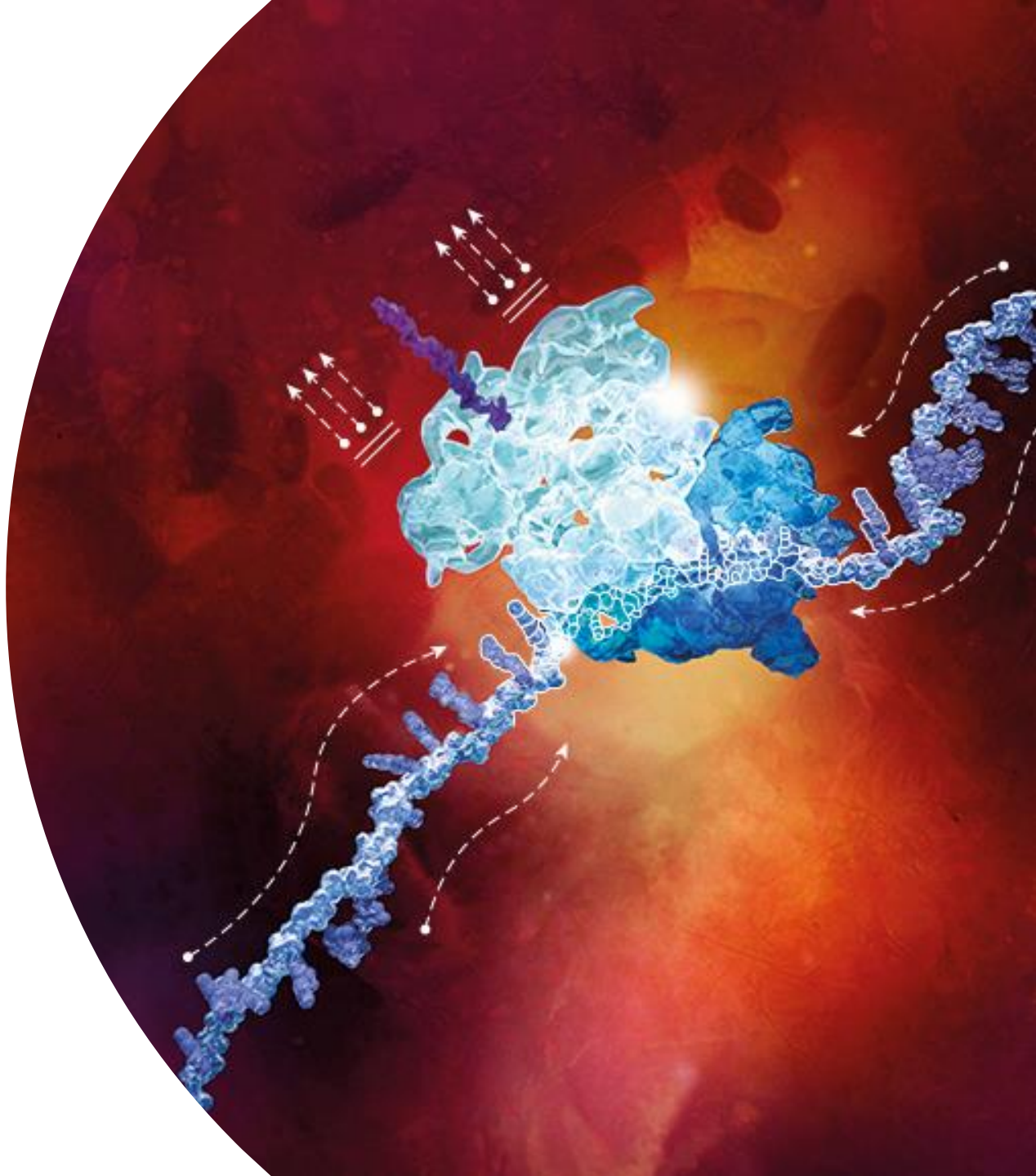
Diversity matters – the risks of standardisation

Natasha Karp

Quantitative Biology, Discovery Science, R&D, AstraZeneca, UK



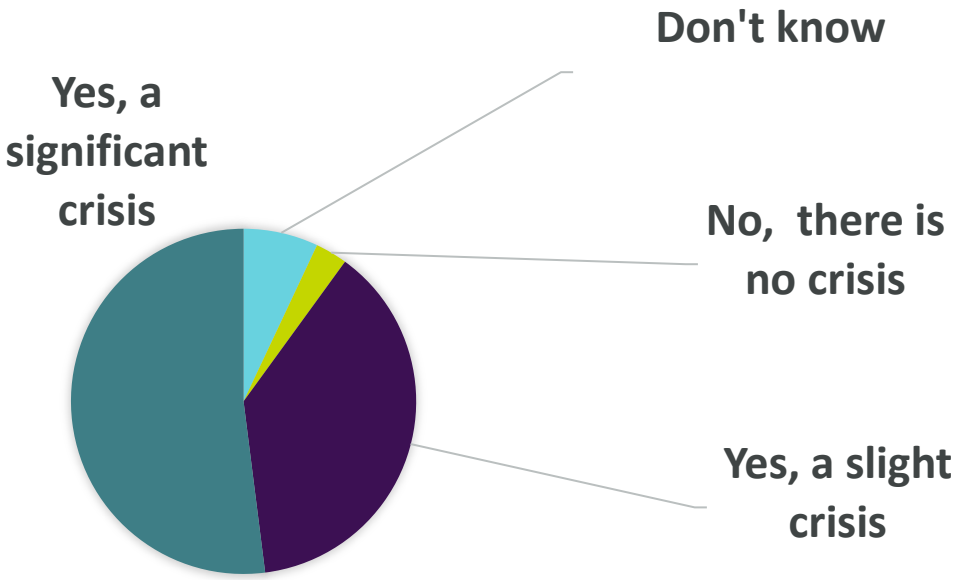
Agenda

1. Reproducibility
2. Standardisation
3. Phenotypic plasticity
4. Sex as a biological variable
5. Batch impact on reproducibility
6. Introducing noise
7. Conclusions

1

Reproducibility issue

Is there a reproducibility crisis?



