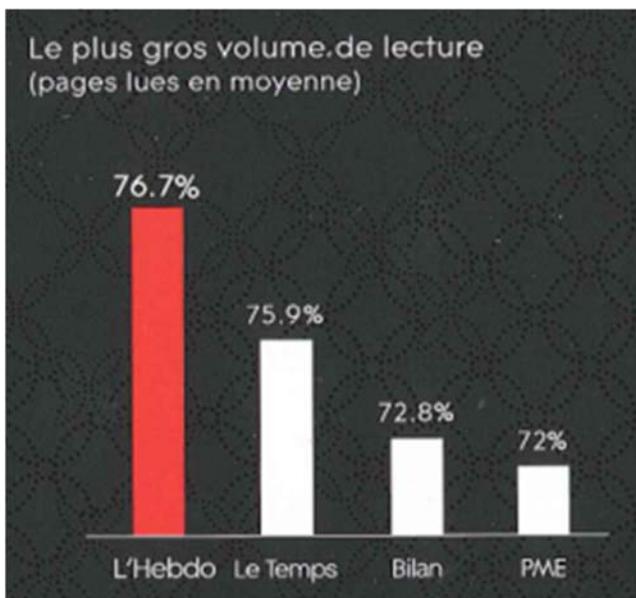
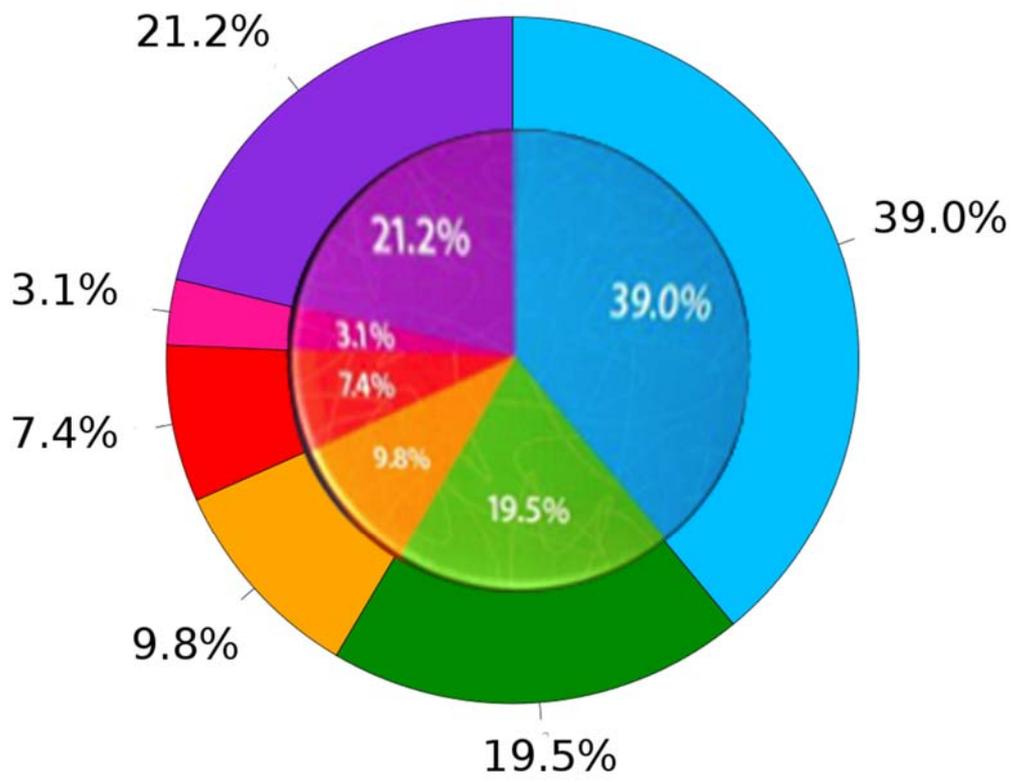
U.S. SmartPhone Marketshare



