FORMATION DES MAITRES D'EXPERIENCE

10 SEPTEMBRE 19

bruxelles environnement .brussels

Classification & Reporting of Severity

Notes de cours préparées par Anne VERMEYLEN
Université de Namur
et Anne-Dominique Degryse, David Anderson, David Smith

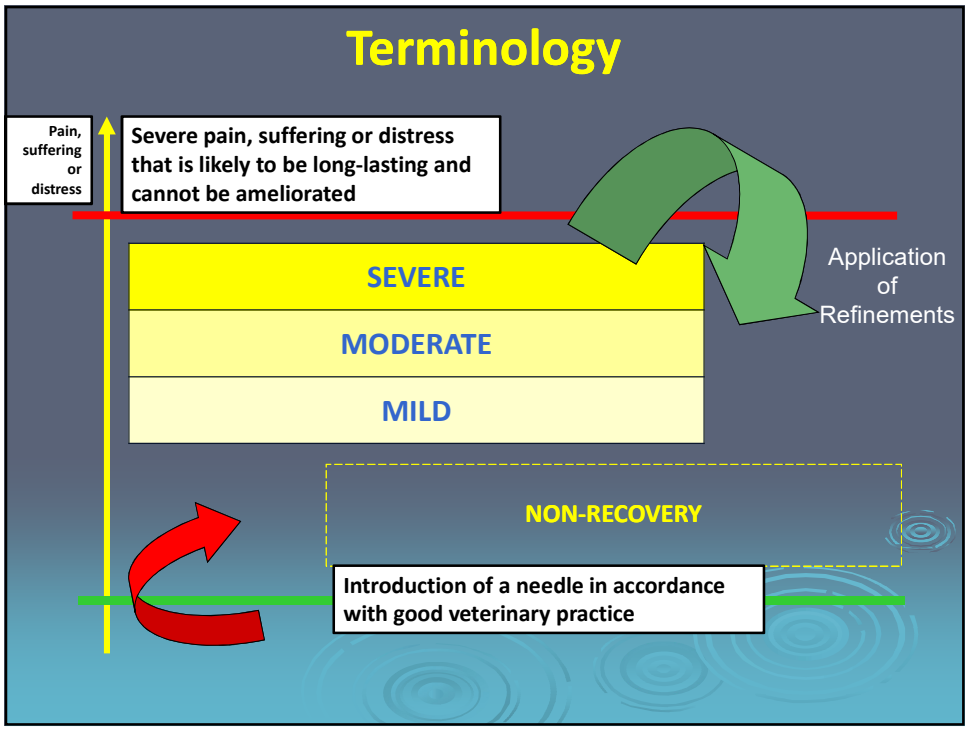
felasa ESLAV eclam bruxelles environnement .brussels

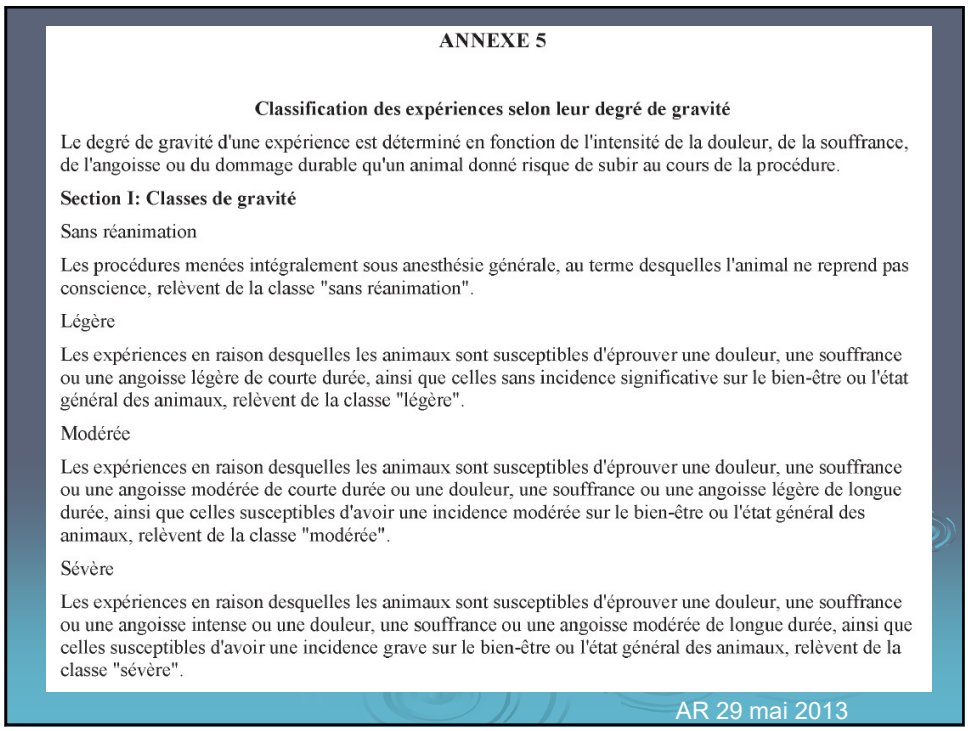
Severity Assessment
Legal framework under Dir 2010/63/EC

- **Article 4(3)** "…..eliminating or **reducing to the minimum** any possible pain, suffering, distress or lasting harm to the animals."
- **Article 15(1)** " …..**all procedures are classified** as 'non-recovery', 'mild', 'moderate', or 'severe' on a case- by-case basis using the assignment criteria set out in Annex VIII."
- **Article 54(2)** " … statistical information on the use of animals in procedures, including information on the **actual severity** of the procedures"
- **Article 16** – "**actual severity**" a factor to be considered in requests to **re-use**
- **Article 16(1)(d)** – "…, taking into account the **lifetime experience** of the animal."

Severity Assessment

- Classification of Severity of procedures (Art. 15, Ann. VIII)
  - « […] *a procedure is not to be performed if it involves severe pain, suffering or distress that is likely to be long-lasting and cannot be ameliorated.* »
  - **Prospective:** Mild – Moderate – Severe - Non-recovery
- Statistical Reporting (Art. 54)
  - Use of animals in procedures, including information on the **Actual severity** of the procedures and on the origin and species of NHP
  - Reporting on **annual basis**
- Reuse (Art. 16)
  - Animal's general state fully restored
  - **Restriction** according to first use (until *'moderate')*
  - **Further** procedure up to *'moderate'* or 'non-recovery'
  - Reuse possible **only** in accordance with veterinary advice, taking into account lifetime experience (**Cumulative suffering**)
  - Derogation for re-use after "severe" possible after veterinary examination





ANNEXE 5

Classification des expériences selon leur degré de gravité

Le degré de gravité d'une expérience est déterminé en fonction de l'intensité de la douleur, de la souffrance, de l'angoisse ou du dommage durable qu'un animal donné risque de subir au cours de la procédure.

Section I: Classes de gravité

Sans réanimation

Les procédures menées intégralement sous anesthésie générale, au terme desquelles l'animal ne reprend pas conscience, relèvent de la classe "sans réanimation".

Légère

Les expériences en raison desquelles les animaux sont susceptibles d'éprouver une douleur, une souffrance ou une angoisse légère de courte durée, ainsi que celles sans incidence significative sur le bien-être ou l'état général des animaux, relèvent de la classe "légère".

Modérée

Les expériences en raison desquelles les animaux sont susceptibles d'éprouver une douleur, une souffrance ou une angoisse modérée de courte durée ou une douleur, une souffrance ou une angoisse légère de longue durée, ainsi que celles susceptibles d'avoir une incidence modérée sur le bien-être ou l'état général des animaux, relèvent de la classe "modérée".

Sévère

Les expériences en raison desquelles les animaux sont susceptibles d'éprouver une douleur, une souffrance ou une angoisse intense ou une douleur, une souffrance ou une angoisse modérée de longue durée, ainsi que celles susceptibles d'avoir une incidence grave sur le bien-être ou l'état général des animaux, relèvent de la classe "sévère".

AR 29 mai 2013

BIJLAGE 5

Indeling naar ernst van de dierproeven

De ernst van een dierproef wordt bepaald aan de hand van mate van, pijn, lijden, angst of blijvende schade die een individueel dier tijdens de procedure naar verwachting zal ondervinden.

Deel I: Categorieën ernst

Terminaal:

Dierproeven die worden uitgevoerd onder algemene verdoving en aan het eind waarvan het dier niet meer bij bewustzijn komt, worden ingedeeld als terminaal.

Licht:

Dierproeven waarbij de dieren waarschijnlijk gedurende korte tijd een lichte vorm van pijn, lijden of angst zullen ondervinden, en dierproef die geen significante hinder voor het welzijn of de algemene toestand van de dieren opleveren, worden ingedeeld als licht.

Matig:

Dierproeven waarbij de dieren waarschijnlijk gedurende korte tijd een matige vorm van pijn, lijden of angst, dan wel langdurig een lichte vorm van pijn, lijden of angst zullen ondervinden en dierproeven die waarschijnlijk een matige hinder voor het welzijn of de algemene toestand van de dieren zullen opleveren, worden ingedeeld als matig.

Ernstig:

Dierproeven waarbij de dieren waarschijnlijk een ernstige vorm van pijn, lijden of angst, dan wel langdurig een matige vorm van pijn, lijden of angst zullen ondervinden en dierproeven die waarschijnlijk ernstige hinder voor het welzijn of de algemene toestand van de dieren zullen opleveren, worden ingedeeld als ernstig.

KB 29 mei 2013

Severity Assessment
When is it important?

- **Prospective** – offers opportunity to consider appropriateness of design & ensure suffering is minimised
- **Ongoing** – important to remain within limits of harm/benefit evaluation
- **Retrospective** – offers review of actual severity before further work undertaken ; allows publication of actual severity

Severity Assessment
Why is it important?

- Ongoing opportunities in particular to **implement Refinement** and reduce suffering
- Improved animal **welfare**
- Improved **scientific data quality** due to better welfare
- Improved **communication** between those responsible for using, caring for and monitoring animals
- Input to **retrospective project assessment** when this is required
- Improved **transparency**

Procedure planning

What does this study involve doing to the animals?	What will the animals experience? How much suffering might it cause? What might make it worse?	How will suffering be reduced to a minimum?	
	Adverse effects	Methodology and interventions	End-Points

PDF-XChange Viewer Click to buy NOW! www.docu-track.com

CONCLUSION

bruxelles environnement .brussels

CONCLUSION

- Encore des questions?
- Envoyez-les à :

- bienetreanimal@environnement.brussels
- dierenwelziin@leefmilieu.brussels

Merci pour votre attention!

FORMATION DES MAITRES D'EXPERIENCE

10 SEPTEMBRE 19

bruxelles environnement .brussels

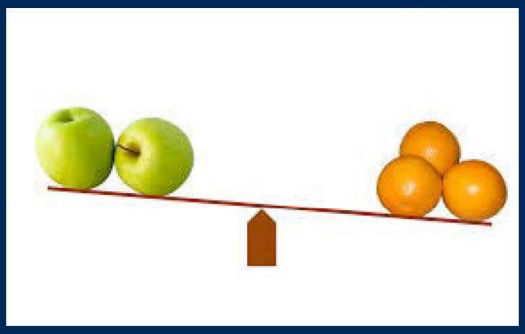
HARM-BENEFIT ANALYSIS

Notes de cours préparées par Anne VERMEYLEN Université de Namur et Javier Guillèn AAALAC

bruxelles environnement .brussels

Version originale de A. Vermeylen (UNamur), 2019

Ethical Balance of Different Things



Risk – Benefit

↓

Cost – Benefit

↓

Harm – Benefit

Protocol Evaluation Approach

"Technical" Details

Scientific objectives
Alternatives?
Personnel
Animal
Exp. Design & statistics
Housing conditions
Procedures
Pain/suffering relief
Humane endpoints
Euthanasia
Safety

Harm/Benefit

Harm-Benefit Analysis Framework

Directive 2010/63/EU. Art. 38, 2, d.

2. The project evaluation shall consist in particular of the following:
(a), (b), (c), (e), (f)
(d) a harm-benefit analysis of the project, to assess whether the harm to the animals in terms of suffering, pain and distress is justified by the expected outcome taking into account ethical considerations, and may ultimately benefit human beings, animals or the environment;

Harm-Benefit Analysis Framework

Guide for the Care and Use of Laboratory Animals
(NRC, 2011) p.27

"…the IACUC is obliged to weigh the objectives of the study against potential animal welfare concerns."

Harm-Benefit Analysis Framework

https://www.oie.int/doc/ged/D10905.PDF

Terrestrial Code Chapter 7.8.
Use of Animals in Research and Education

Ethical review: means consideration of the validity and justification for using *animals* including: an assessment and weighing of the potential harms for *animals* and likely benefits of the use and how these balance;…

Harm-benefit analysis: means the process of weighing the likely adverse effects (harms) to the *animals* against the benefits likely to accrue as a result of the proposed project.

Project Proposal Review: …(i) ethical considerations such as the application of the Three Rs and a harm/benefit analysis; the benefits should be maximised and the harms, in terms of pain and distress, should be minimized;

Harm-Benefit Analysis Framework

http://iclas.org/wp-content/uploads/2013/03/CIOMS-ICLAS-Principles-Final.pdf

International Guiding Principles for Biomedical Research Involving Animals (2012)

P.I: "…Decisions regarding the welfare, care and use of animals should be guided by scientific knowledge and professional judgement, reflect ethical and societal values, and consider the potential benefits and the impact on the well-being of the animals involved".

P.X: "The oversight framework… should promote a harm-benefit analysis for animal use, balancing the benefits derived from the research or educational activity with the potential for pain and/or distress experienced by the animal".

Harm-Benefit Analysis Framework

http://www.aaalac.org/accreditation/faq_landing.cfm#B3)

AAALAC International expects that IACUC's (or comparable oversight body), as part of the protocol review process, will weigh the potential adverse effects of the study against the potential benefits that are likely to accrue as a result of the research. This analysis should be performed prior to the final approval of the protocol, and should be a primary consideration in the review process.

Dimensions of Harm (literature)

- **Species**, choice of animals
- Sentience and consciousness
- Quality of animals
- Duration
- Duration related to lifespan
- **Number** of animals
- Origin, acquisition or transport
- Care, housing factors, handling, health care
- Possibility to express Normal Behaviour
- **Staff** competence and quality
- Hunger and Thirst
- Discomfort
- **Pain**
- Injury or Disease
- Fear, anxiety and distress
- Frequency of procedures
- Severity of procedures
- Risk of harm = probability x severity
- Deaths (caused by the experiment)
- Intrinsic value and animal rights
- Genetic modulation of animals - respect for nature

- **Aim, Realistic potential**
- **Scientific Quality**
- **Non-publishing of negative results**

Dimensions of Harm (literature)

The Five Freedoms

1. Freedom from Hunger and Thirst
2. Freedom from Discomfort
3. Freedom from Pain, Injury or Disease
4. Freedom to Express Normal Behaviour
5. Freedom from Fear and Distress

- **Brambell, R., Five Freedoms. 1965, Farm Animal Welfare Council**
- **Mellor&Reid, Concepts of animal well-being and predicting the impact of procedures on experimental animals 1994**

”There is a danger that with focus largely on suffering we could overlook a broader view of welfare which may be more informative and safeguard more effectively the interests of the experimental animals ”

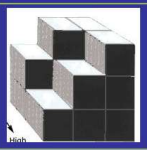
Dimensions of Benefit (literature)

- **Benefits for humans**
- **Benefits for animals**
- **Benefits for environment**
- Economic interests
- **Health interests**
- **Safety interests**
- **Knowledge interests**
- **Educational interest**

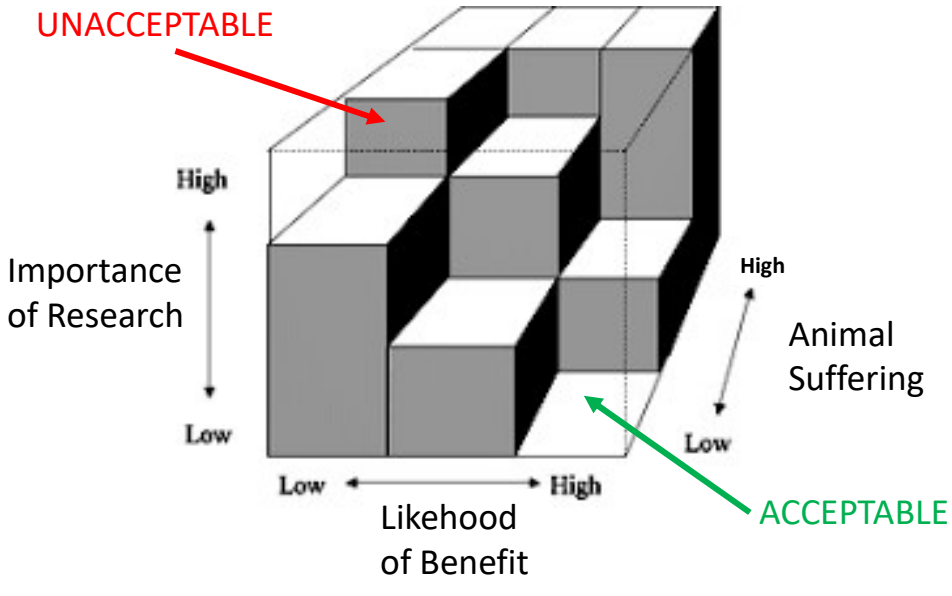
- Primary (direct) versus secondary (indirect) benefits
- ”Surrogate outcomes" versus "health outcomes”
- Originality
- Dissemination of results

- **Aim, Realistic potential**
- **Quality, "good science"**
- **Non-publishing of negative results**

Models of Harm-Benefit Analysis (literature)

	Strengths	Weaknesses
Tables, spreadsheet	**Categories** are useful to simplify a complex picture. Stimulate actions to avoid severe categories.	The categories do not fit all cases
$E = mc^2$ $\Sigma = \pi e^{HBA}$	**Algorithms** are helpful in guiding a decision	Moral dilemmas cannot/shall not be solved by arithmetic's
	**Graphic** representations have pedagogic value in visualizing the concept and relationship between harm and benefit	Depend on defined categories (eg. low-middle-high) Not operational (too simple?)
	**Process** oriented models structure the HBA process, how to balance different opinions and question quality of the analysis. Generic	Does not provide an answer on what model (as previous) to use or provide solutions for conclusions (too generic?)