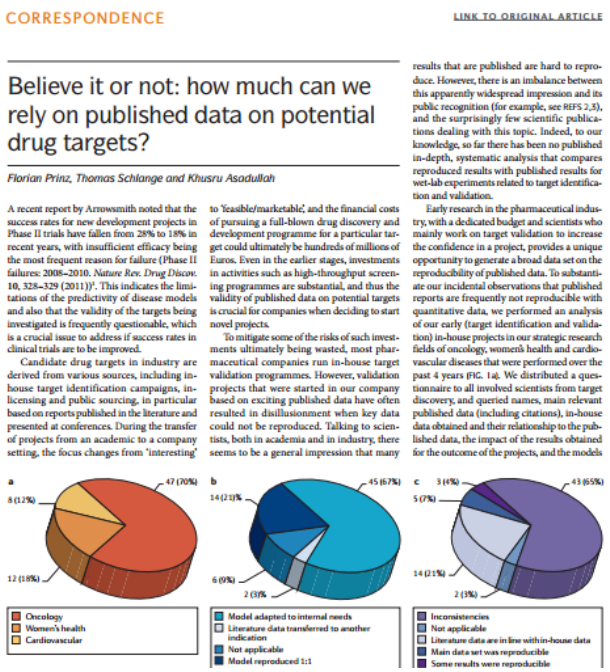
RESEARCHERS SURVEYED (N=1,576)

- In field of biology, 76% have failed to replicate another scientists experiments
- In field of biology, 60% failed to replicate their own experiments.

Baker, Monya. "1,500 scientists lift the lid on reproducibility." (Nature 2016).

Pre-clinical reproducibility issues

Bayer



CORRESPONDENCE

LINK TO ORIGINAL ARTICLE

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

Florian Prinz, Thomas Schlange and Khusru Asadullah

A recent report by Arrowsmith noted that the success rates for new development projects in Phase II trials have fallen from 28% to 18% in recent years, with insufficient efficacy being the most frequent reason for failure (Phase II failures: 2008–2010. *Nature Rev. Drug Discov.* 10, 328–329 (2011))[1]. This indicates the limitations of the predictivity of disease models and also that the validity of the targets being investigated is frequently questionable, which is a crucial issue to address if success rates in clinical trials are to be improved.

Candidate drug targets in industry are derived from various sources, including in-house target identification campaigns, in-licensing and public sourcing, in particular based on reports published in the literature and presented at conferences. During the transfer of projects from an academic to a company setting, the focus changes from 'interesting' to 'feasible/marketable,' and the financial costs of pursuing a full-blown drug discovery and development programme for a particular target could ultimately be hundreds of millions of Euros. Even in the earlier stages, investments in activities such as high-throughput screening programmes are substantial, and thus the validity of published data on potential targets is crucial for companies when deciding to start novel projects.

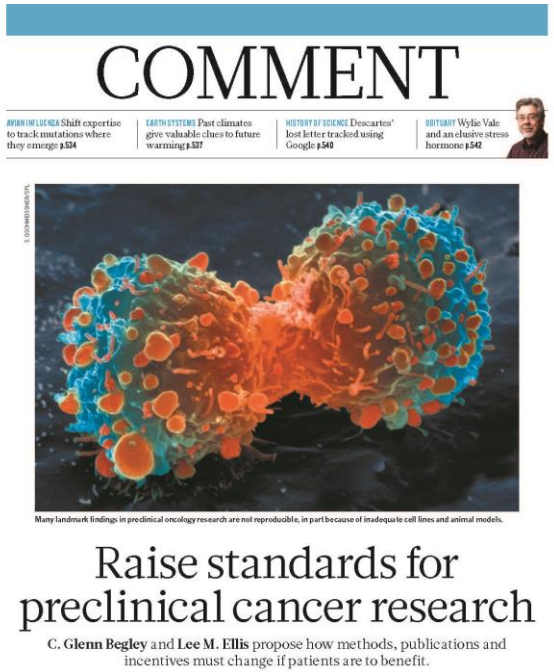
To mitigate some of the risks of such investments ultimately being wasted, most pharmaceutical companies run in-house target validation programmes. However, validation projects that were started in our company based on exciting published data have often resulted in disillusionment when key data could 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. However, there is an imbalance between this apparently widespread impression and its public recognition (for example, see REFS 2,3), and the surprisingly few scientific publications dealing with this topic. Indeed, to our knowledge, so far there has been no published in-depth, systematic analysis that compares reproduced results with published results for wet-lab experiments related to target identification and validation.

Early research in the pharmaceutical industry, with a dedicated budget and scientists who mainly work on target validation to increase the confidence in a project, provides a unique opportunity to generate a broad data set on the reproducibility of published data. To substantiate our incidental observations that published reports are frequently not reproducible with quantitative data, we performed an analysis of our early (target identification and validation) in-house projects in our strategic research fields of oncology, women's health and cardiovascular diseases that were performed over the past 4 years (FIG. 1a). We distributed a questionnaire to all involved scientists from target discovery, and queried names, main relevant published data (including citations), in-house data obtained and their relationship to the published data, the impact of the results obtained for the outcome of the projects, and the models

Findings confirmed in only 14 out of 67 studies (21%)

Prinz et al Nature reviews Drug discovery 10.9 (2011): 712-712.