Steve Jobs, MacWorld 2008

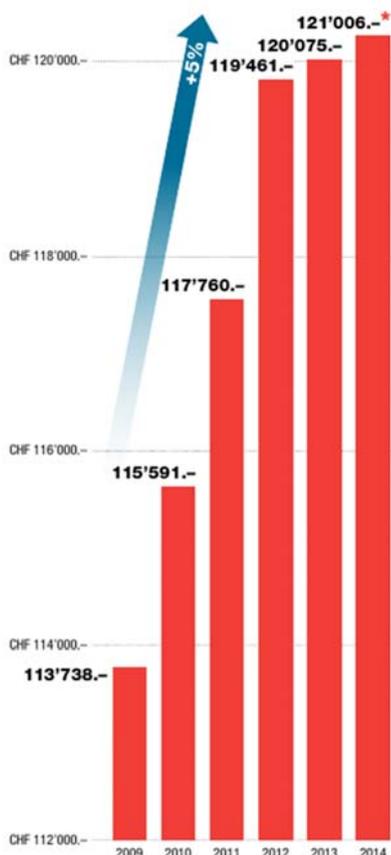
Photo: Engadget.com



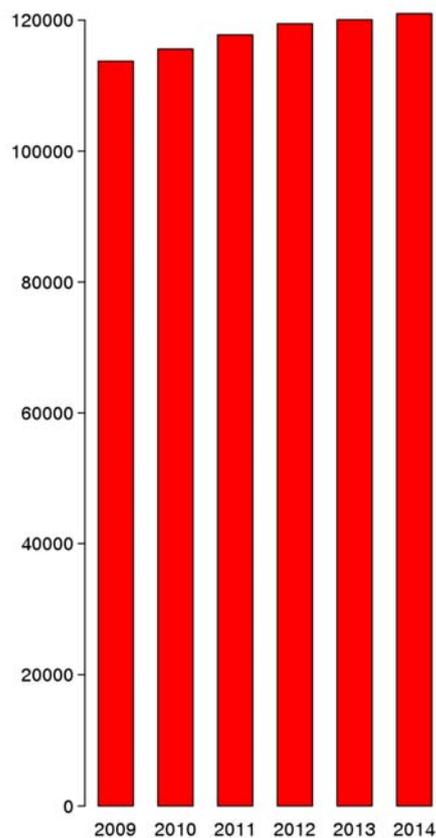


Augmentation incroyable des salaires à l'État (2009 - 2014)