BATESON'S CUBE



Harm-Benefit Analysis Extensive Summary

- HBA is a systematic way to assess and compare harms, benefits and how they are balanced
- HBA must be transparent and verifiable
- HBA identifies harm – and stimulate researchers to seek alternative approaches
- HBA is a tool to make sure that animals are only used when it is justified because of potential benefit
- HBA clarifies if harm is necessary for achieving certain benefits
- HBA is important for public relations
- HBA is important to avoid uncritical use of animals even for the cause of the good
- HBA provides an ethical framework and is an essential part of the ethical review
- Harm Benefit analysis is based on utilitarian consequence ethics
- HBA stimulates ethical reflection and discussion
- HBA is dependent on and limited to the current context (external factors)
- HBA is influenced by subjective opinions (“affective heuristics”)

https://www.blv.admin.ch/blv/en/home/tiere/tierversuche/schweregrad-gueterabwaegung.html

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA
Federal Food Safety and Veterinary Office FSVO

Federal Department of the Environment, Transport, Energy and Communications DETEC
Federal Office for the Environment FOEN
Soil and Biotechnology Division

01.05.2017

Dignity of the animal: guide to the 'weighing of interests'

1. **Description of the aim of the proposed intervention**

2. **Presentation of the facts**

3. **Question of suitability**

   Can the intended aim be achieved by the proposed intervention?

   - Yes -> carry out a weighing of interests
   - No -> do not carry out intervention, no need for a weighing of interests

4. **Question of necessity**

   Is the proposed intervention necessary in order to achieve the intended aim, or can the aim be achieved by means that entail less or no strain for the animal?
   - No alternative available -> carry out a weighing of interests
   - Alternative entailing no strain for the animal is available -> choose alternative, no need for a weighing of interests

5. **Identification and assessment of strain**

If an alternative is to be assessed alongside the proposed intervention, this can be done in the same table (use different colours for the proposed intervention and the alternative).

	Are the following forms of strain present?	No	Yes	Which, specifically?	Weighting		
					*	**	***
1	Pain, suffering, anxiety						
2	Harm, especially impairment of growth, reproductive capacity, adaptive capacity, mobility, species-specific social behaviours						
3	Major interference with the appearance						
4	Humiliation, excessive instrumentalisation						
5	Other						

6. Identification and evaluation of legitimate interests

	Are there legitimate interests in the following areas?	No	Yes	Which, specifically?	Weighting *	**	***	****
1	Human and/or animal health							
2	Expanding scientific knowledge							
3	Preservation and improvement of environment							
4	Protection against violation of fundamental rights such as economic freedom, freedom of ownership, freedom of research, freedom of association							
5	Other							

7. Comparison: strain vs. legitimate interests

If an alternative is to be assessed alongside the proposed intervention, this can be done in the same table (use different colours for the proposed intervention and the alternative).
The crucial factors are: the severest form of strain and the most significant interest.

Dignity of the animal is respected		Legitimate interests *	**	***	****
Strain	*	no	yes	yes	yes
	**	no	no	yes	yes
	***	no	no	no	yes

Conclusion:

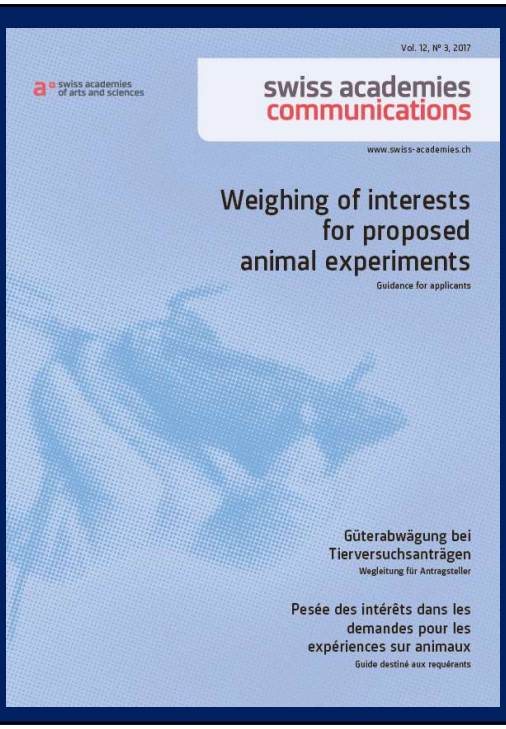
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA
Federal Food Safety and Veterinary Office FSVO

Federal Department of the Environment, Transport, Energy and Communications DETEC
Federal Department of Environment FOEN
Soil and Biotechnology Division

01.05.2017

Dignity of the animal
Explanatory notes on the 'weighing of interests'



Harm-Benefit Analysis: the AALAS-FELASA Working Group Proposal

Working Party Report

Current concepts of Harm–Benefit Analysis of Animal Experiments – Report from the AALAS–FELASA Working Group on Harm–Benefit Analysis – Part 1

Aurora Brønstad[1], Christian E Newcomer[2], Thierry Decelle[3], Jeffrey I Everitt[4], Javier Guillen[5] and Kathy Laber[6]

Laboratory Animals
2016, Vol. 50(1S) 1–20

DOI: 10.1177/0023677216642398
la.sagepub.com
SAGE

- Consideration of harms is based upon the Five Freedoms and a set of "modulating factors" that may have "mitigating" or "aggravating" effects
- Consideration of benefits is based upon a specific set of domains (what, who, how, when) and "modulating factors" that may have "mitigating" or "aggravating" effects

KEY

Benefit = Domains
Social benefits • Human health • Animal health • Environmental health
Socioeconomic benefits
Scientific benefits
Educational benefits
Safety and efficacy

Harm = Impact on Five Freedoms
Freedom from Pain/Injury
Freedom from Fear/Distress
Freedom from Hunger/Thirst
Ability to express normal behavior
Freedom from Discomfort/appropriate husbandry

Working Group Report

Recommendations for Addressing Harm–Benefit Analysis and Implementation in Ethical Evaluation – Report from the AALAS–FELASA Working Group on Harm–Benefit Analysis – Part 2

Kathy Laber[1], Christian E Newcomer[2], Thierry Decelle[3], Jeffrey I Everitt[4], Javier Guillen[5] and Aurora Brønstad[6]

Laboratory Animals
2016, Vol. 50(1S) 21–42

DOI: 10.1177/0023677216642397
la.sagepub.com
SAGE

CONCLUSION

bruxelles environnement .brussels

CONCLUSION

- Encore des questions?
- Envoyez-les à :

- bienetreanimal@environnement.brussels
- dierenwelzijn@leefmilieu.brussels

Merci pour votre attention!

PDF-XChange Viewer
Click to buy NOW!
www.docu-track.com

Experimental design and statistics

Etienne Hanon
Stat Solutions
Etienne.hanon@gmail.com

Question :

- What is the historical connection between
  - experimental design
  - statistics
  - beer

Answer :

- Statistics suitable for experiments based on small sample size are born in the brewery

William Gosset, pseudonyme : Student

The historical article of Gosset :

The Probable Error of a Mean

Student

Biometrika, Volume 6, Issue 1 (Mar., 1908), 1-25.

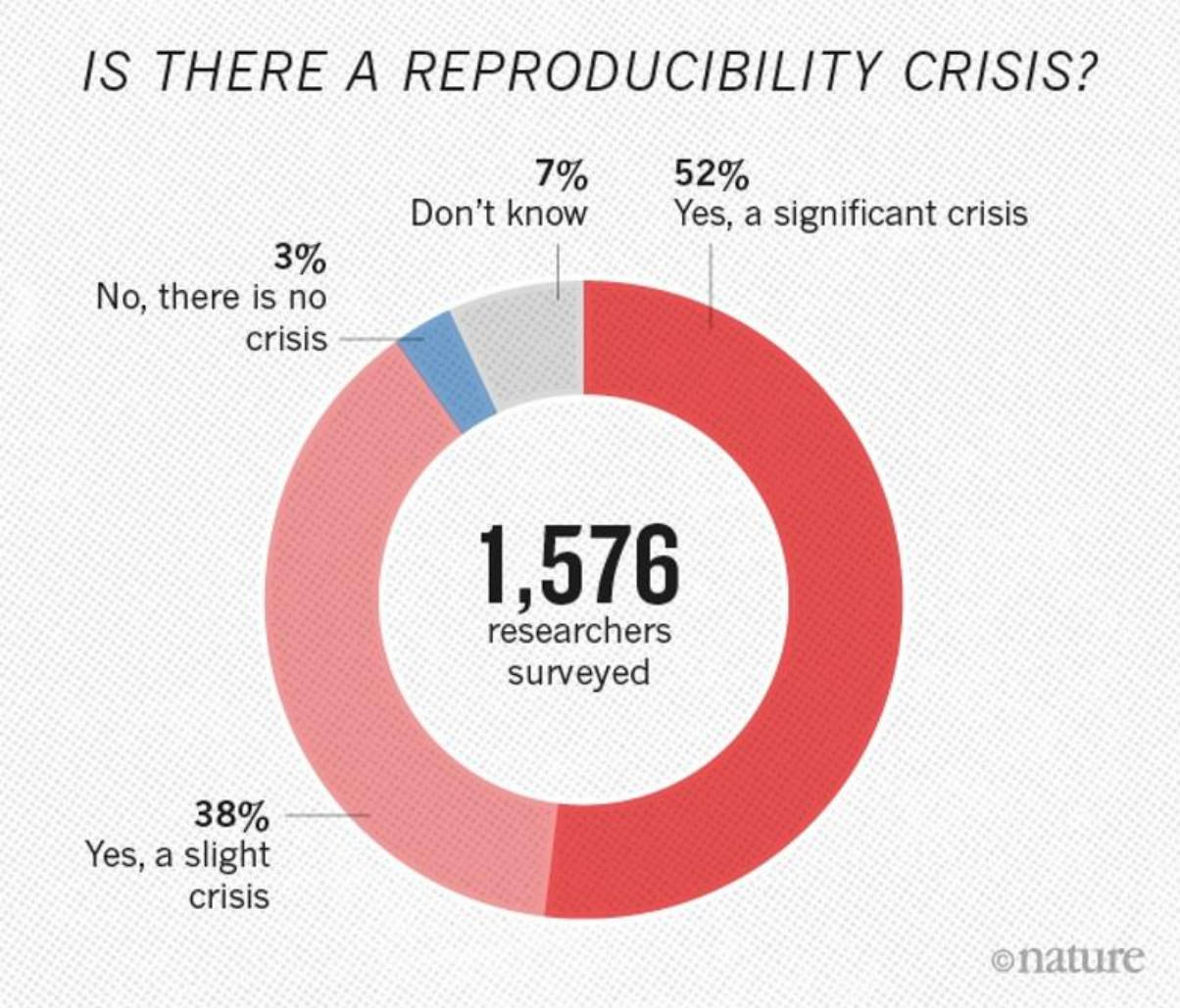
- Fisher: "The value for which P = .05, or 1 in 20... it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant."
- Gosset was more concerned with whether a result was practically meaningful than whether it was statistically "significant."

Scientific results and reproducibility ?

- Who has already experienced difficulties in reproducing his own experiences?

- Who has already experienced difficulties in reproducing published experiences?

1,500 scientists lift the lid on reproducibility



Monya Baker

25 May 2016 | Corrected: 28 July 2016

A crisis?

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

Published online 2005 Aug 30. doi: 10.1371/journal.pmed.0020124

Why Most Published Research Findings Are False

John P. A. Ioannidis

▸ Author information ▸ Copyright and License information Disclaimer

John P. A. Ioannidis is a physician-scientist, professor at Stanford , involved in evidence-based medicine, epidemiology, data science and clinical research. Pioneered the field of meta-research (research on research).

Festing's list of methodological issues that lead to false positive results includes:

- Selective publication of positive results
- Incorrect randomization
- Failure to blind wherever possible
- Psuedo-replication and incorrect identification of the animal used
- Failure in quality control of experimental materials
- Inadequate external validity
- Inadequate description of methods (e.g., strain nomenclature)
- Incorrect statistical analysis (e.g., no analysis; multiple testing without adjustment; wrong statistical model; incorrect treatment of outliers; cherry-picking the data).

Michael F.W. Festing Author of « The Design of Animal Experiments (Reducing the use of animals in research through better experimental design)

NATURE | NEWS

Poorly designed animal experiments in the spotlight

High-status journals or institutions no guarantees of carefully-reported trials.

Daniel Cressey

13 October 2015

Preclinical research to test drugs in animals suffers from a "substantial" risk of bias because of poor study design, even when it is published in the most-acclaimed journals or done at top-tier institutions, an analysis of thousands of papers suggests.

ACCESS Freely available online

PLOS

How to Make More Published Research True

John P. A. Ioannidis[1,2,3,4]*

1 Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, United States of America, **2** Department of Medicine, Stanford Prevention Research Center, Stanford, California, United States of America, **3** Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, United States of America, **4** Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, United States of America

possibilities include :

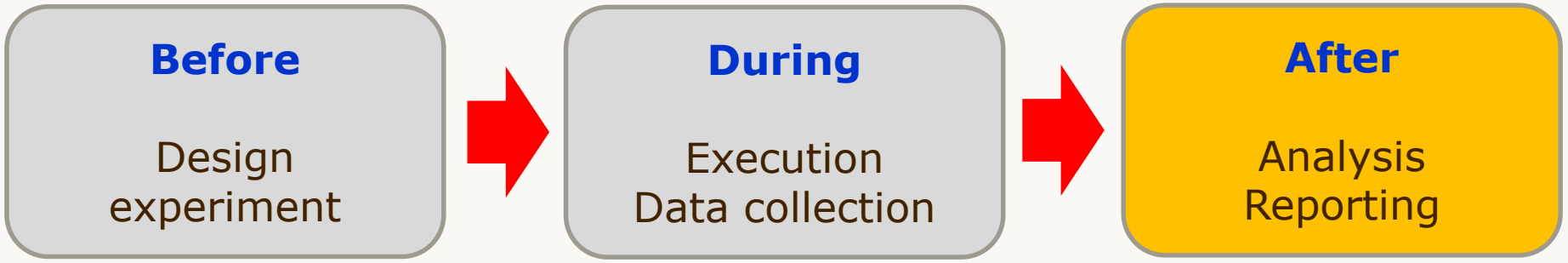
- the adoption of large-scale collaborative research
- replication culture
- registration
- standardization of definitions and analyses
- more appropriate (usually more stringent) statistical thresholds
- improvement in study design standards

How to improve ?

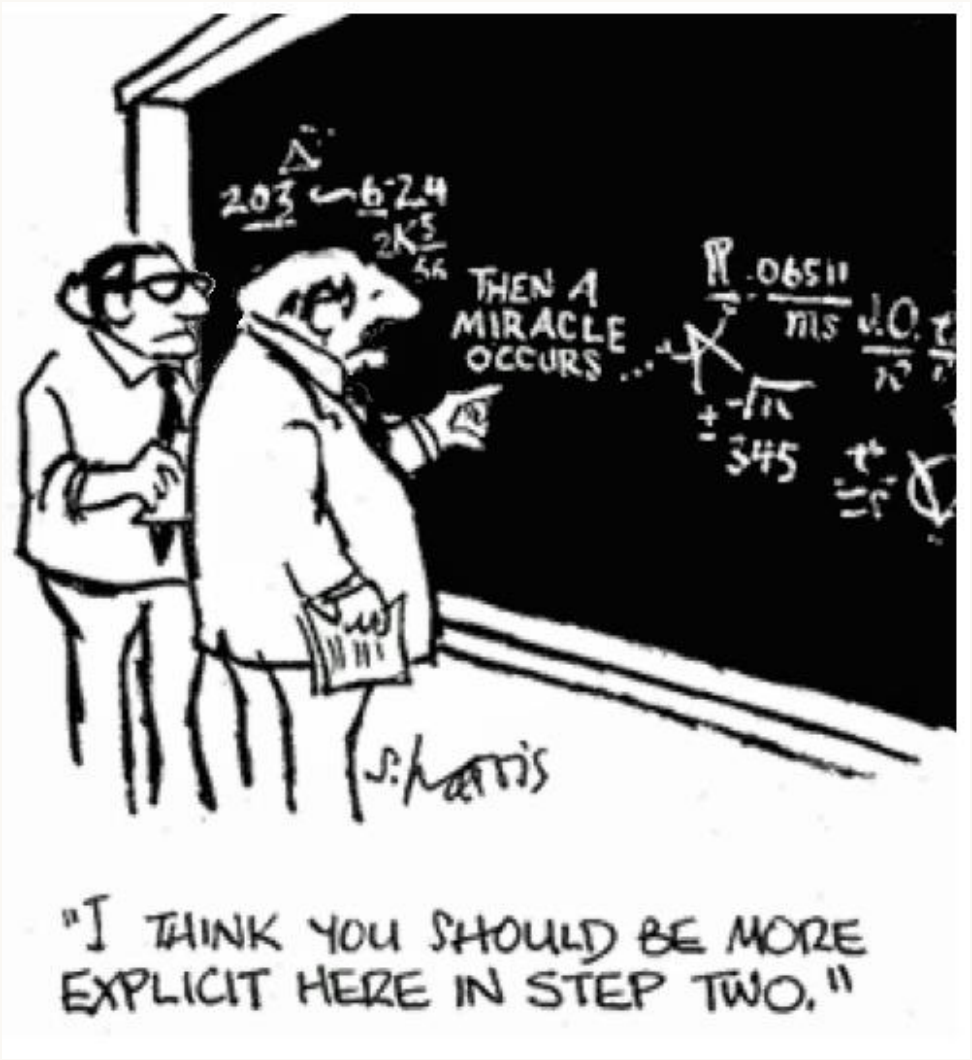
Some initiatives to improve reproducibility and robustness of biomedical studies :

- **PREPARE (Planning Research and Experimental Procedures on Animals : Recommendations for Excellence**
- **ARRIVE Guidelines Checklist Reporting In Vivo Experiments**
- **EQUATOR Network : Enhancing the Quality and Transparency Of health Research (Reporting guidelines)**
- **Conferences dedicated to the reproducibility of experiments and meta-analyzes, organized by Uliège(Ezio Tirelli, Pierre Drion)**
- **Books :**
  - **Experimental Design for Laboratory Biologists (Maximising information and improving reproducibility) by Stanley E. Lazic**
  - **The Design of Animal Experiments (Reducing the use of animals in research through better experimental design) by Michael F.W. Festing**

Improving data quality : Scientific Workflow, where to act ?