Amgen



COMMENT

AVIAN INFLUENZA Shift expertise to track mutations where they emerge p.534 | EARTH SYSTEMS Past climates give valuable clues to future warming p.537 | HISTORY OF SCIENCE Descartes' lost letter tracked using Google p.540 | OBITUARY Wylie Vale and an elusive stress hormone p.542

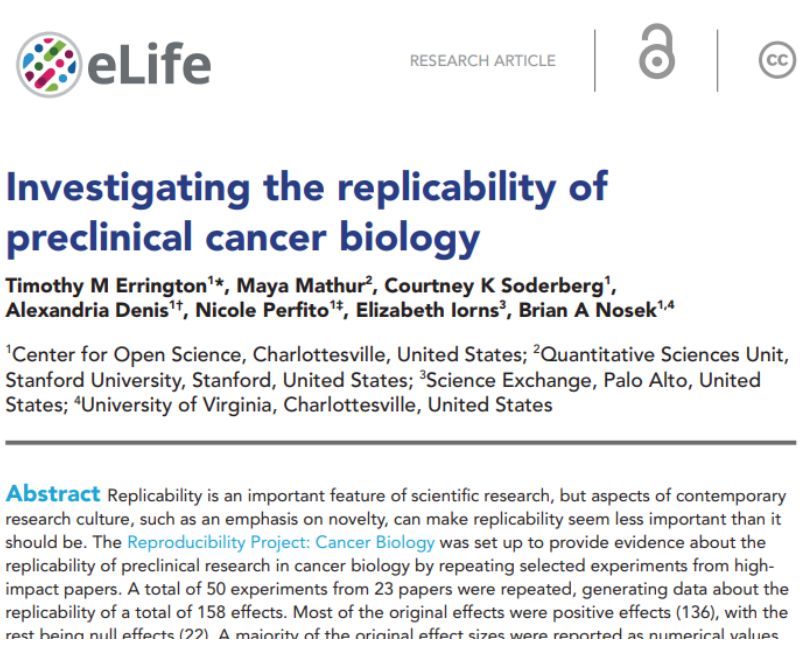
Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Findings confirmed in only 6 out of 53 studies (11%)

Begley and Ellis Nature 483.7391 (2012): 531-533.



eLife

RESEARCH ARTICLE

Investigating the replicability of preclinical cancer biology

Timothy M Errington[1]*, Maya Mathur[2], Courtney K Soderberg[1], Alexandria Denis[1†], Nicole Perfito[1‡], Elizabeth Iorns[3], Brian A Nosek[1,4]

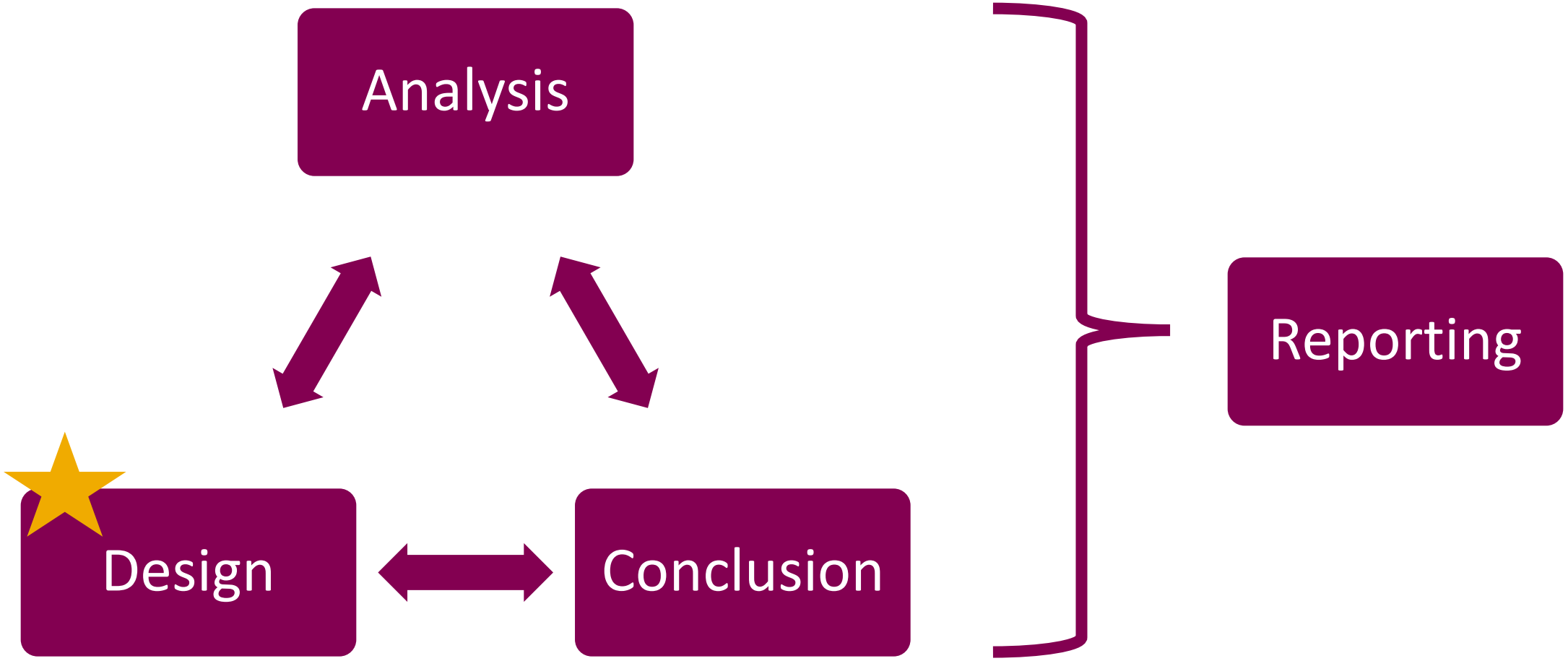
[1]Center for Open Science, Charlottesville, United States; [2]Quantitative Sciences Unit, Stanford University, Stanford, United States; [3]Science Exchange, Palo Alto, United States; [4]University of Virginia, Charlottesville, United States

Abstract Replicability is an important feature of scientific research, but aspects of contemporary research culture, such as an emphasis on novelty, can make replicability seem less important than it should be. The Reproducibility Project: Cancer Biology was set up to provide evidence about the replicability of preclinical research in cancer biology by repeating selected experiments from high-impact papers. A total of 50 experiments from 23 papers were repeated, generating data about the replicability of a total of 158 effects. Most of the original effects were positive effects (136), with the rest being null effects (22). A majority of the original effect sizes were reported as numerical values