Salaire brut moyen du personnel de l'État

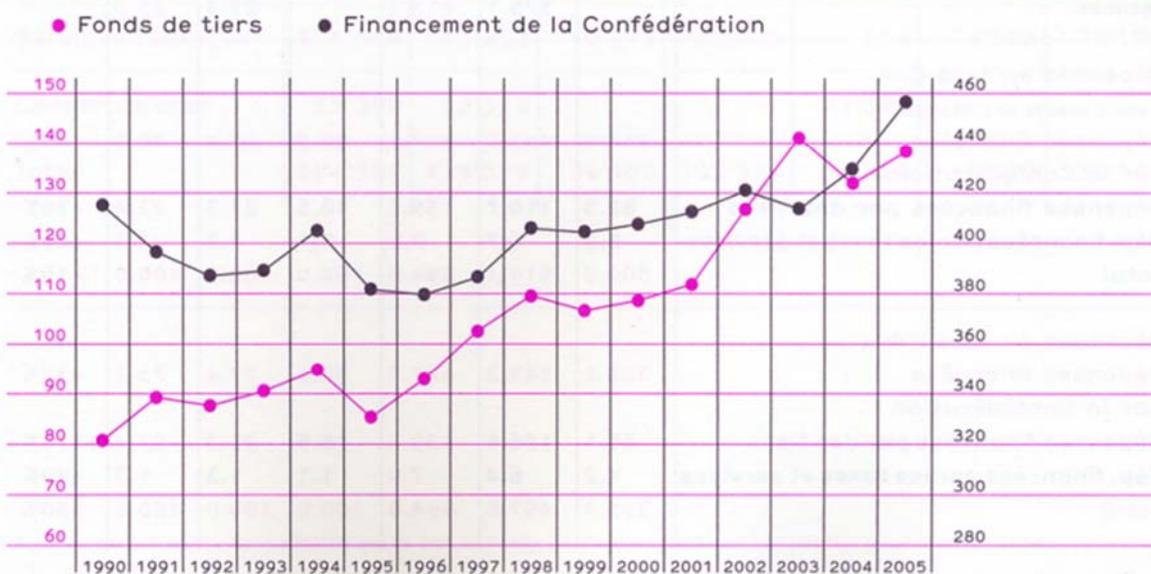


Sources: OFS (2015), sans primes de performance/allocations. *Extrapolation 1^{er} - 3^{ème} trimestre

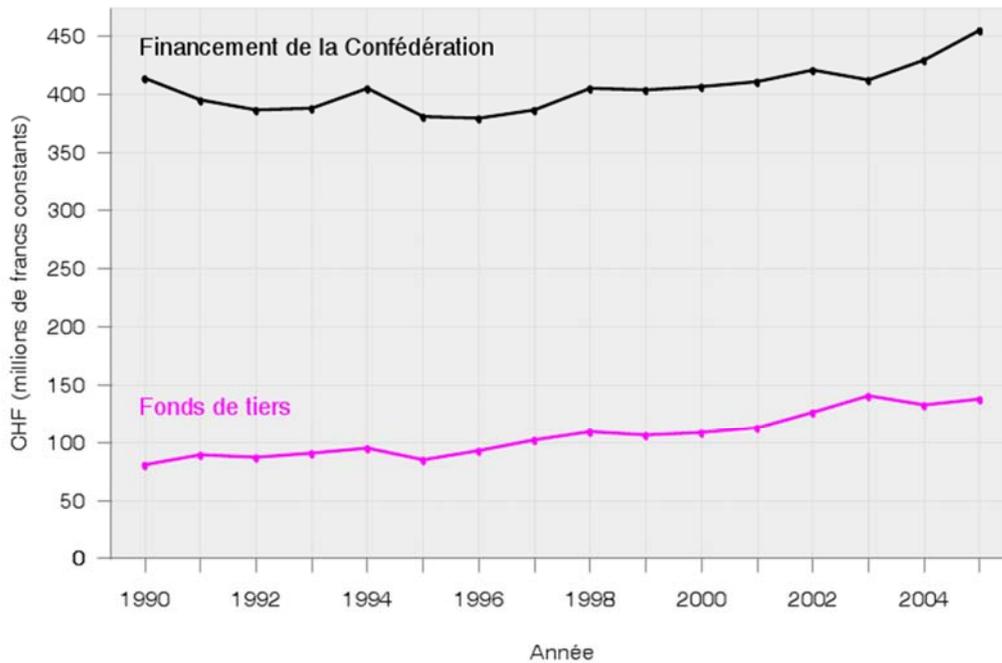


Resources

(mios CHF constants)