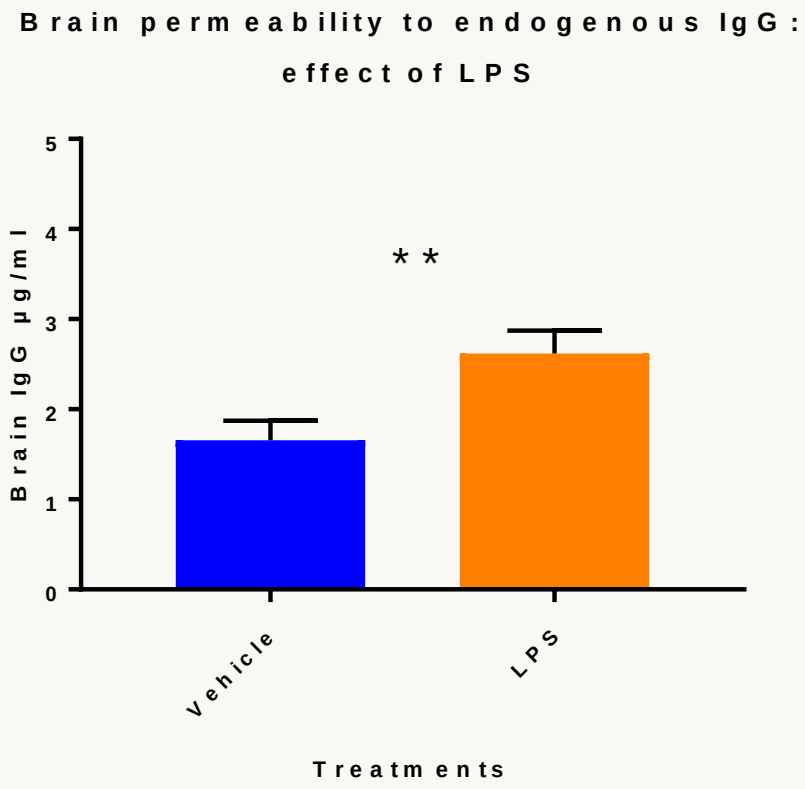


Reporting
Improve Communication of scientific results



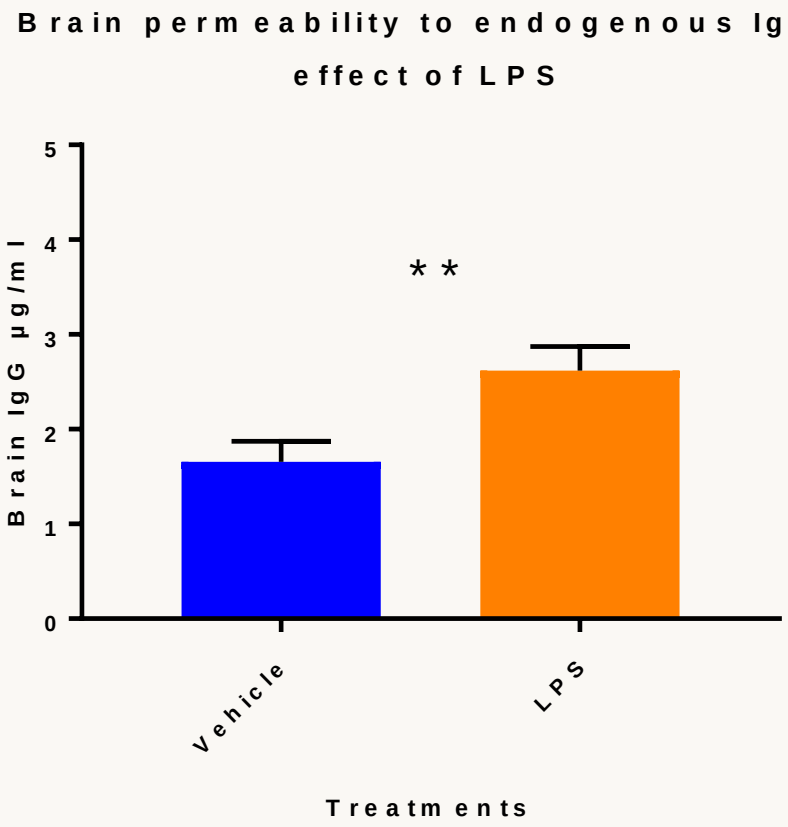
Improve presentation of the data

Blood brain barrier opening in the rat in order to increase penetration of large molecules into the brain



What is your feeling with this graph?

Mandatory information :



Means + SEM

N=16

Mann-Withney test, p value = 0.0056

Student t test, p value = 0.0078

SEM versus SD

- When sample size increases how is the SD expected to evolve :

  Increase, decrease or stay about the same?

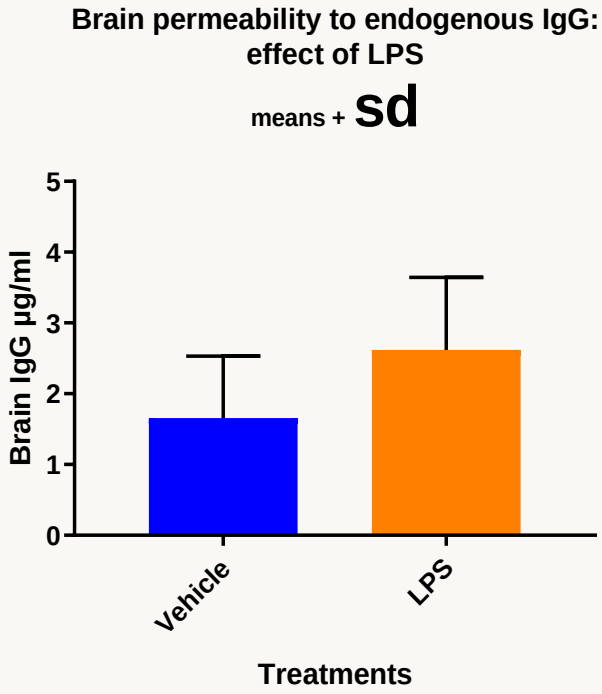
- When sample size increases how is the SEM expected to evolve :

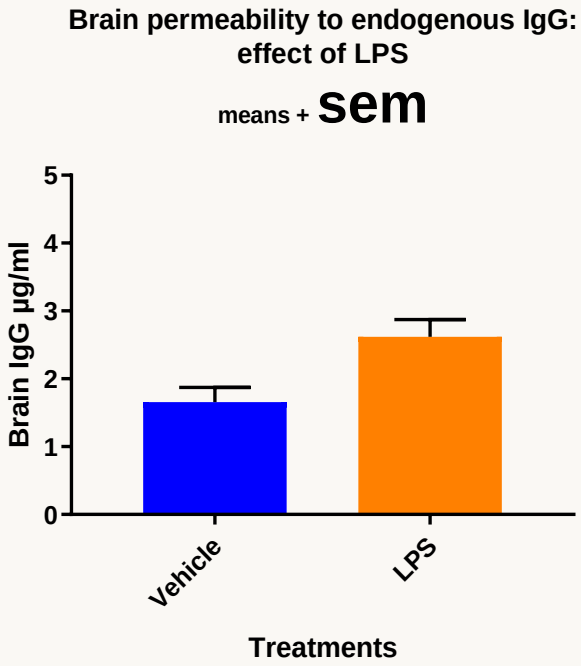
  Increase, decrease or stay about the same?

Error bars : Sd or Sem ? That is the Question.

Science or Propaganda?

Improving Transparency



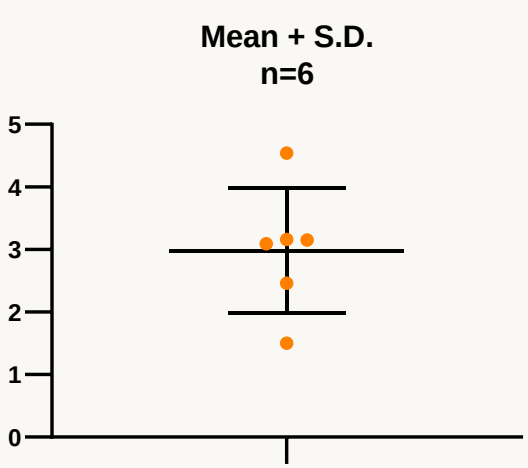


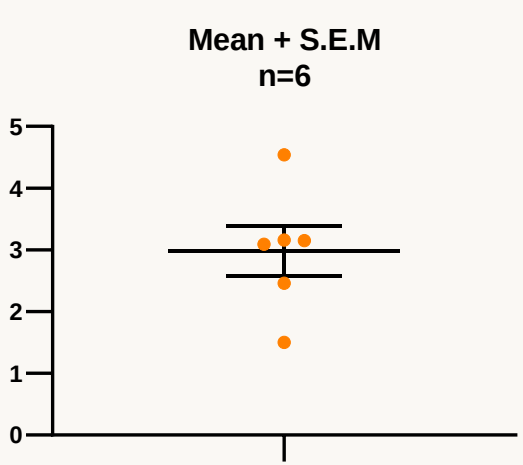
SD : *standard deviation quantifies* the dispersion of the data around the mean (roughly 2/3 of the data lie in the range : mean ± sd)

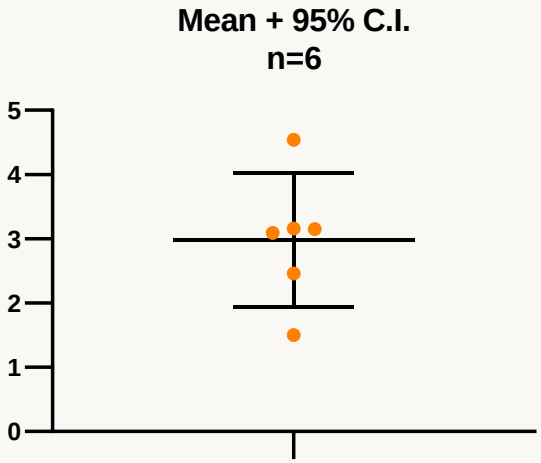
SEM : *standard error of the mean or standard error* quantifies how precisely you estimate the population mean (linked to confidence interval)

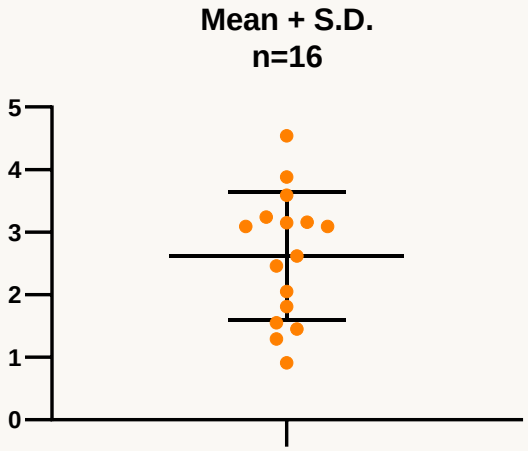
$$sem = \frac{s_d}{\sqrt{n}}$$

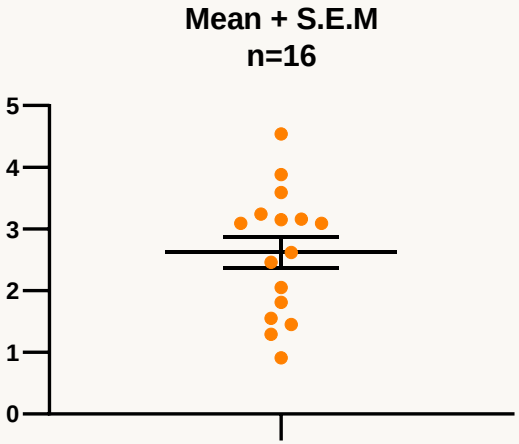
Errors Bars Comparison

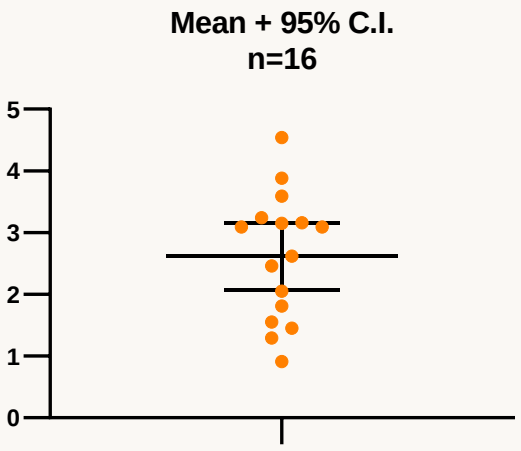




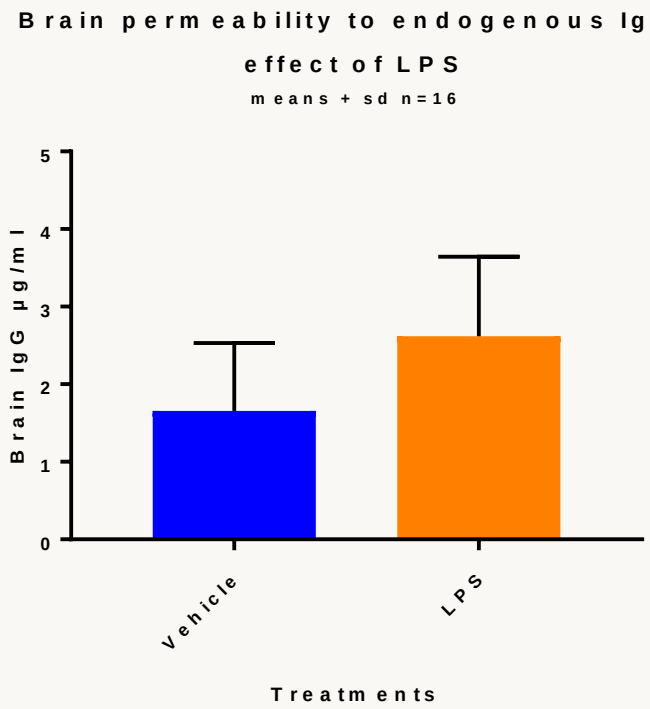


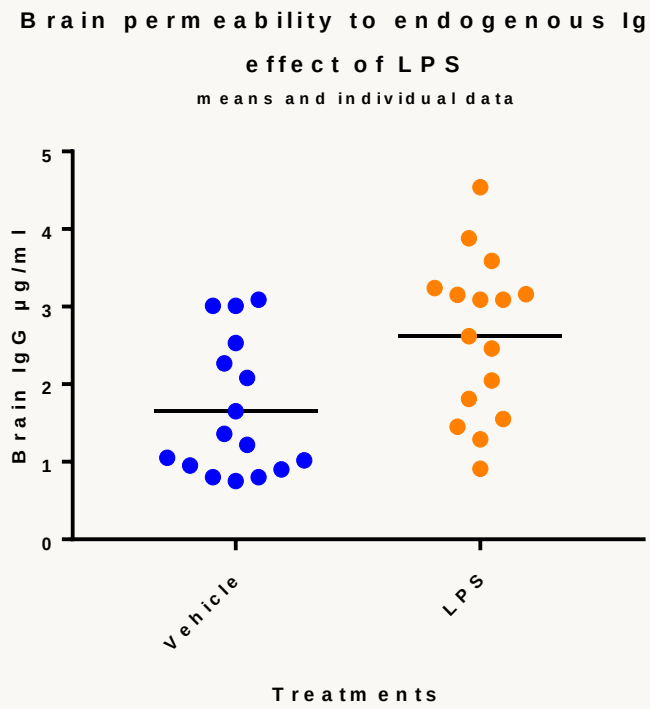






The Full Transparency

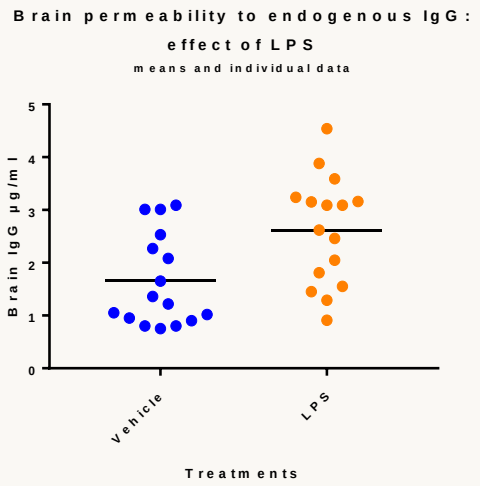




Avoid Scrubing the Data !
Do not remove Outliers (for your hypothesis !)

The MagicEraser of the most Resistant Outliers

How to express the effect amplitude ?