Findings confirmed in 46% studies ES: 85% smaller

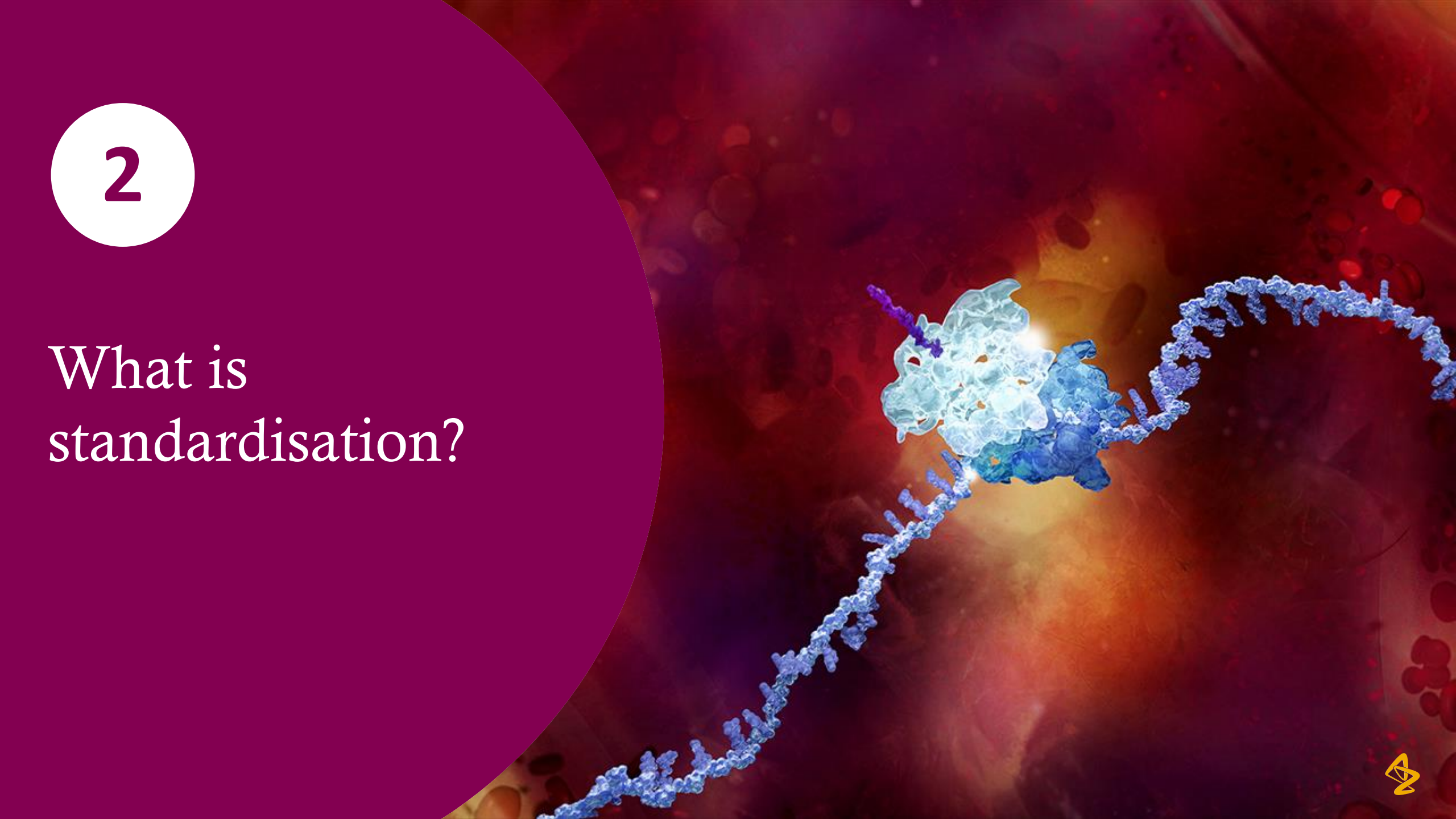
Errington et al Elife 10 (2021): e71601.