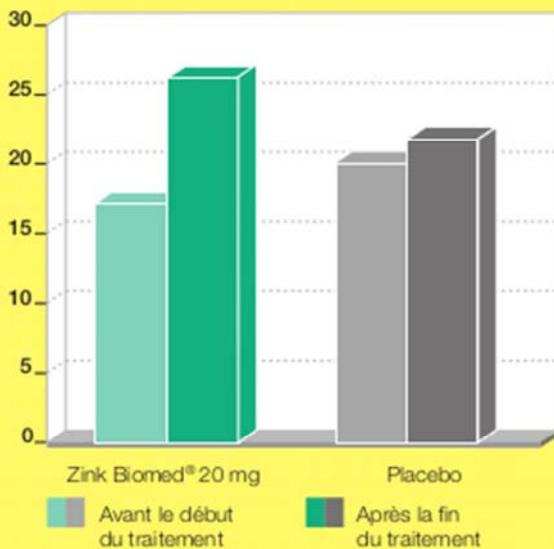


Ressources de l'EPFL



Amélioration des facultés gustatives

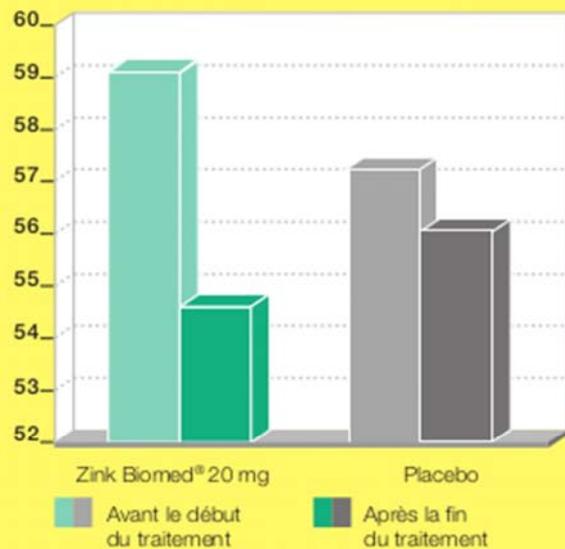
Nombre de bandes de papier filtre dont le goût a été correctement identifié.



Résultat: amélioration significative avec le zinc par rapport au placebo ($p < 0,001$).

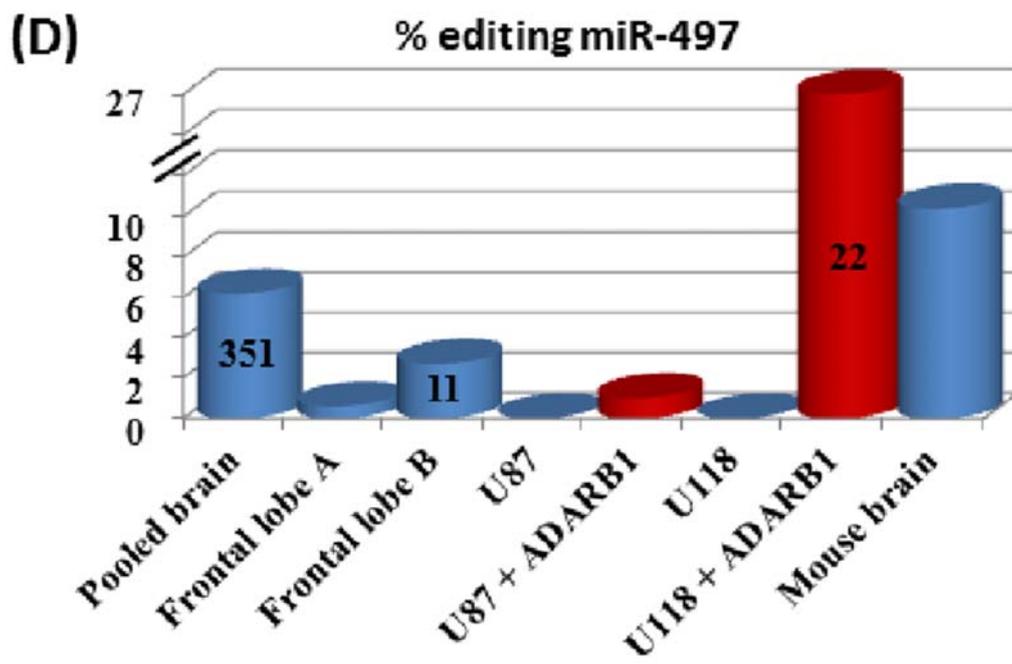
Diminution de la dysgueusie

Auto-évaluation de la dysgueusie (données en % d'une échelle visuelle analogique).