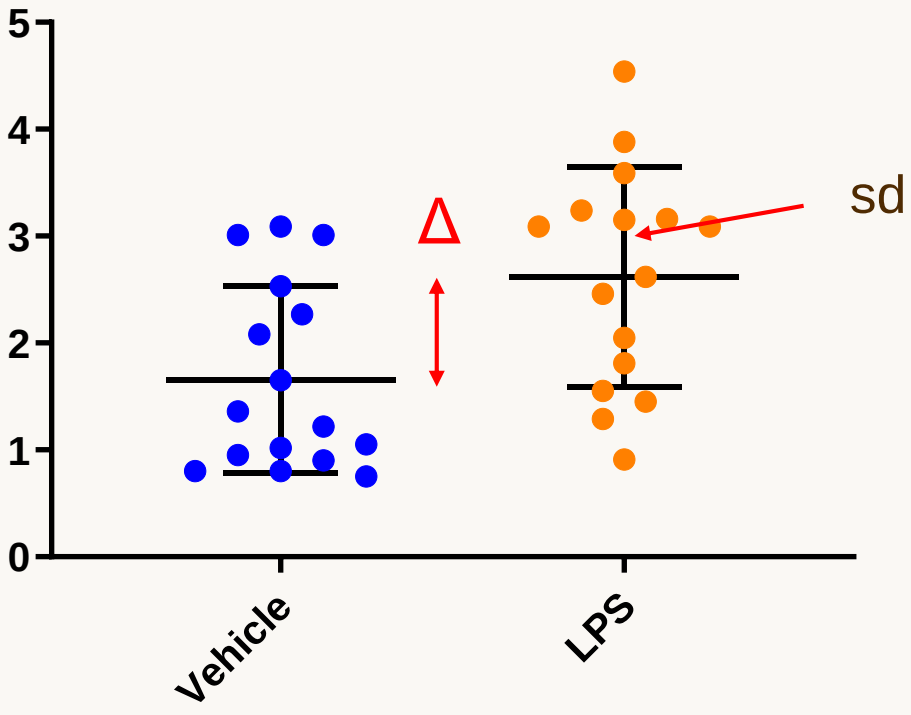
Brain permeability IgG µg/ml		
	Treatments	
	Control	LPS
Average	1.66	2.62
Standard dev.	0.87	1.03
n	16	16

Effect Amplitude:

Brain permeability IgG µg/ml		
	Treatments	
	Control	LPS
Average	1.66	2.62
Standard dev.	0.87	1.03
n	16	16

Difference	Difference %
0.96	58.1

Effect size or Standardized Effect Size or Standardized Difference

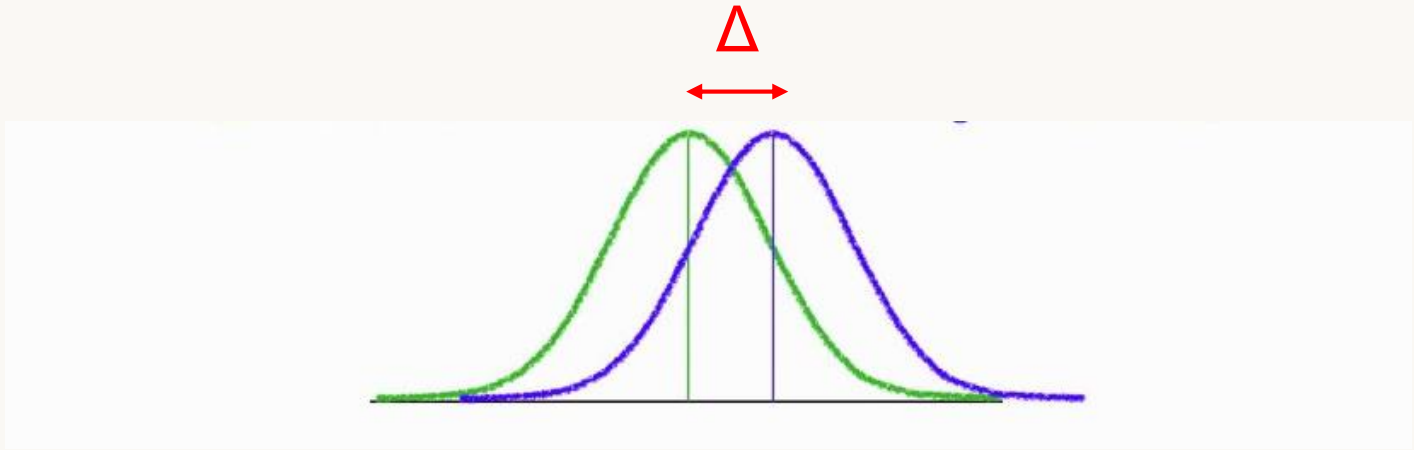


$$ES = \frac{\overline{x1} - \overline{x2}}{sd} = \frac{\Delta}{sd}$$

- The effect size represents the magnitude of the effect on the scale of variability ≈ the scale of physiological variation
- Effect size is independent of the sample size

Standardized Effect size

$$ES = \frac{\overline{x1} - \overline{x2}}{sd} = \frac{\Delta}{sd}$$

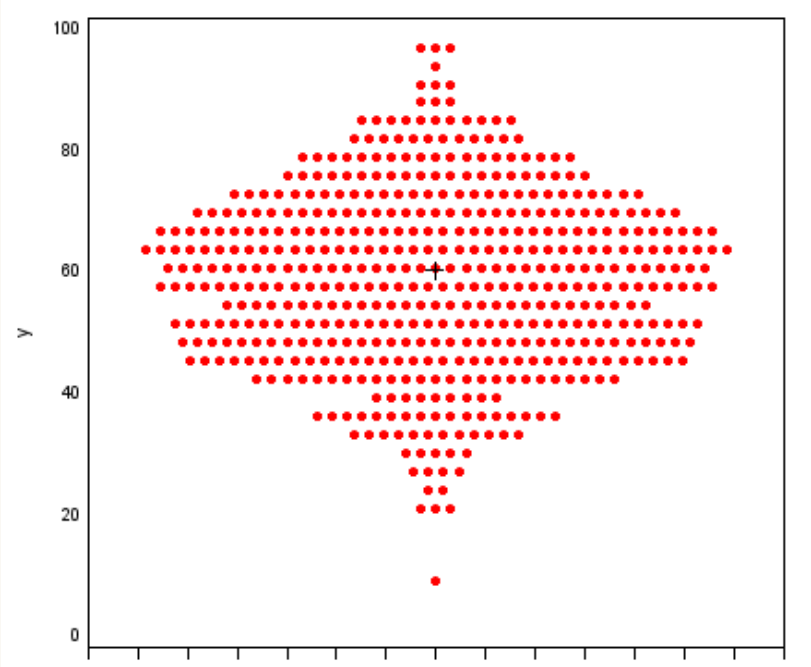


- ! Cohen's d (american psychologist and statistician)

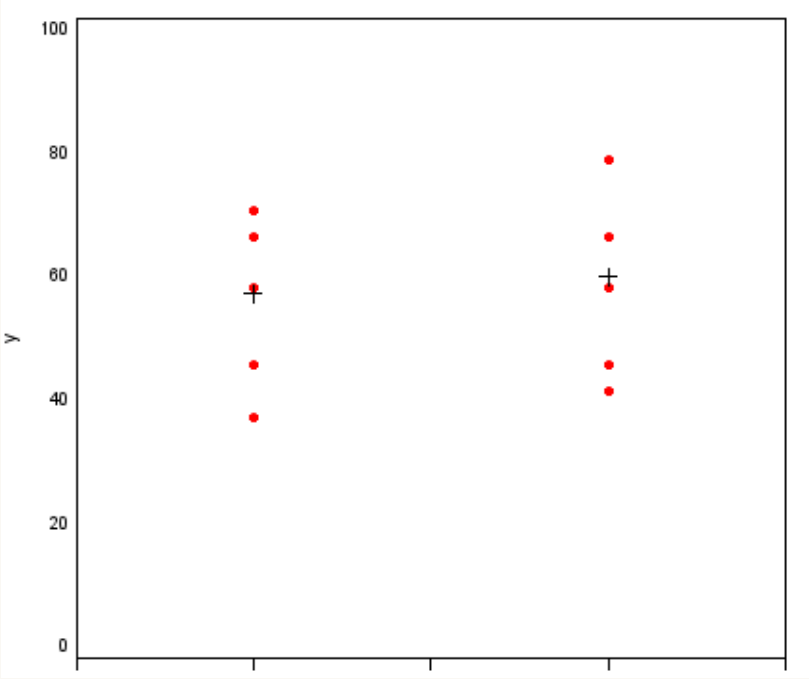
Since ancient Egypt

P value, the meaning :

Population



2 random samples (n=5) from the Population



The P value is the probability to get by chance a difference as the one observed (or greater) between the 2 samples (n=16) if they were randomly drawn from a single infinite population (equivalent to : 2 identical populations = The null hypothesis)

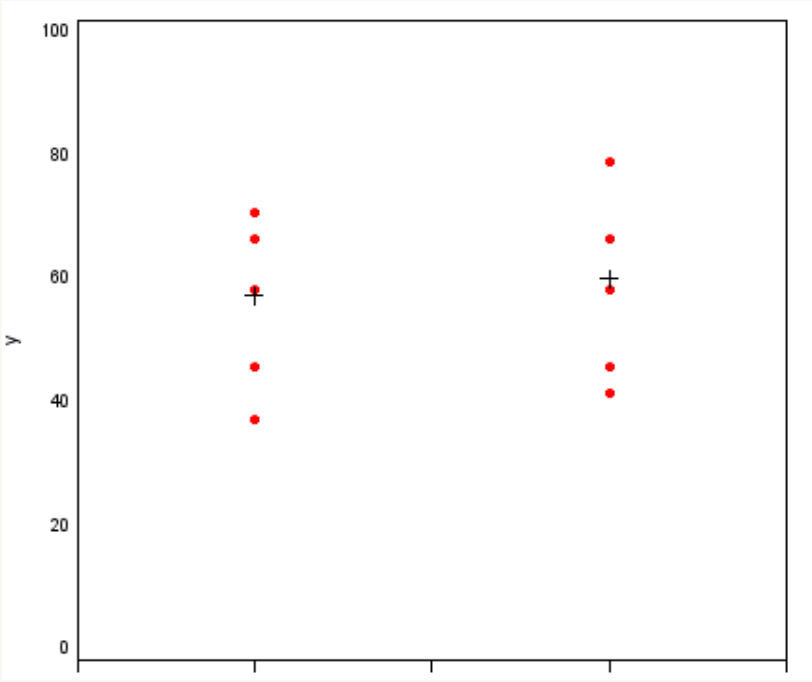
For the 2 extracted samples, Student T test

n	mean 1	mean 2	d	sd	es	p value
5	57.4	60.2	2.9	14.6	0.20	0.764

P Values examples

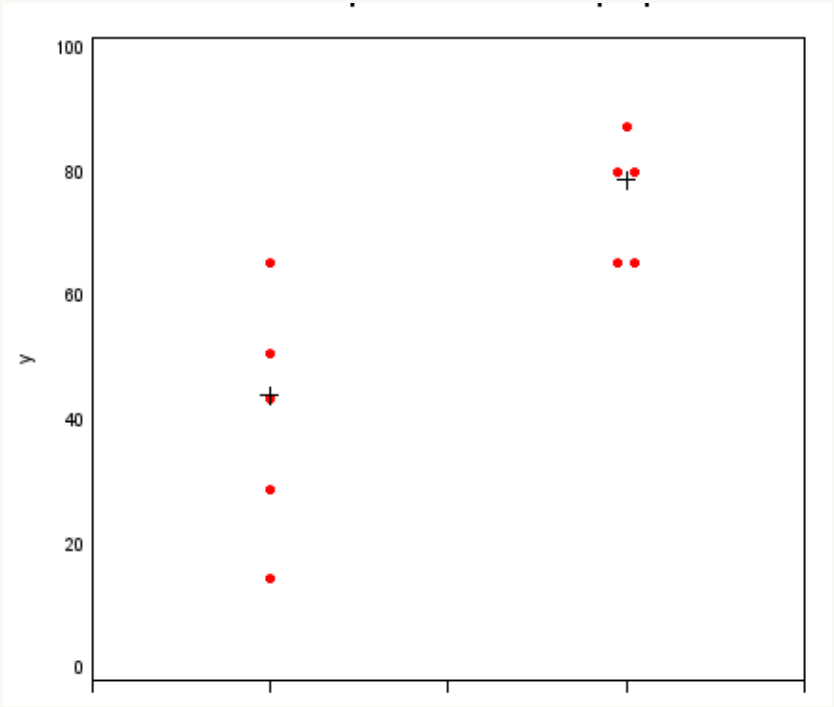
What is the least likely drawing? (Keeping in mind the original population)

Drawing A



n	mean 1	mean 2	d	sd	es	p value
5	57.4	60.2	2.9	14.6	0.20	0.764

Drawing B



n	mean 1	mean 2	d	sd	es	p value
5	44.4	79.0	34.7	14.7	2.36	0.006

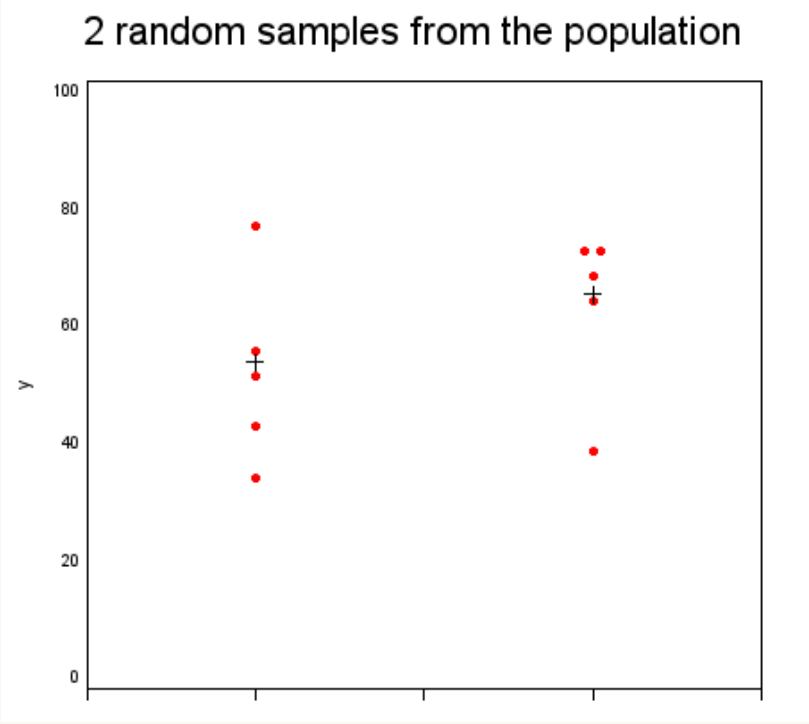
P Values examples

What is the least likely drawing? (Keeping in mind the original population)

Drawing A

n	mean 1	mean 2	d	sd	es	p value
25	54.6	66.1	11.4	14.3	0.80	0.007

Drawing B

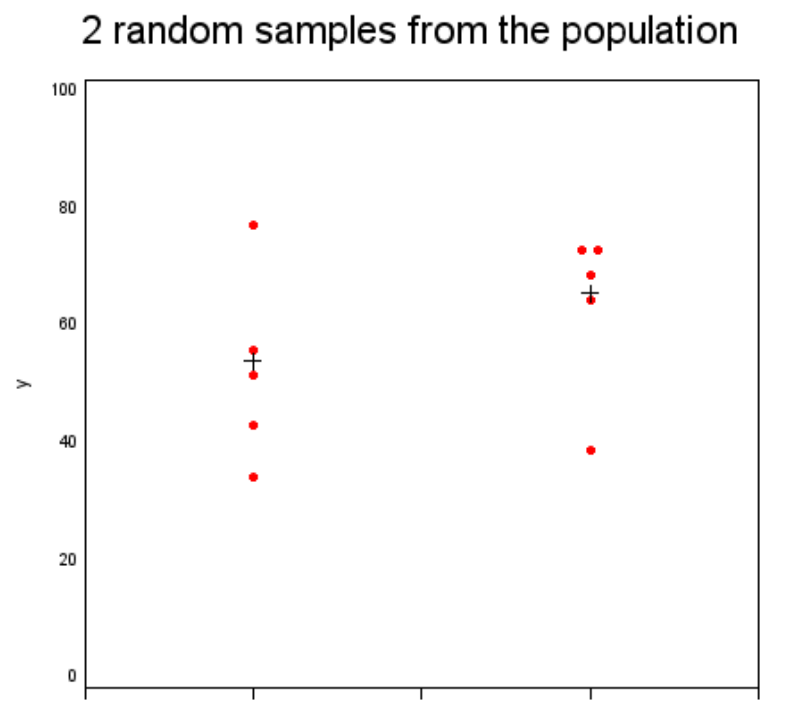


n	mean 1	mean 2	d	sd	es	p value
5	54.1	65.7	11.6	15.2	0.76	0.264

P Values examples

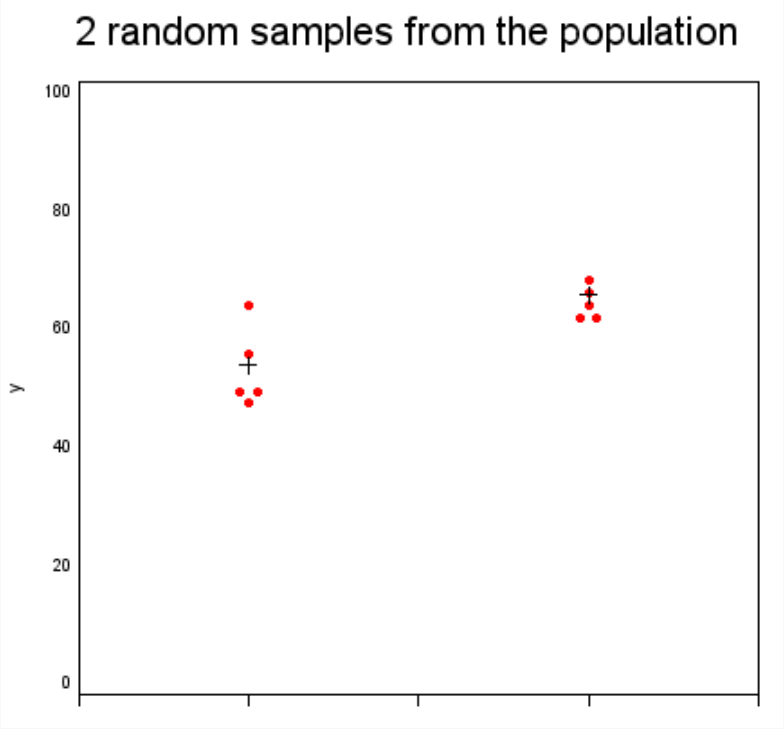
What is the least likely drawing? (Keeping in mind the original population)