Multi-faceted



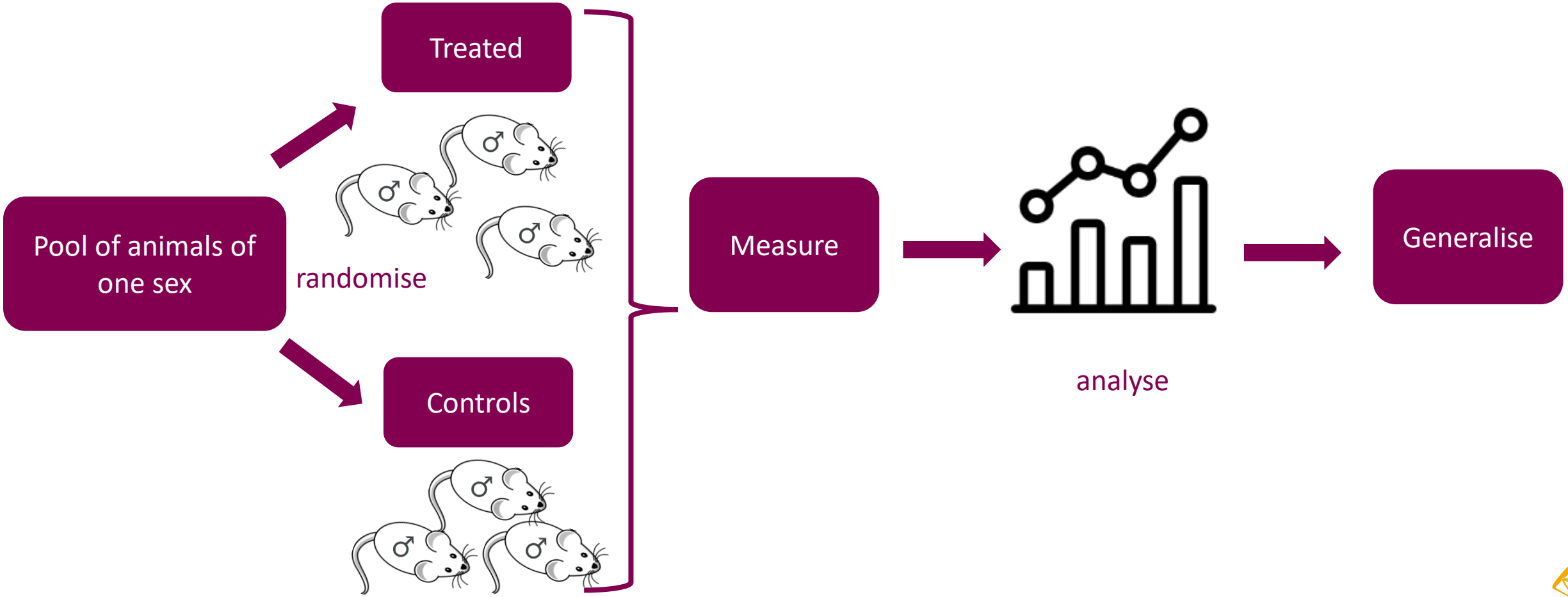
2

What is standardisation?

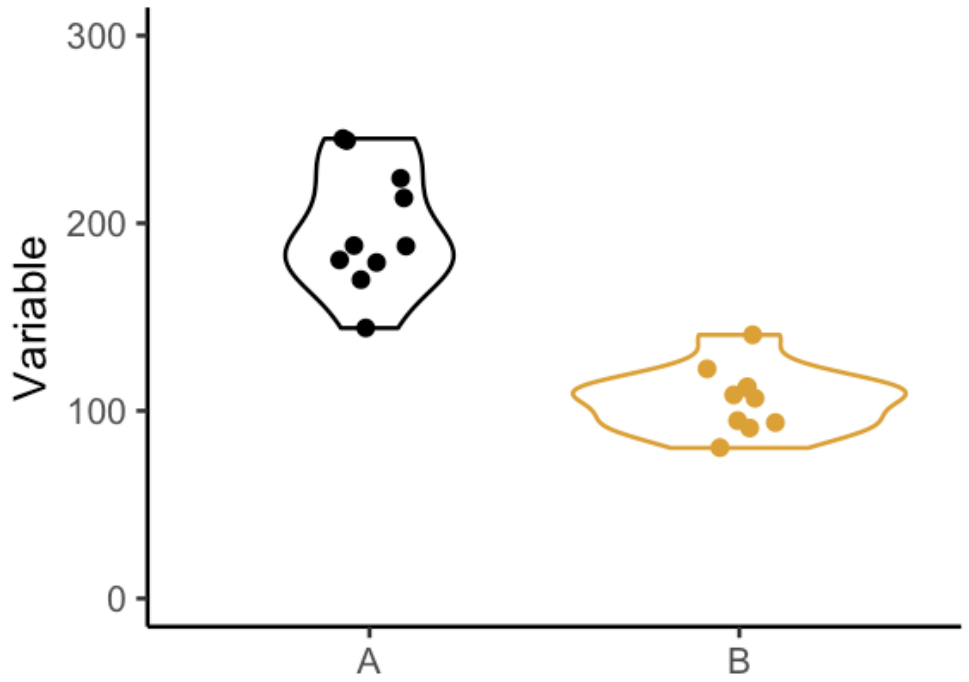


Scientific inference: requires simplification

Isolating cause and effect - Feasible testing space



Research objective: isolate cause and effect

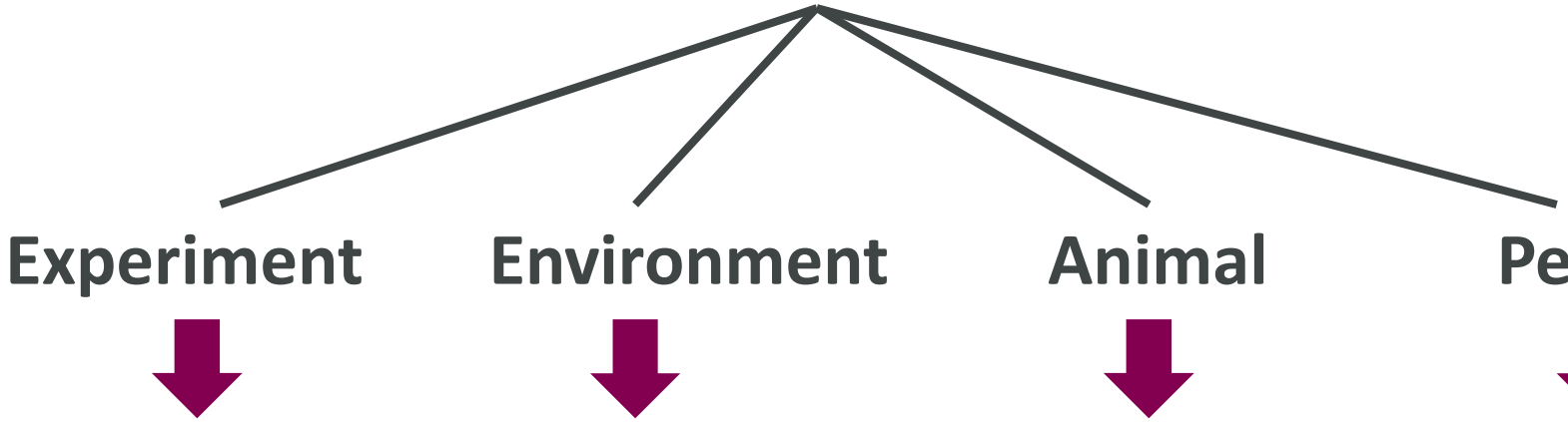


Have high **Internal validity** when we are confident that no other explanation for the observed effect

If there are systematic differences which correlate with the outcome variable then we cannot disentangle the intervention effect from the second systematic differences and the experiment is **confounded**.