Résultat: amélioration significative avec le zinc par rapport au placebo ($p = 0,048$).

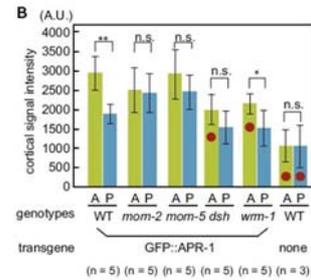
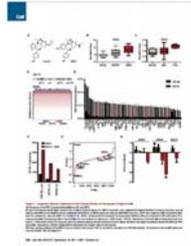
Figure 3: Résultats de l'étude de Heckmann et al. ayant évalué la supplémentation en zinc (2005) (85)



**The first rule of data analysis:
Show (and do not hide) your data !**



A (small) sample of graphics published in Cell and Genes and Development

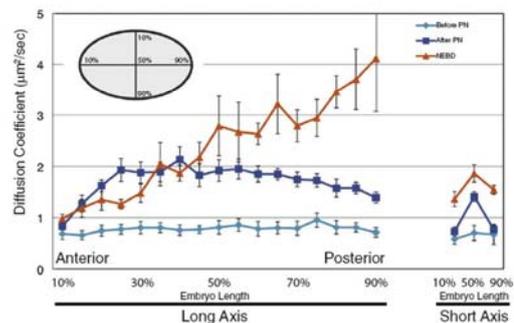
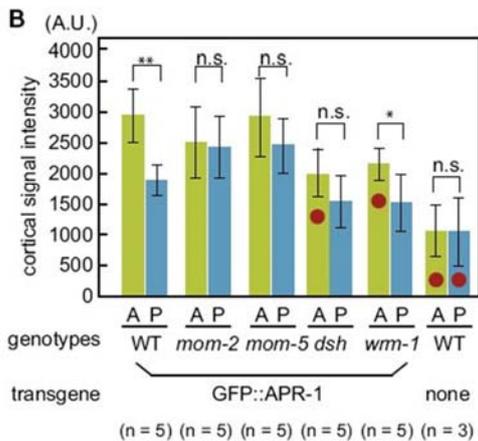
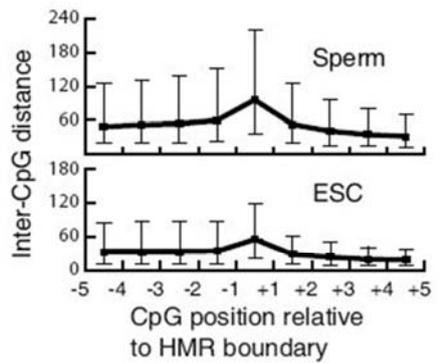
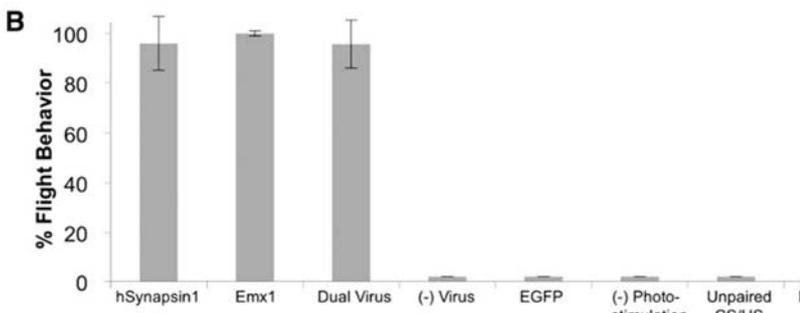


36 research articles...