Drawing A



n	mean 1	mean 2	d	sd	es	p value
5	54.1	65.7	11.6	15.2	0.76	0.264

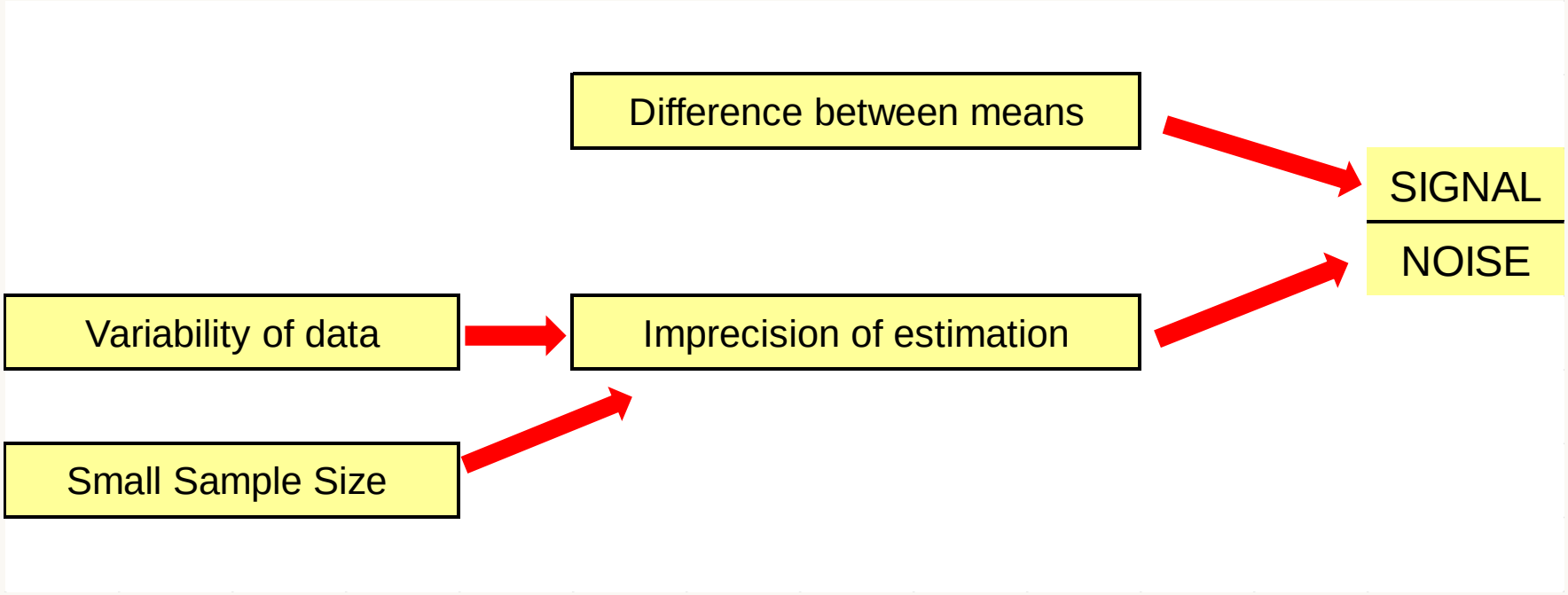
Drawing B



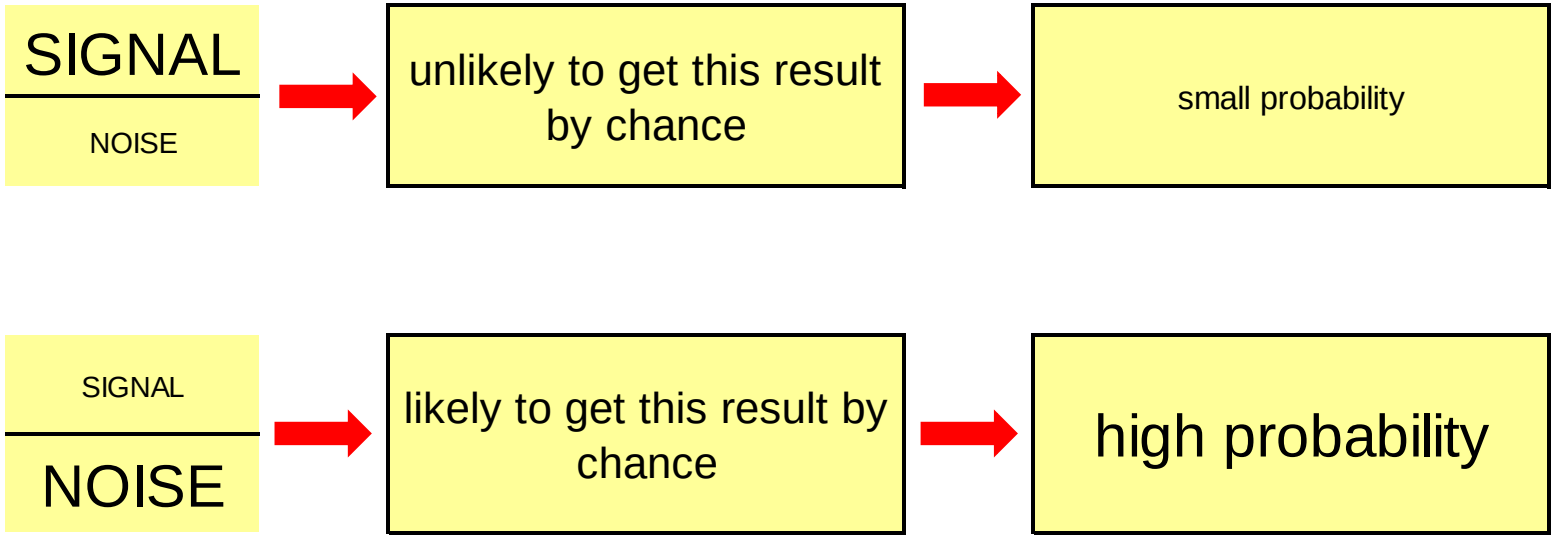
n	mean 1	mean 2	d	sd	es	p value
5	54.1	65.7	11.6	4.8	2.41	0.005

Conclusion : the logic of the P Value

P value is a function of the ratio $\frac{\text{SIGNAL}}{\text{NOISE}}$



Relationship signal/noise ➡ P Value



Avoid using avoid using the P Value in a reductive way :

P>0.05
Not Significant
No effect

P<0.05
Significant
Important
effect

Do not dichotomize the conclusions !!!!

- http://www.askanalytics.in/2015/10/understanding-p-value.html

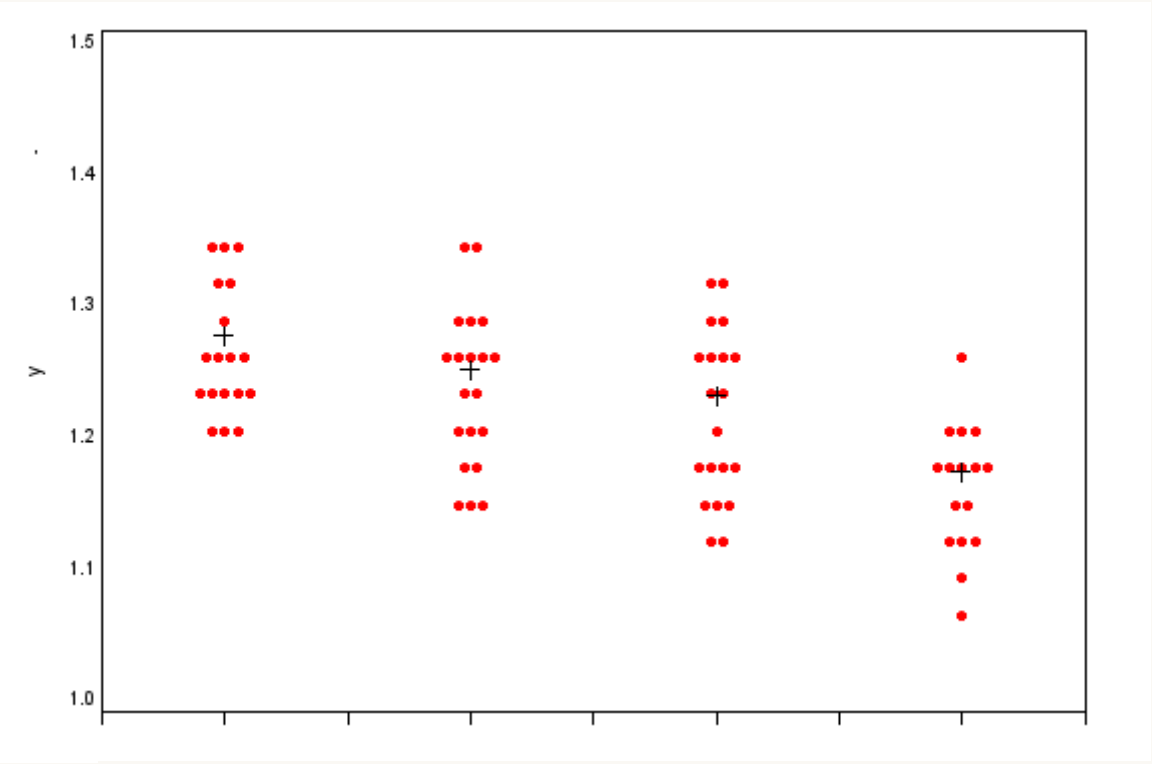
How to use the P Value?

- The P value gives an idea of the chance of obtaining such a result by hazard if the null hypothesis is true (= drawing from 2 identical populations)

- The P Value gives no idea of the physiological relevance

P value interpretation

Brain weight (g) of rats daily exposed to x during development



Control Dose1 Dose 2 Dose3

Comparisons with control group (Anova & Dunnett)

P values: 0.38 0.046 <0.0001

x	mean	sd	diff to ctrl
0	1.28	0.05	
d1	1.25	0.06	-0.02
d2	1.23	0.06	-0.04
d3	1.17	0.05	-0.10

Anova linear contrast	p<0.0001
Trend test	p<0.0001

P value interpretation

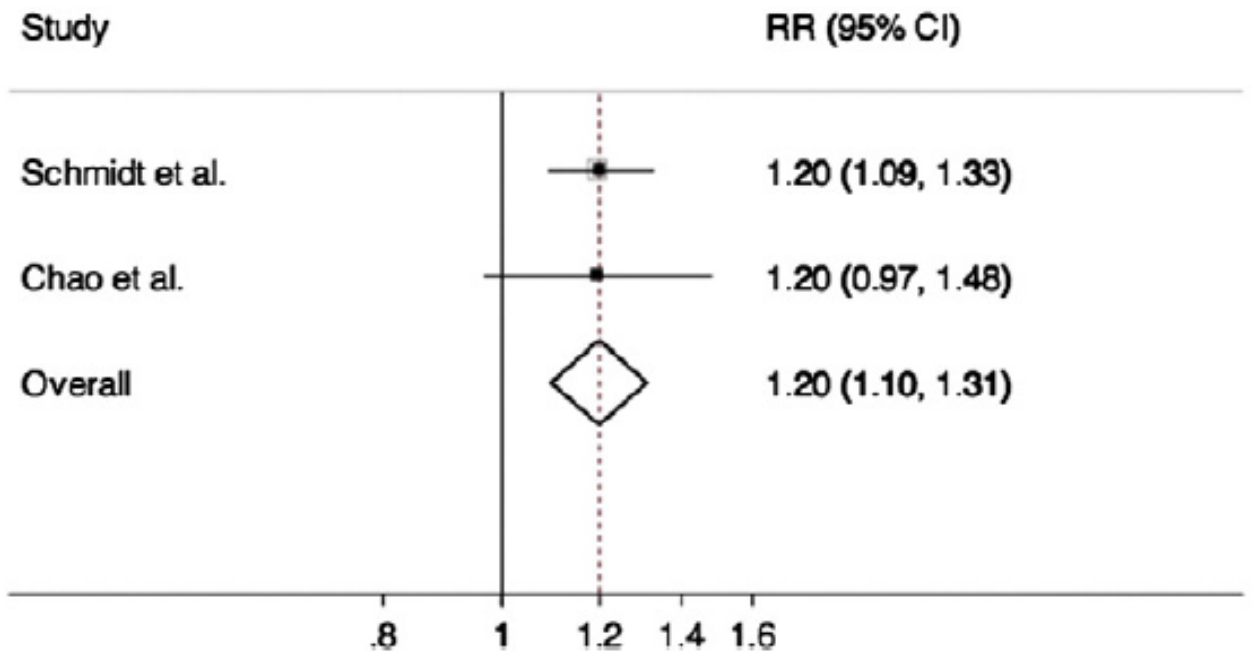


Fig. 2. Meta-analysis of the association between use of COX-2 inhibitors and atrial fibrillation as reported by Schmidt et al. and Chao et al.

- Morten Schmidt a,b,*, Kenneth J. Rothman c,d

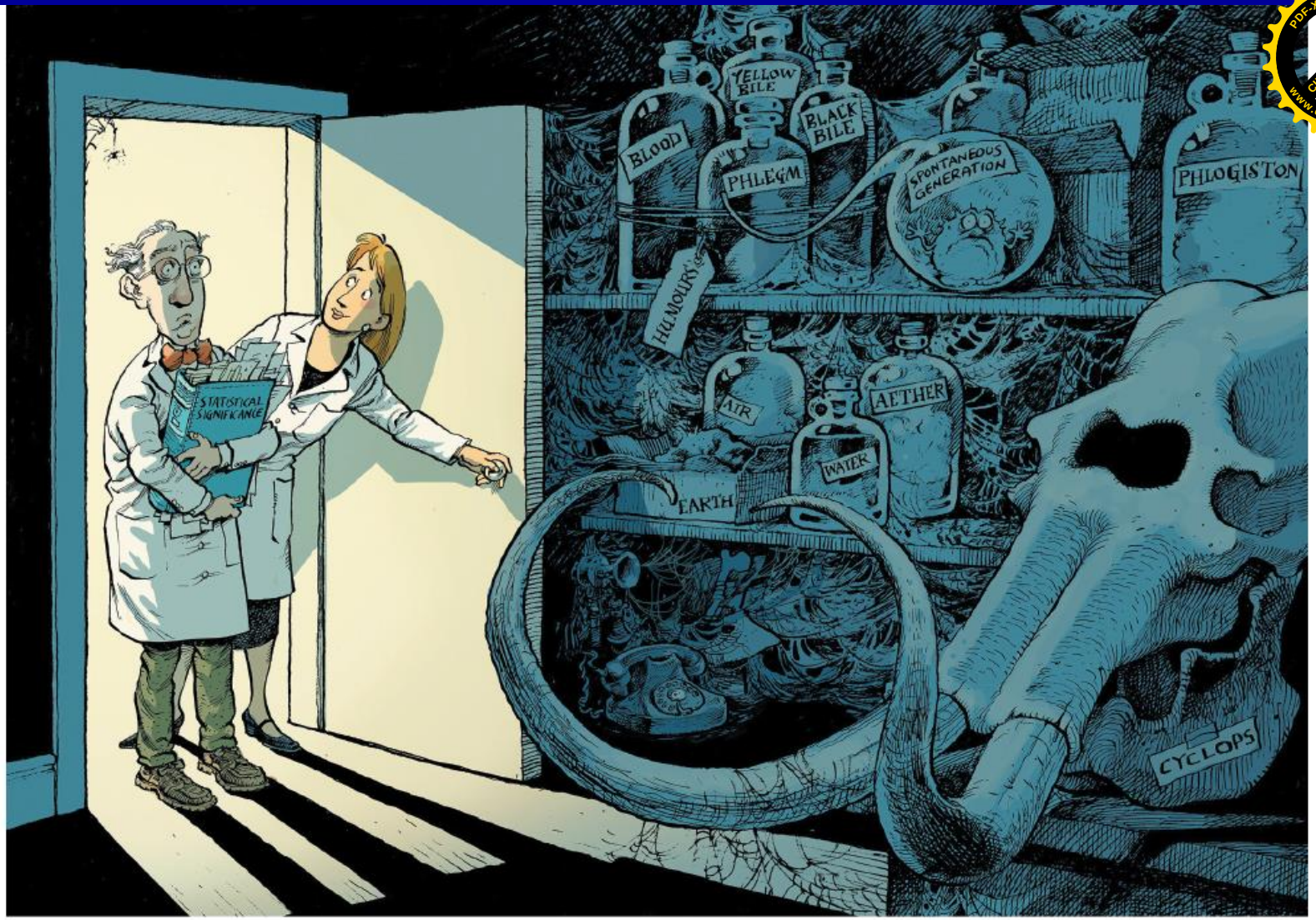
AJG "The Negative Issue" November 2016

"The Negative Issue" of The American Journal of Gastroenterology

The American Journal of Gastroenterology Presents "The Negative Issue"

Twenty-five Negative Studies that Remind Readers that "Negative is Positive"

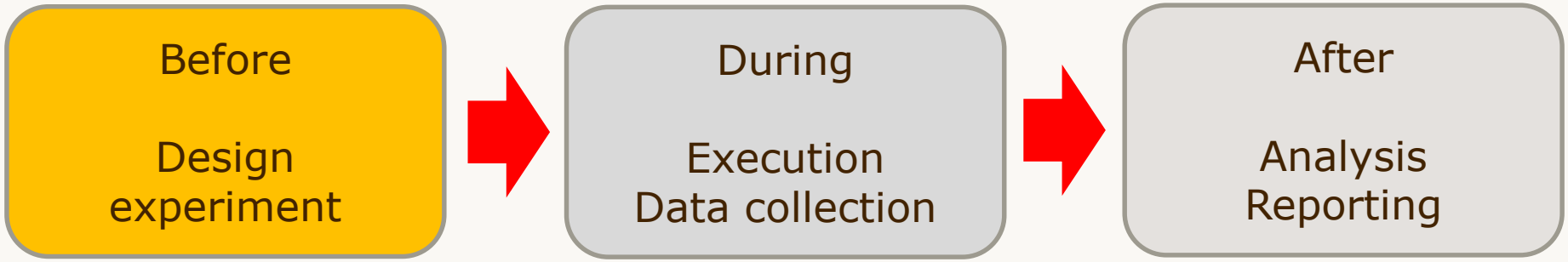
"We want to know what medicines don't work, what diets miss the mark, what risk factors are irrelevant, what supplements are no better than placebo, what diagnostic tests are unrevealing, unhelpful, or even harmful, and anything else that may be terrifically non-contributory in gastroenterology and liver diseases," the Co-Editors asked in the request for manuscripts.



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Data quality improvement : Scientific Workflow, where to act ?