Statistical tests are used to understanding sources of variation and assess whether effect is a significant difference.

Many potential sources of variation that can alter the response



Experiment

Treatment
Dose level
Time
Shelf life
Instrument calibration
Sampling

Environment

Temperature
Noise
Light
Smell
Cage Size
Bedding
Enrichment activity
No. of animals
Diet
Handling

Animal

Species
Strain
Sex
Batch
Age
Weight
Health status

People

Technician
Surgeon
Researcher

Experimental strategies to manage variation

Strategy 1: Factor of interest

- Planned systematic variation
- Will be compared to noise (chance-like variability)

Risk: Confounding (unplanned, systematic variation)

Strategy 2: Control (aka standardise)

- removes variation from potential confounders

Strategy 3: Randomizing: converts into chance-like variability

Strategy 4: Block

- converts into planned, systematic variability.
- Within the block standardise vary between block
- Source of noise you want to account for to increase generalisability but maintain sensitivity

In-vivo research has a strong push to minimise N

- 3Rs - guiding principles for ethical use of animals in testing
  - Replacement, Reduction, and Refinement

	Standard	**Contemporary**
Reduction	Methods which minimise the number of animals used per experiment	Appropriately designed and analysed animal experiments that are robust and reproducible, and truly add to the knowledge base

https://www.nc3rs.org.uk/the-3rs

Why the push to standardise?

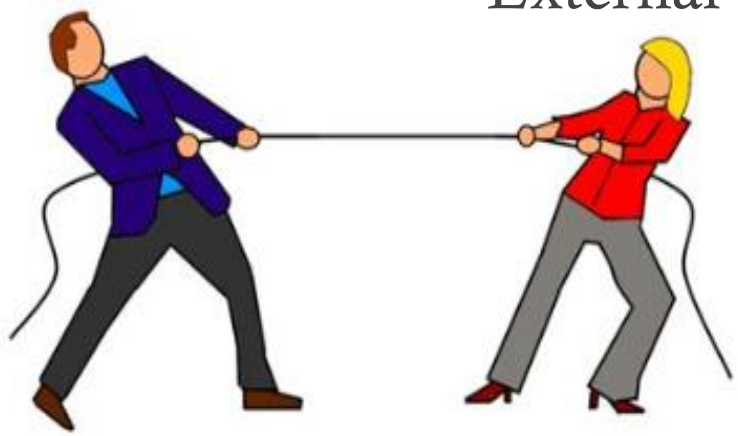
- Simple strategy to manage potential confounders

- If you lower variability then you will increase sensitivity
  - stat test: signal/variability