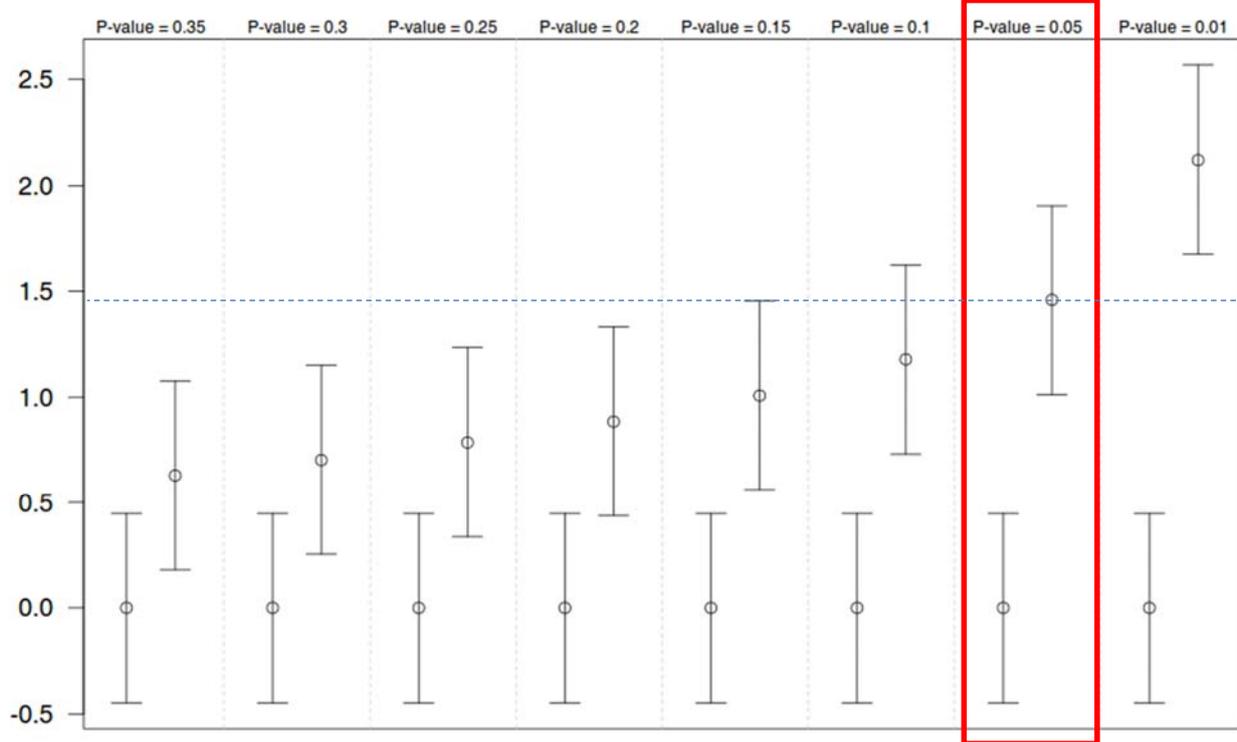
... 33 containing (statistical) graphics...

... 31 containing error bars.

Error bars are ubiquitous in the scientific literature



Which pair of error bars correspond to a p -value of 0.05 for a t -test ?



Researchers Misunderstand Confidence Intervals and Standard Error Bars

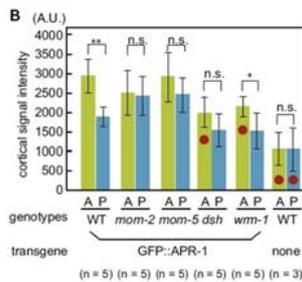
Sarah Belia, Fiona Fidler, Jennifer Williams, and Geoff Cumming
La Trobe University

Little is known about researchers' understanding of confidence intervals (CIs) and standard error (SE) bars. Authors of journal articles in psychology, behavioral neuroscience, and medicine were invited to visit a Web site where they adjusted a figure until they judged 2 means, with error bars, to be just statistically significantly different ($p < .05$). Results from 473 respondents suggest that many leading researchers have severe misconceptions about how error bars relate to statistical significance, do not adequately distinguish CIs and SE bars, and do not appreciate the importance of whether the 2 means are independent or come from a repeated measures design. Better guidelines for researchers and less ambiguous graphical conventions are needed before the advantages of CIs for research communication can be realized.