Improve the design : 6 approaches

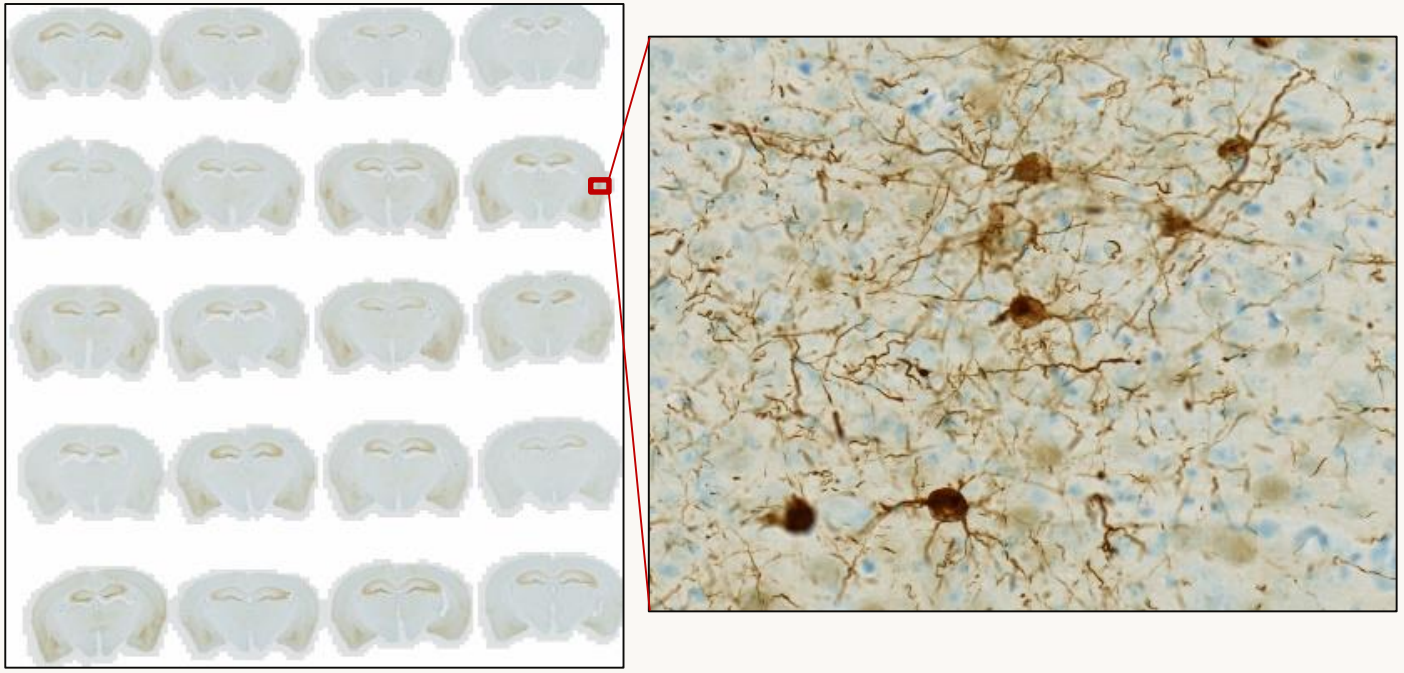
- 1 Avoid bias : randomize, blind ...
- 2 Calculate sample size
- 3 Favor paired design
- 4 Reduce variability of measurements
- 5 Use historical data
- 6 Split experiment in mini sub-experiments (Randomize block design)

1 Avoid bias

- Randomize
- Blind observations
- Process automation (Reduce technical variability)
- Scrutinize all the technical details

Avoid bias

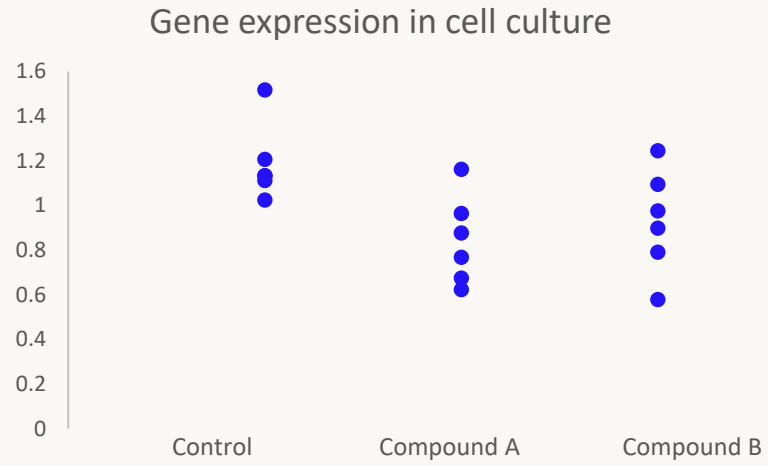
Example : Immunohistochemistry staining, Simultaneous Processing of slices from 20 animals



Avoid bias

Example : Gene expression measured in 18 cell cultures

Gene expression in cell cultures (qPCR)					
Treatments	sample	date culture	gene x	Mean	sd
Control	1	2017-12	1.11		
	2	2017-12	1.21		
	3	2017-12	1.13		
	4	2017-12	1.52		
	5	2017-12	1.13		
	6	2017-12	1.02	1.19	0.17
Compound A	7	2018-02	1.16		
	8	2018-02	0.77		
	9	2018-02	0.62		
	10	2018-02	0.96		
	11	2018-02	0.67		
	12	2018-02	0.88	0.84	0.20
Compound B	13	2018-02	1.25		
	14	2018-02	0.98		
	15	2018-02	0.90		
	16	2018-02	1.10		
	17	2018-02	0.79		
	18	2018-02	0.58	0.93	0.23



Anova P value : 0.027

2 Sample size calculation

Questions:

How large should my sample be?

Answer:

It depends...

...large enough to be an accurate representation of the population

...large enough to achieve statistically significant results

Educational Research Basics by Del Siegle - University…

Why Calculate the Sample Size?

- Scientific reasons
  - Level of expected confidence
  - Size of the effect to be detected:
    Biologically significant ≠ Statistically significant

- Ethical reasons
  - Avoid too many experiences (save animals, patients)
  - Avoid unuseful experiences if not enough data !!!!!!

- Financial reasons

Why Calculate the Sample Size?
Effect of sample size on statistical significance

Example (simulation):

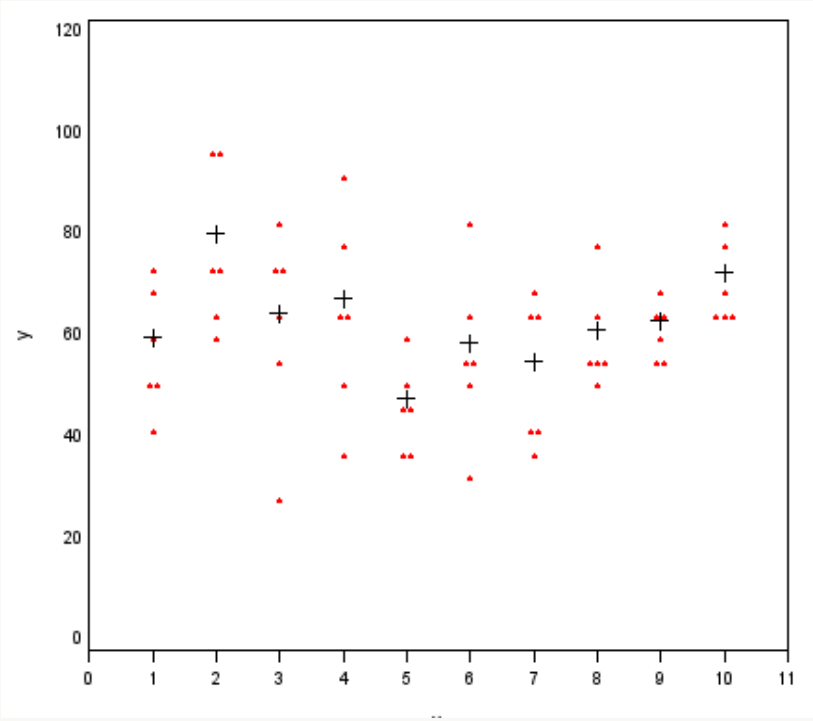
Comparison of blood pressures		
Trial 1	Treatment A	Treatment B
Mean	100	119
Sd	15	14
n	5	5
p	**0.08**	
Trial 2	Treatment A	Treatment B
Mean	100	119
Sd	15	14
n	20	20
p	**0.0001**	

Effect of sample size on variability of means (simulated data)

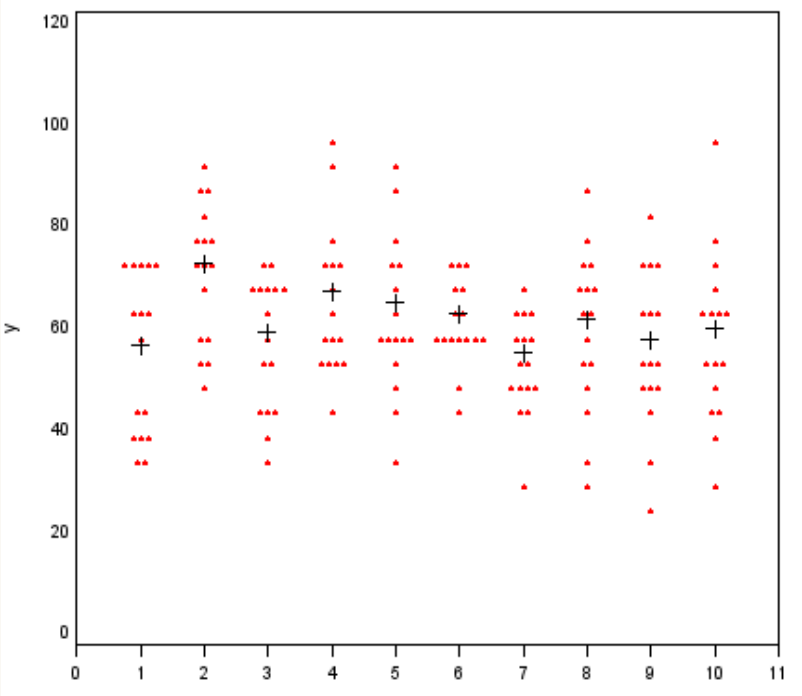
random samples drawn from a gaussian population (mean=60 sd=15)

10 random samples n=6

Means and individual data



10 random samples n=16



n=6

Minimal mean = 48

Maximal mean = 72

n=16

Minimal mean = 53

Maximal mean = 67

Sample Size Calculation: Comparison of 2 Means

$$n = 2\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \cdot \frac{s_d^2}{\Delta^2} \qquad n = 2\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \cdot \frac{CV^2}{\Delta^2}$$

- n : minimum number of observations
- $Z_{1-\alpha}$: critical value in Gaussian distribution corresponding to the probability threshold value α:
  - 0,05 one-sided test $Z_{1-\alpha}$ = 1,64
  - 0,05 two-sided test (0,025) $Z_{1-\alpha}$ = 1,96
- $Z_{1-\beta}$: critical value in Gaussian distribution corresponding to the selected power probablility.
  - Power 80% $Z_{1-\beta}$ = 0,84
  - Power 90% $Z_{1-\beta}$ = 1,28
- s_d : standard deviation and △ smallest difference to be detected
  - Sd and △ are expressed in absolute values or
  - CV and △ are expressed in percentages
  - Sd estimate : literature (comparison !), pilot study

Sample size according to effect size (comparison of 2 means)

- $Z_{1-\alpha}$: 0,05 two-sided test (0,025) $Z_{1-\alpha} = 1,96$
- $Z_{1-\beta}$: Power 80% $Z_{1-\beta} = 0,84$

$$n = 2\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \cdot \frac{s_d^2}{\Delta^2}$$

$$n = 15.7 \frac{1}{(ES)^2}$$

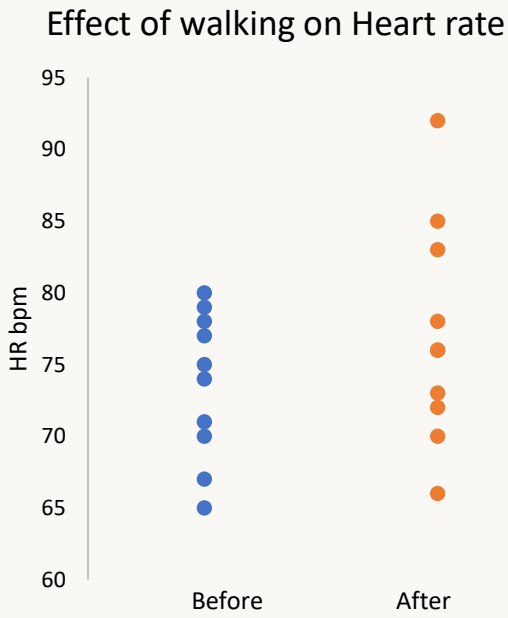
Effect size	0.5	0.75	1	1.25	1.5	2
n (per group)	63	28	16	11	7	4

3 Paired design : the benefit

- Example:
  - Measures performed on same subjects before and after treatment or subjects who receive 2 treatments sequentially
  - Measures performed on different matched subjects (similarity in genotype or in any relevant feature)

Immediate effect of walking on heart rate		
	Heart rate (bpm)	
Subject	before	after
1	65	66
2	67	70
3	70	73
4	71	72
5	74	76
6	75	78
7	77	76
8	78	85
9	79	83
10	80	92

mean	73.6	77.1
sd	5.2	7.7
var	26.7	59.9
Common sd	6.6	
Diff of means	2.6	
Effect size	0.39	

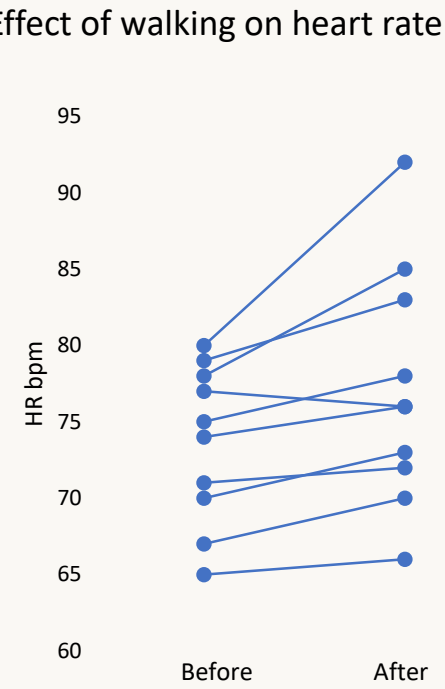


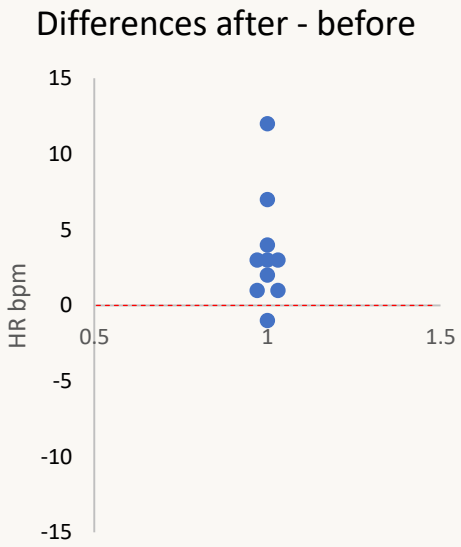
(Forgetting that the data are paired)
Non Paired Student t test P Value: 0.24

Benefit of paired design

before	after	difference
65	66	1
67	70	3
70	73	3
71	72	1
74	76	2
75	78	3
77	76	-1
78	85	7
79	83	4
80	92	12

mean	3.5
sd	3.7
Effect size	0.96





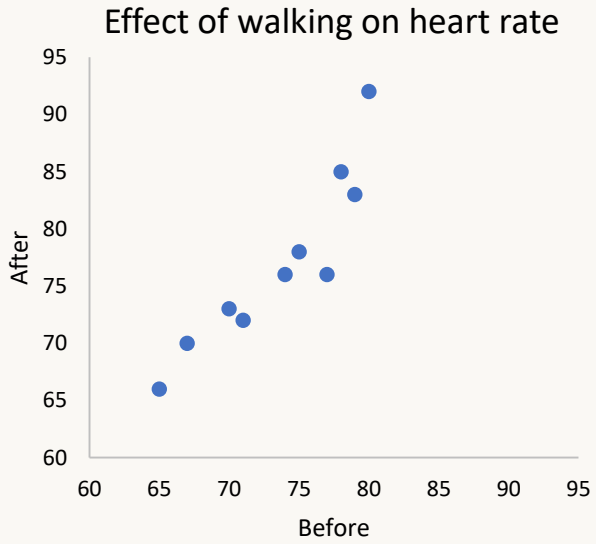
Paired Student t test P Value: 0.014

For comparison : Non Paired Student t test P Value: 0.24

Benefit of paired design

before	after	difference
65	66	1
67	70	3
70	73	3
71	72	1
74	76	2
75	78	3
77	76	-1
78	85	7
79	83	4
80	92	12

mean	3.5
sd	3.7
Effect size	0.96



Correlation r = 0.92

Benefit of Paired Student is linked to correlation

Benefit of paired design : sample size

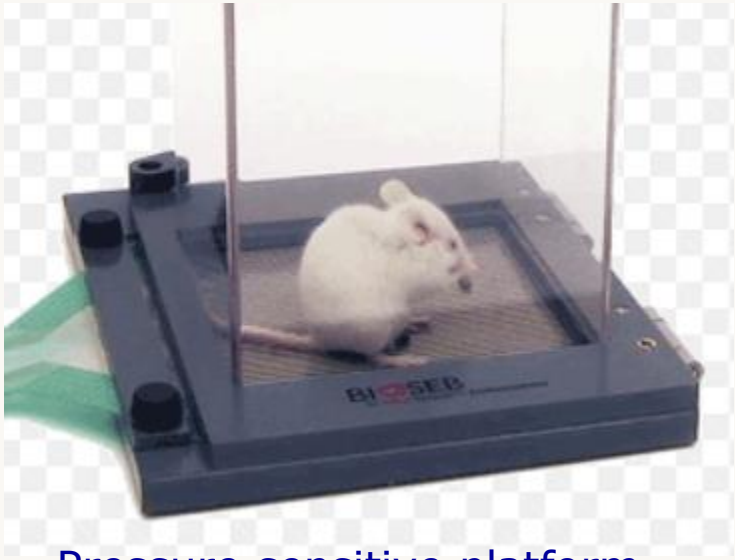
$$n = (z_{1-\alpha} + z_{1-\beta})^2 \frac{S^2}{\Delta^2}$$