  But limited generalisability

Ability to generalise

External Validity



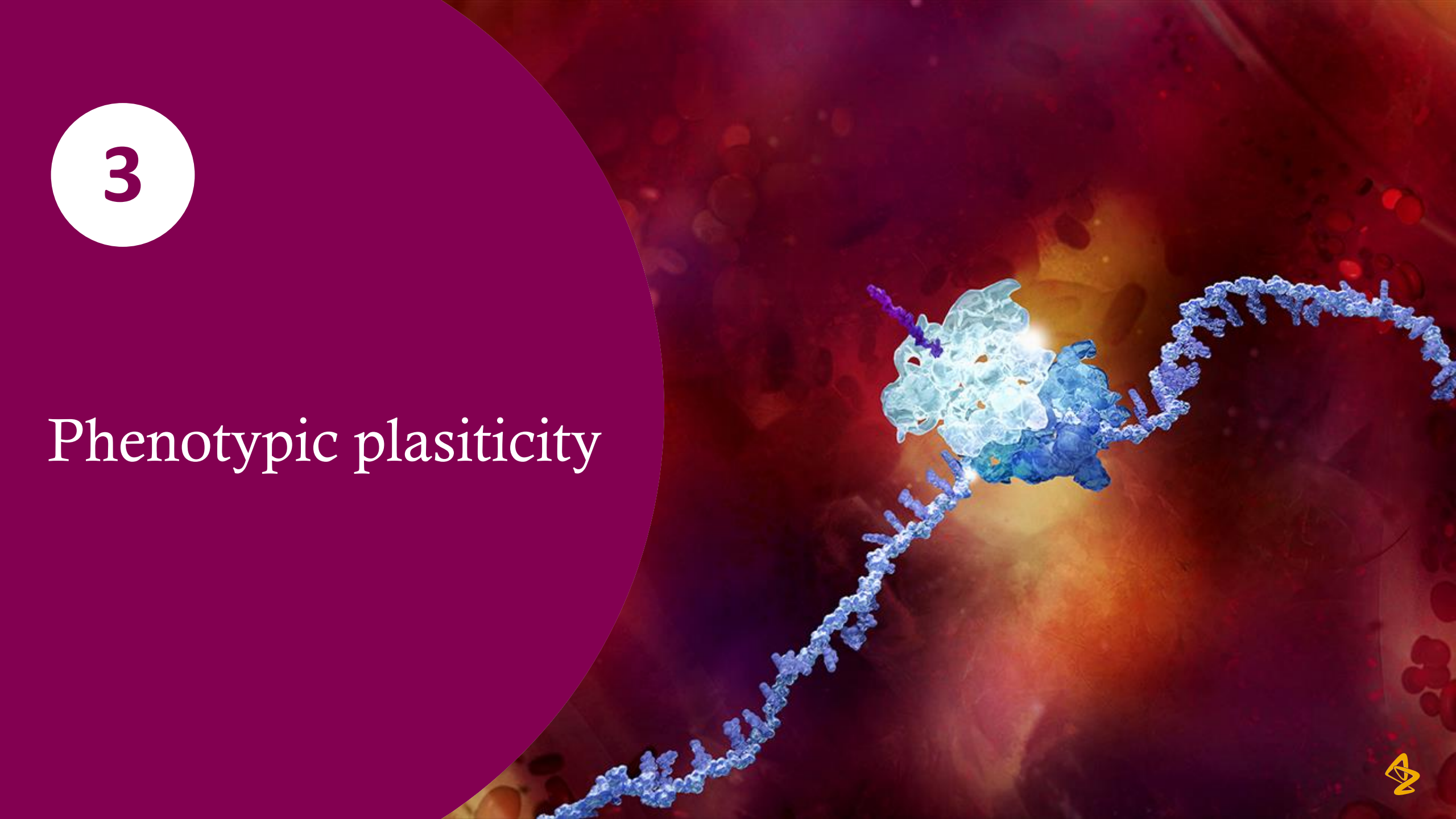
Internal Validity

Ability to isolate cause and effect

Sketch courtesy of Bartosz Bajer

3

Phenotypic plasiticity



Context dependent outcome

- Phenotypic plasticity: Living organisms are highly responsive to the environment with phenotypic changes with both long- and short-term duration. This is adaption and is ensure optimal fit and an essential component for survival.

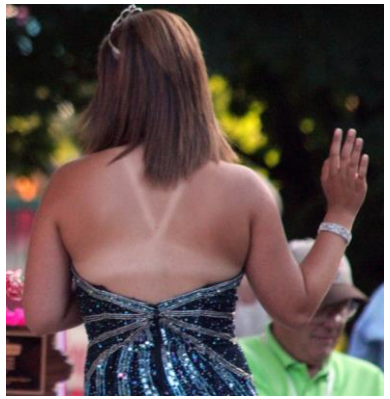


Image: Quinn Dombrowski from Berkeley, USA

- The direction and magnitude of a treatment effect depends not only on the nature, duration, and intensity of the treatment, but also on the animals/cells current phenotype and the experimental context

Aligned to pivotal 1999 Crabbe *et al* study

- 3 laboratories
- 6 strains
- 6 behavioural screens
- Extensive standardisation of environment
- 8 independent outcomes

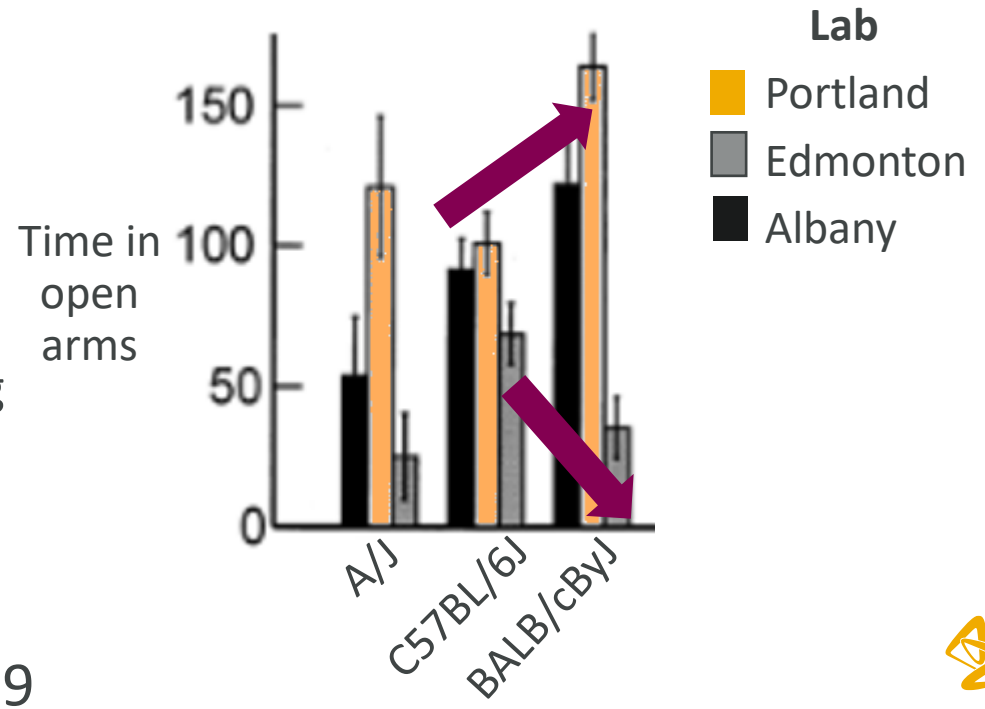
Genotype	Site	Sex	Genotype*Sex	Genotype*Site
8/8	6/8	3/8	2/8	5/8

Standardisation fallacy

There is no pure treatment effect

With every additional variable that is standardised the testing space (inference space) narrows

Biological variation is norm – therefore treatment effect can only be meaningful interpreted relative to biological variation



Crabbe *et al* Science 1999

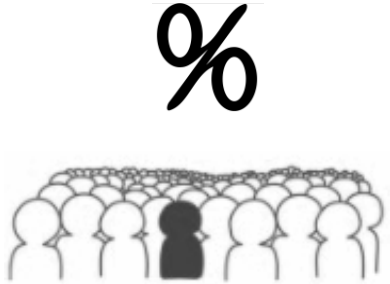
Simplification leads to irreproducible, unrepresentative research

Much of preclinical science is analogous to testing clones of the same middle-aged white American, who lives in the suburbs, doesn't smoke or drink and goes to the gym 3 times a week.