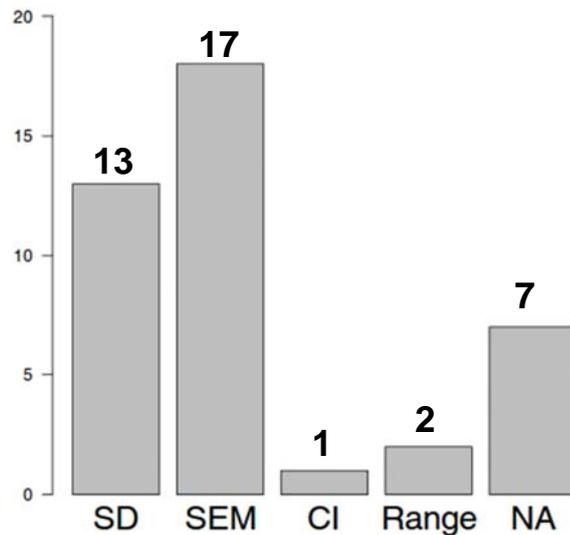
Keywords: confidence intervals, statistical cognition, standard error, error bars, statistical reform

We identified four different problems. First, responses were very widely spread and inaccurate: Only 22% of respondents set the means so the p value was between .025 and .10. Second, respondents overall did not adequately distinguish CIs and SE bars, as if they did not sufficiently recognize that a single graphic is used for two very different indicators of precision. Third, many respondents (overall 31.5%, 99/314) used the incorrect rule that error bars, whether a 95% CI or SE bars, just touch when means are just statistically significantly different ($p < .05$). Finally, for

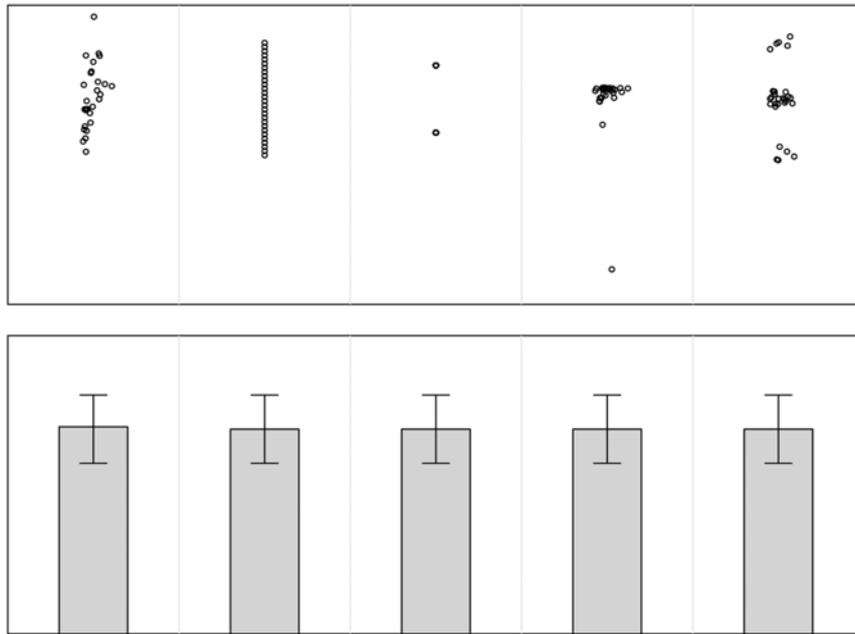
Error bars in our sample



... 31 containing error bars.



Five different datasets ... but five times the same bars and error bars !



Thank you !

Frederic.Schutz@unil.ch