$$n = 7.8 \frac{1}{(ES)^2}$$

Effect size	0.5	0.75	1	1.25	1.5	2
n	32	14	8	6	4	2

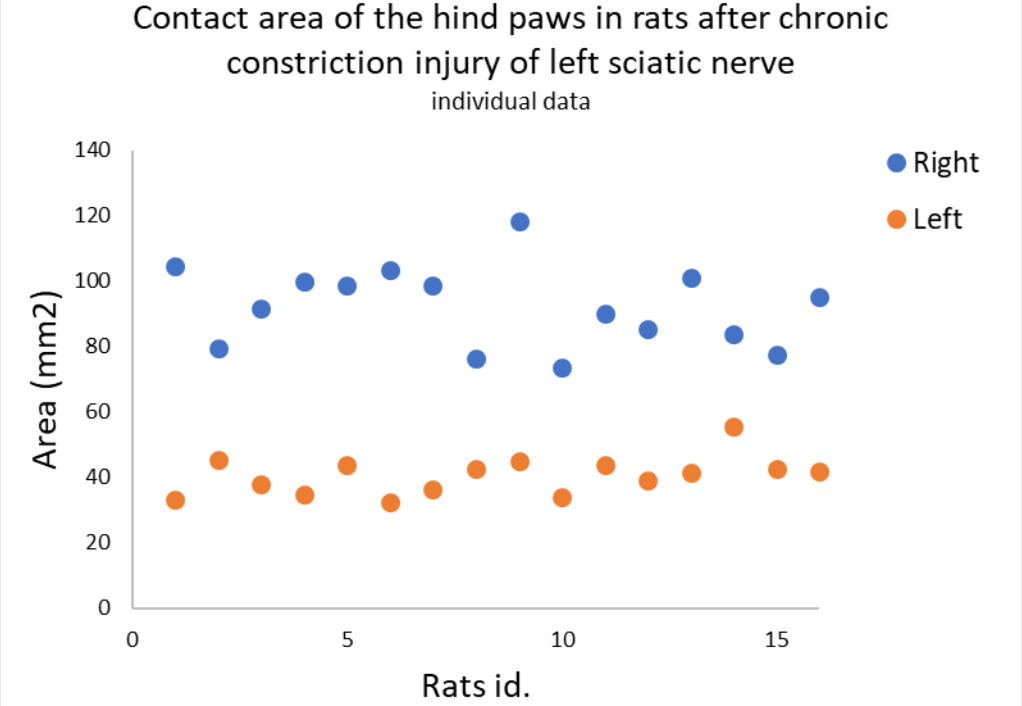
4 Reduce variability of measurements

Model of pain induced in the rat by compression of the sciatic nerve

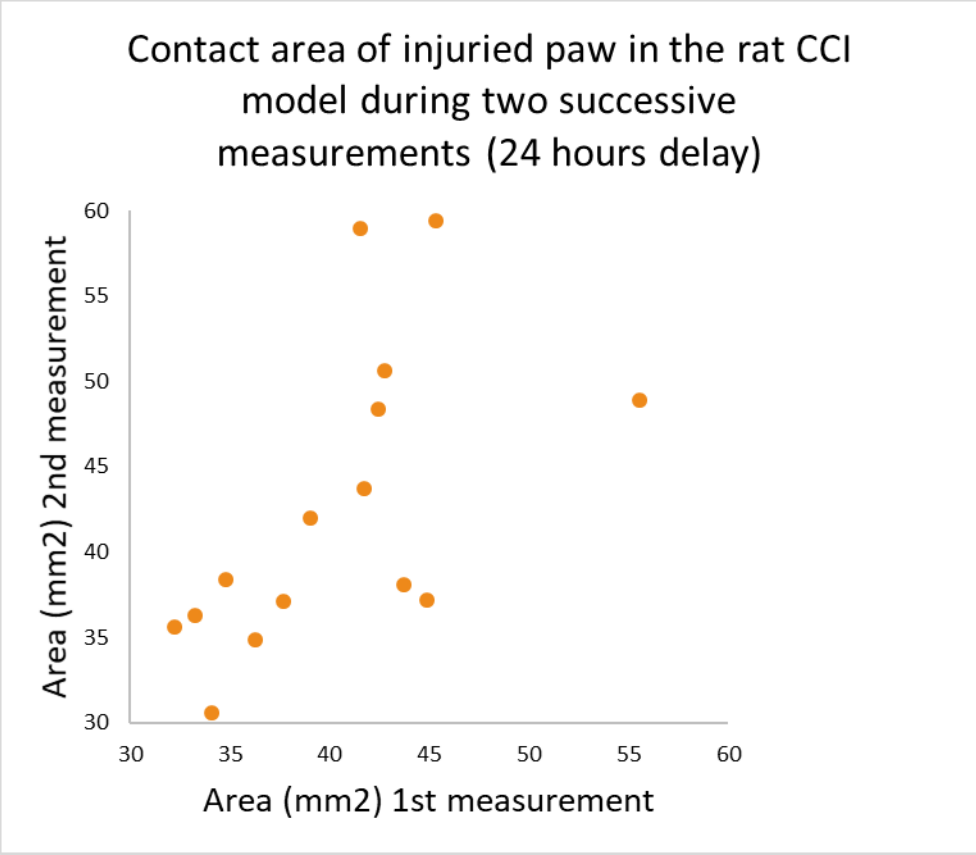


Pressure sensitive platform

Analysis of the model: induced asymmetry

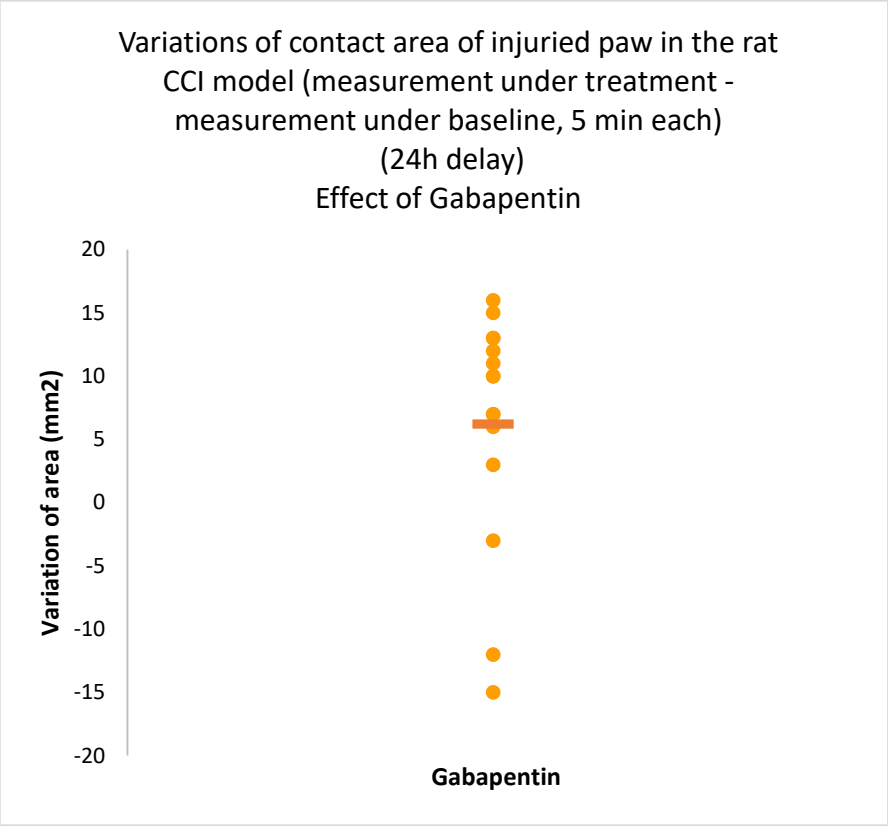


Statistical analysis of the model : correlation between successive measures



Correlation r = 0.59

Analysis of the model : therapeutic window



Mean	6.2
sd	9.4
n	15
Paired t	1.601
Probability	0.132
Effect size	0.66

Conclusion : High level of variability due to poor correlation between successive measurements, great variability of differences (before – afeter)

Sample size calculation based on current data

Mean	6.2
sd	9.4
n	15
Paired t	1.601
Probability	0.132
Effect size	0.66

$$n = 7.8\frac{1}{(ES)^2}$$

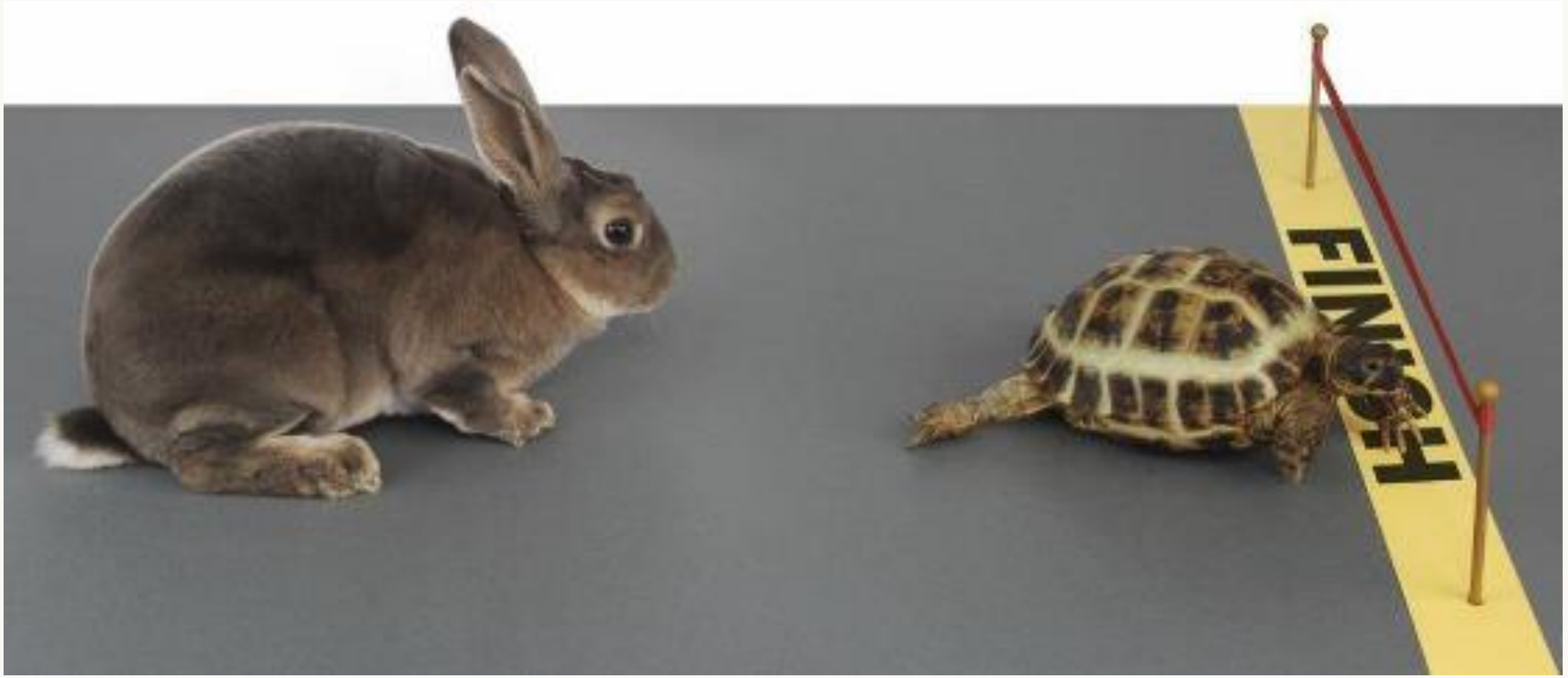
Effect size	0.66
n	**19**

Improvement of the model by reducing variability

Contact area of injuried paw (mm2) 5 min measurement		
meas. before	meas. after	diff
59	44	-15
34	49	15
35	38	3
39	55	16
43	50	7
58	46	-12
41	38	-3
28	41	13
38	44	6
30	40	10
45	55	10
42	54	12
31	42	11
37	44	7
46	59	13
n		**15**
Mean		**6.2**
Sd		**9.4**
Var		**90.7**
ES		**0.66**
Minimum sample size		**19**

Contact area of injuried paw (mm2) 4 X 5 min measurement										
m1	m2	m3	m4	mean before	m5	m6	m7	m8	mean after	diff
59	51	48	39	49.3	44	45	50	49	47.0	-2.3
34	34	39	35	35.5	49	48	49	41	46.8	11.3
35	41	43	46	41.3	38	48	49	40	43.8	2.5
39	38	52	45	43.5	55	62	51	47	53.8	10.3
43	48	38	46	43.8	50	52	52	53	51.8	8.0
58	51	45	58	53.0	46	58	60	50	53.5	0.5
41	43	47	34	41.3	38	44	37	40	39.8	-1.5
28	39	45	38	37.5	41	36	47	39	40.8	3.3
38	38	43	42	40.3	44	45	51	47	46.8	6.5
30	38	35	39	35.5	40	44	40	32	39.0	3.5
45	42	43	47	44.3	55	50	42	50	49.3	5.0
42	41	42	44	42.3	54	57	41	52	51.0	8.8
31	31	44	37	35.8	42	39	38	43	40.5	4.8
37	34	35	26	33.0	44	30	26	30	32.5	-0.5
46	49	41	47	45.8	59	52	54	60	56.3	10.5
n										**15**
Mean										**4.7**
Sd										**4.5**
Var										**18.6**
ES										**1.06**
Minimum sample size										**7**

Speed illusion : the bunny and the turtle

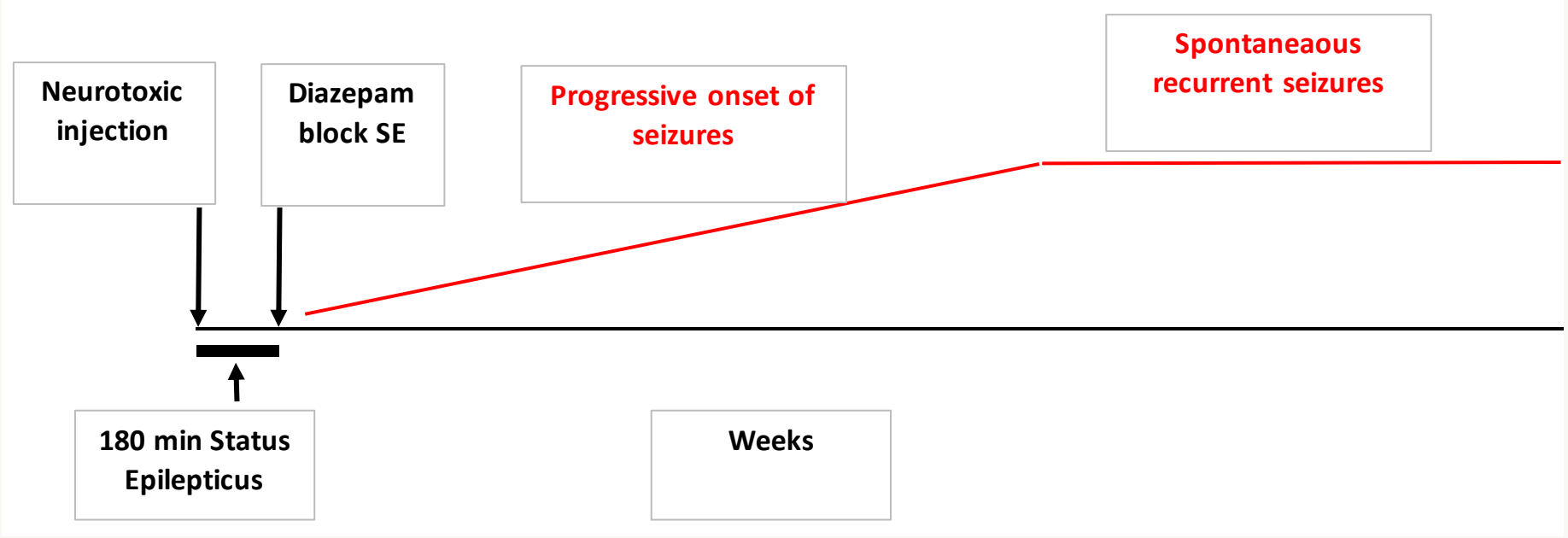


Keep in mind the Harm / Benefit analysis

Animal experimentation	Scientific Information
Discomfort	Relevance of the model
Suffering	Quality of information (robustness , reproducibility)
	Chance of success

5 Use of Historical Data to Improve Power of Design

Chronic epilepsy model



Chronic epilepsy model : kinetic of seizures

Epilepsy Model

Neurotoxic induced seizures in the rat (neurotoxic single injection)

Number of seizures per day

	Baseline														Treatment													
	Days																											
Rat id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	4	4	4				2	3	3	2	2		1	3	1	3				1	2	1		1	1	2	1	
2		1	4	4	1																					1	2	15
3	3			1		3	2						2	5	1	4	9	1	1						3	3	2	1
4	1	1		1					1		3	2		1	2	2	4	6	2				3	1	3	1		
5	3									1	1	2		1	1			1	1		1	3			2	1		4

Nb of seizures	
Controls	
Difference (Treatment -Baseline)	
n	90
Mean	-0.6
Sd	12.1

Sample size calculation based on current data : comparison control versus treated

Comparison of 2 groups (control versus treated) T test comparison of 2 means	
proba 1- alpha/2	0.025
proba beta	0.8
Z alpha	1.96
Z beta	0.84
(za+zb)2	7.85
Delta	10
Sd	12
Effect size	**0.83**
Effect size ^{2}	0.69
n calc	22.6
n	**23**

Sample size calculation based on historical data comparison treated versus historical controls

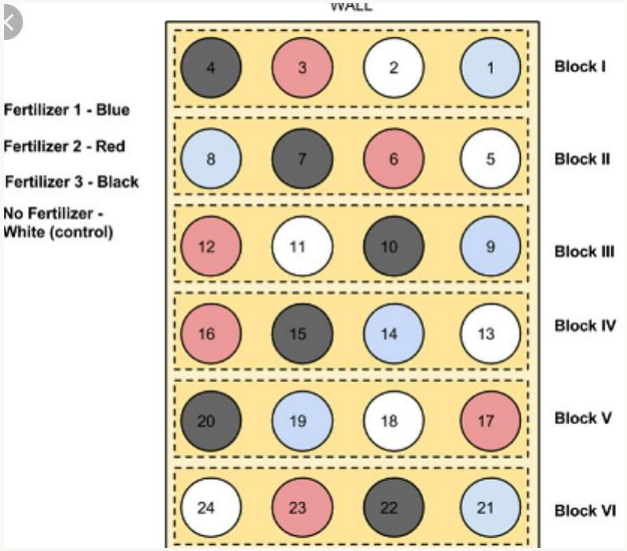
Comparaison de 2 groupes expérimentaux (contrôle versus traité) T test comparaison de 2 moyennes	
proba 1- alpha/2	0.025
proba beta	0.8
Z alpha	1.96
Z beta	0.84
(za+zb)2	7.85
Delta	10
Sd	12
Effect size	0.83
Effect size ^{2}	0.69
$2N = \frac{4(Z_\alpha + Z_\beta)^2 \sigma^2}{\Delta^2}$	
n calc	22.6
n	23

Comparaison aux contrôles historiques (ex n=90) T test comparaison à une référence	
proba 1- alpha/2	0.025
proba beta	0.8
Z alpha	1.96
Z beta	0.84
(za+zb)2	7.85
Delta	10
Sd	12
Effect size	0.83
Effect size ^{2}	0.69
$N_d = \frac{(Z_\alpha + Z_\beta)^2 {\sigma_d}^2}{{\delta_d}^2}$	
n calc	11.4
n	12

2N = sample size for the 2 groups

6 Split in sub-experiments : Randomised Block Design

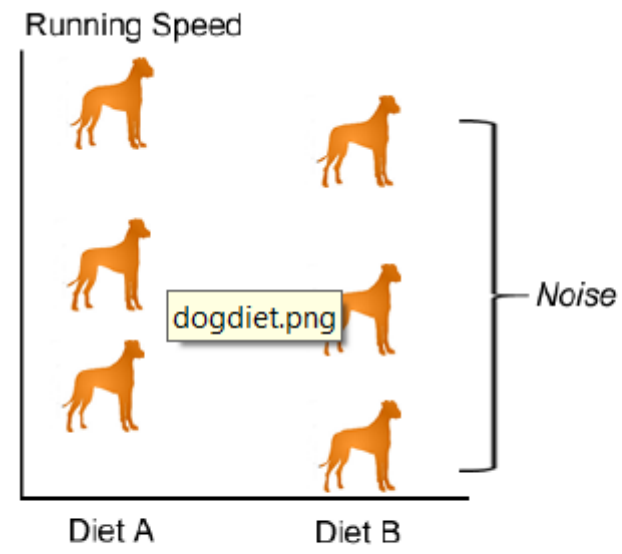
- **In an RB Design the experiment is split up into a number of small experiments**
- **Advantages of the RB Design:**
  - **Increase homogeneity of tested "material" (matched)**
  - **If the blocks are independent, smooth out uncontrolled variations and increase representativity and robustness**



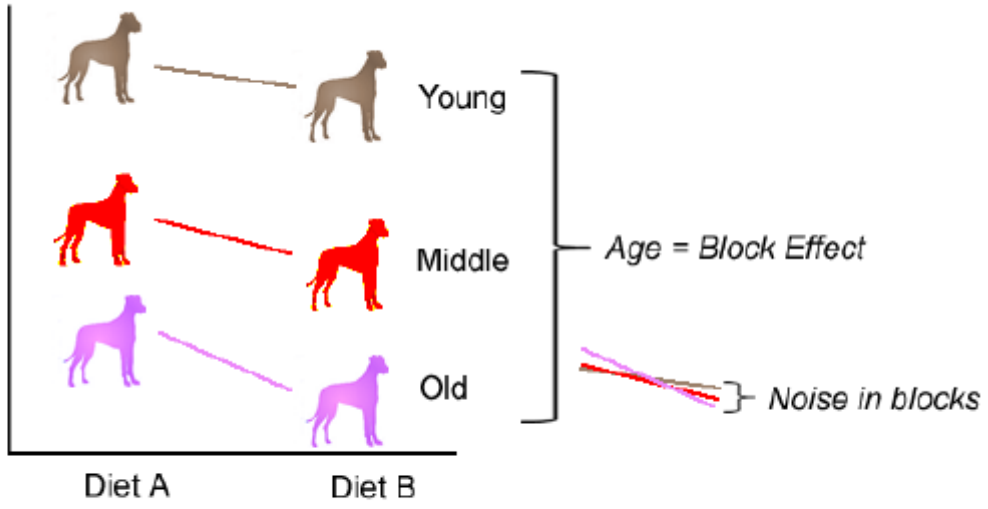
Design coming from the agronomy practice

Randomised Block Design Increases Homogeneity

Compare with the paired test



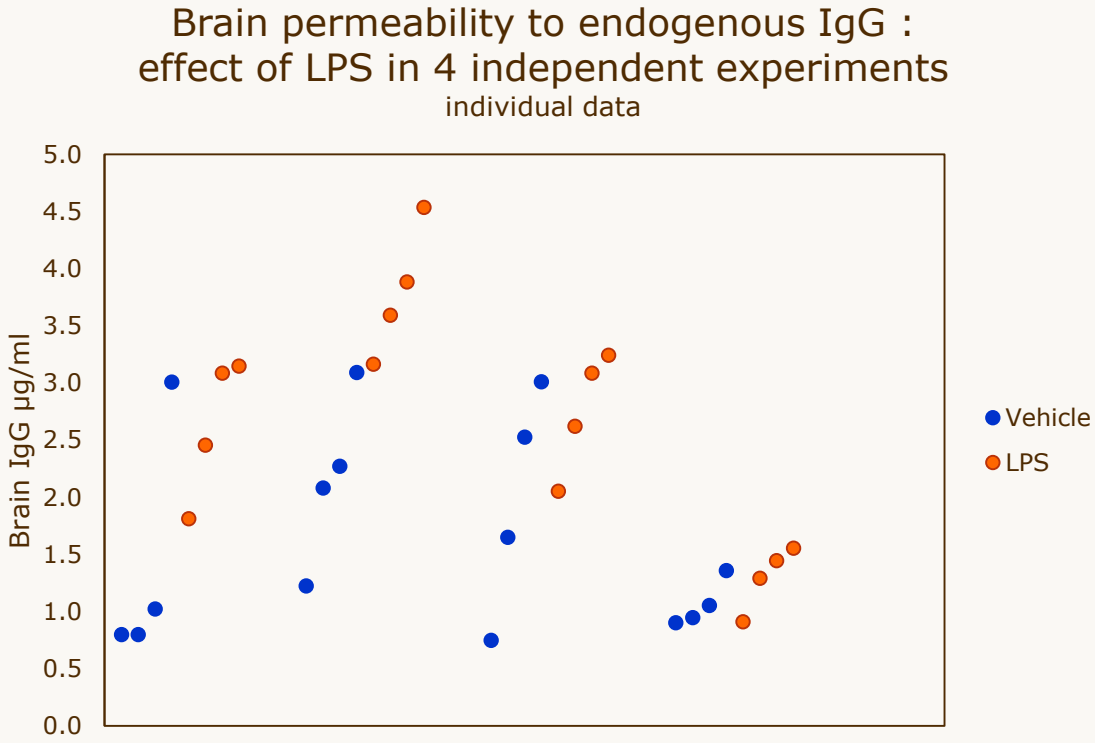
Total Variability = Treatment Effect + Noise

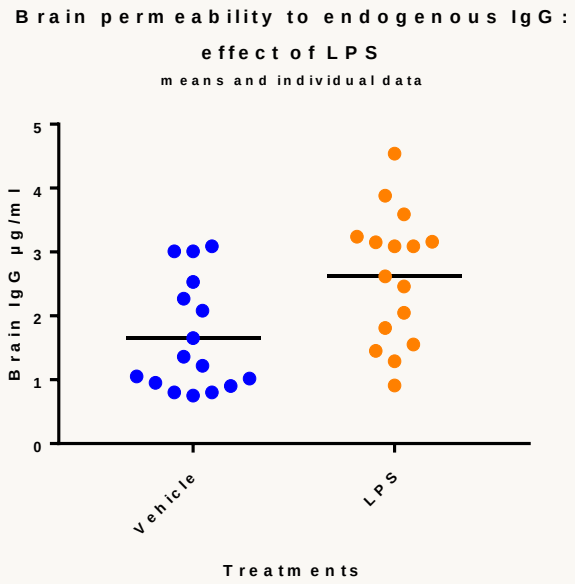


Total Variability = Treatment Effect + Block Effect + Within block Noise

Randomised Block Design Increases robustness

Example of uncontrolled variations (animals, operator, environment, reagents....)

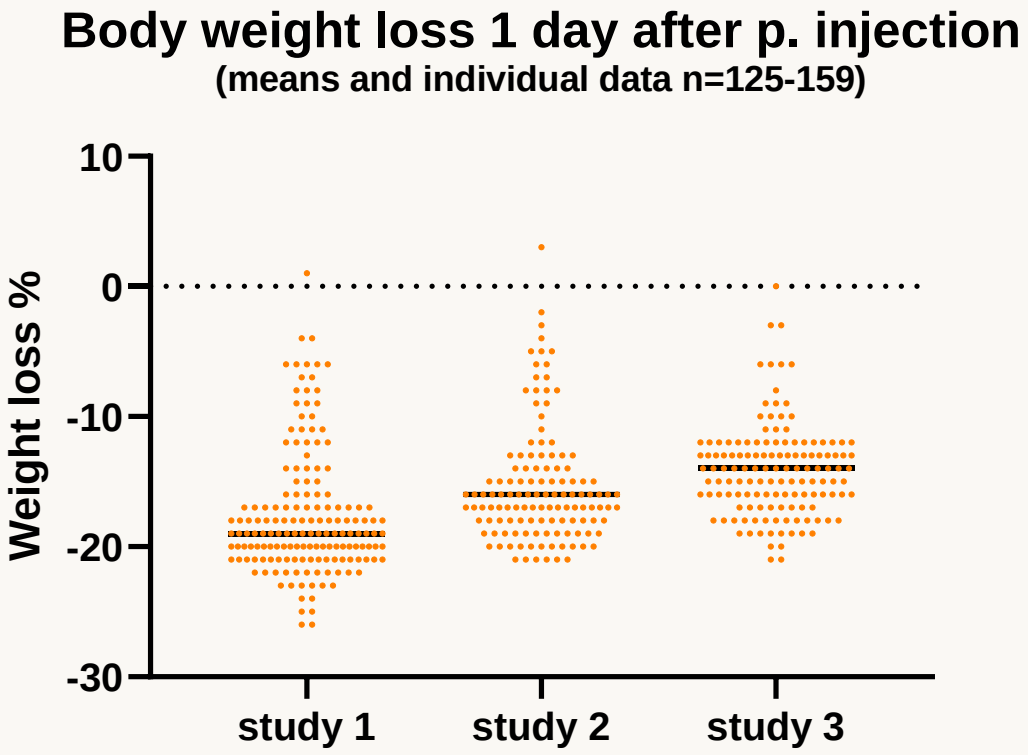




Same data

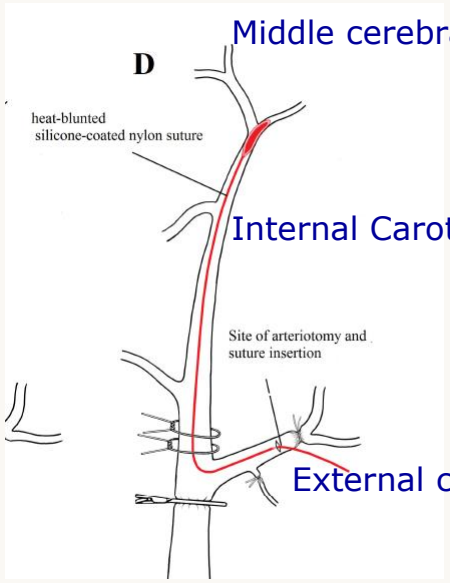
Randomised Block Design in order to avoid :

Example of uncontrolled drift in 3 independent studies conducted at 3-month intervals



Randomised block design improves Stability

Middle Cerebral Artery Occlusion Model of Stroke in Rodents



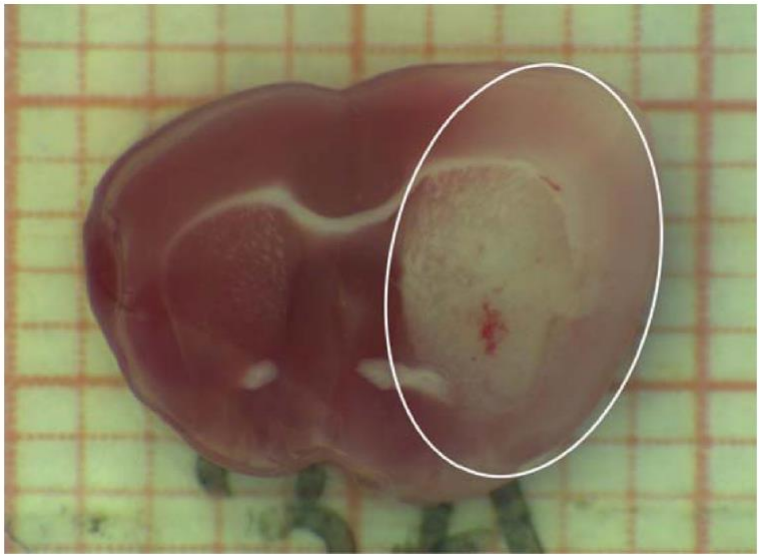


Figure 5. Representative image of TTC staining resulting in infarcted tissue unstained white (white circle) while viable tissue is strongly stained (brick red).

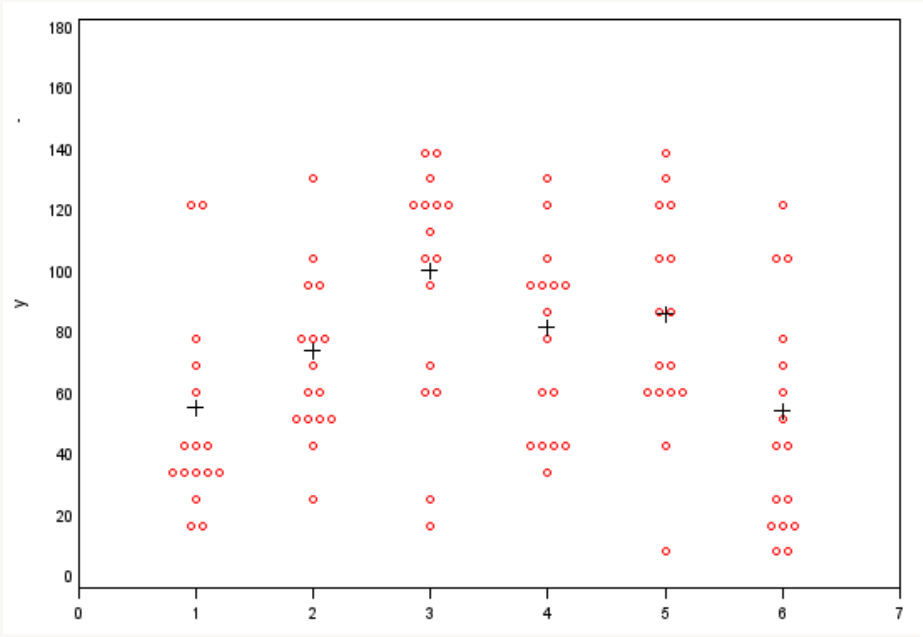
Middle Cerebral Artery Occlusion
Infarct volume in control rats (n=16 per group)

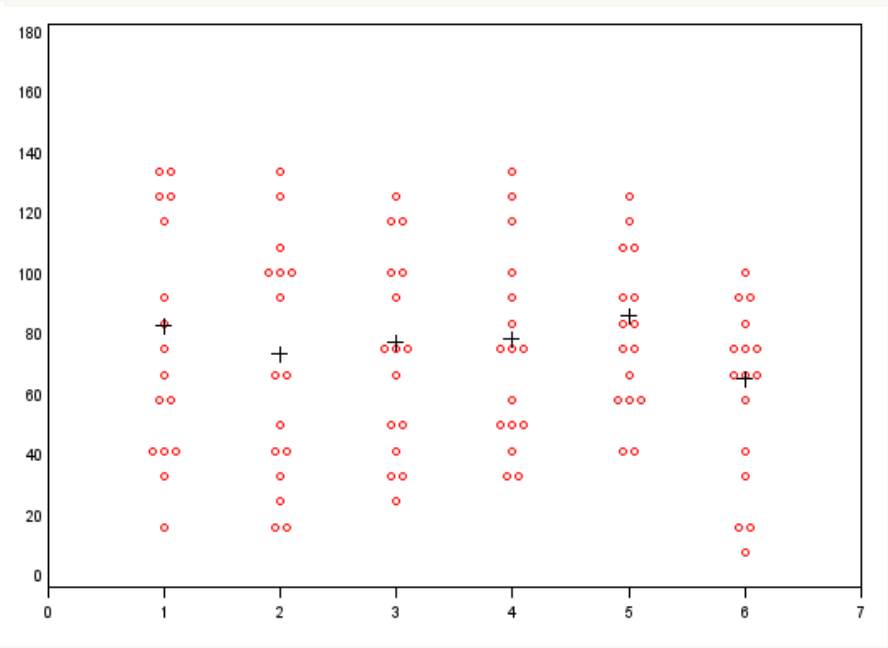
6 experiments performed

during 6 consecutive weeks
(one experiment per week)

6 experiments formed of 4 sub-experiments performed during 4 differents weeks

Randomized bd





Observed data, not simulated!

Conclusions :

Simple approaches to improve reproducibility of studies :

- Tansparency in the reporting of the results : graphics and expression of effects (effect size)

- Improvement of design:
  - Elimination of bias
  - Adequate Sample Size
  - Paired Design
  - Reduction of the Measuring Variability
  - Historical database
  - Randomized block design

PDF-XChange Viewer
Click to buy NOW!
www.docu-track.com

Thank You

Etienne Hanon

Stat Solutions

Etienne.hanon@gmail.com