4

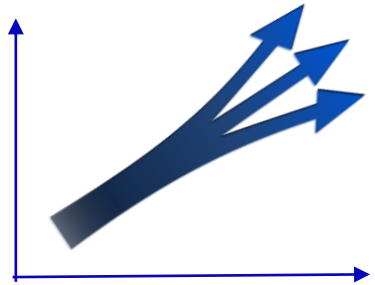
Sex as a biological variable

Sex matters clinically

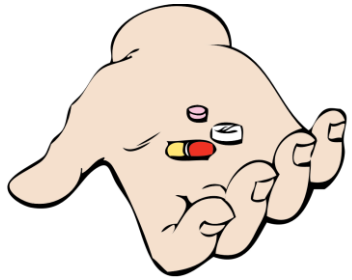


Prevalence

Symptoms



Progression



Side Effects

COVID-19 [Bwire 2020; Doerre & Doblhammer 2022]

- Prevalence higher in ♀ but higher morbidity and mortality in ♂
  - Biological differences ?
    - Higher expression ACE 2 receptor for coronavirus in ♂
    - Immunological differences driven by sex hormone and X chromosome
- Gender differences
  - ♀ - more contacts, work in care roles
  - ♂ higher rates of smoking and drinking
  - ♂ Lower uptake of preventative measures

Embedded neglect of sex within preclinical research

- Reporting:
  - In vivo: Sex not specified – 22% did not specify Yoon et al 2014
  - In vitro: 75% did not report the sex Shah 2014
- Experimental design:
- In vivo: comparison across 9 fields of biology, 2009 to 2019 Beery 2020
  - Inclusion increased: 28 to 49%; 6/9 fields significant improvement.
  - In vitro: 69 -80% male only Taylor 2011, Shah 2014
- Analysis (In vivo):
  - When both sexes (N=356), only 42% sex-based analysis Beery 2020
  - Those reporting sex differences: 1/3 did not test statistically

Garcia-Sifuentes & Maney 2021

Take home: Two problems – including both sexes and appropriate analysis

International Mouse Phenotyping Consortium

nature COMMUNICATIONS

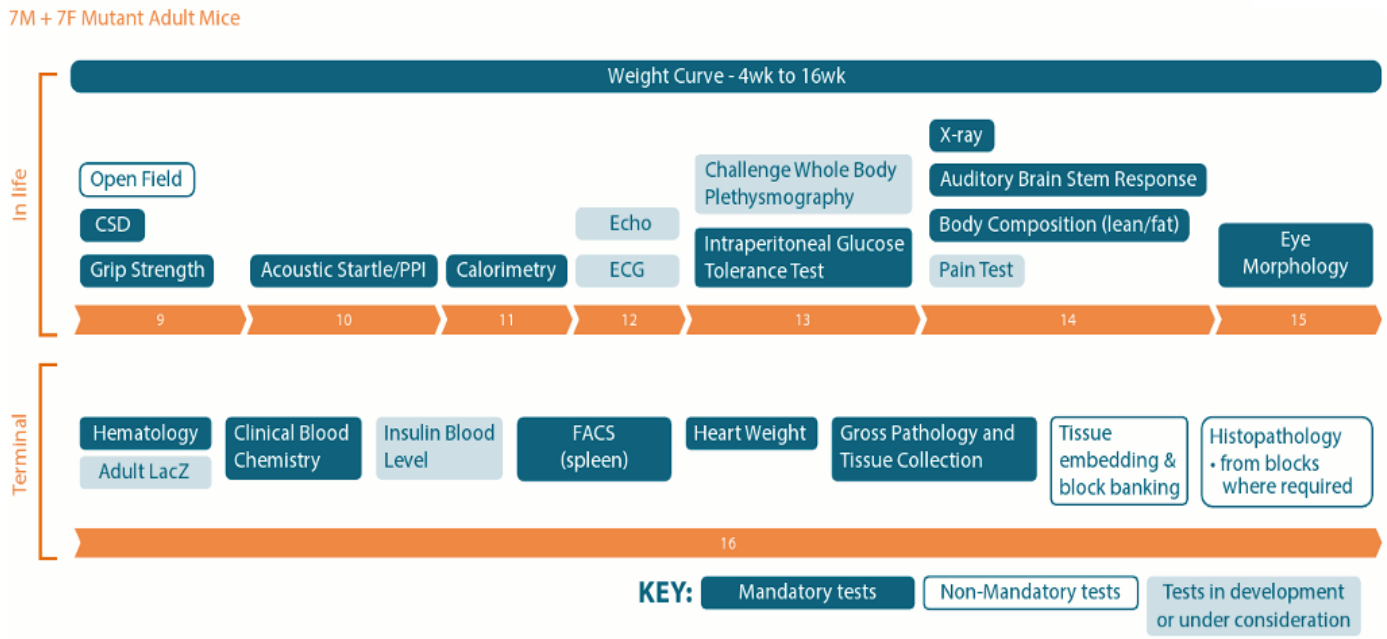
ARTICLE

Received 27 Oct 2016 | Accepted 30 Mar 2017 | Published 26 Jun 2017

DOI: 10.1038/ncomms15475 OPEN

Prevalence of sexual dimorphism in mammalian phenotypic traits

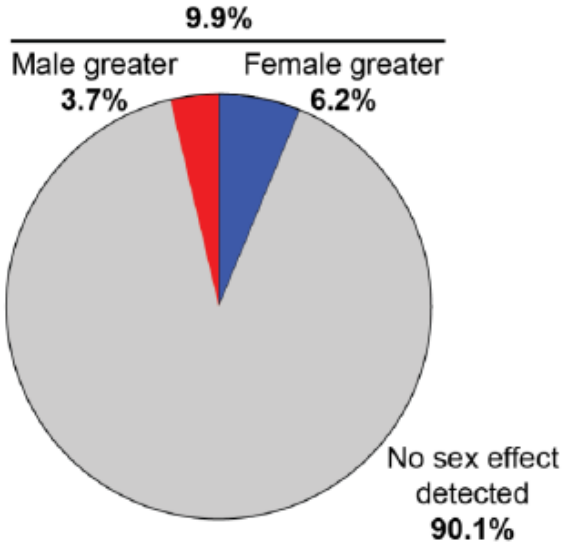
Natasha A. Karp[1,2], Jeremy Mason[3], Arthur L. Beaudet[4], Yoav Benjamini[5], Lynette Bower[6], Robert E. Braun[7],



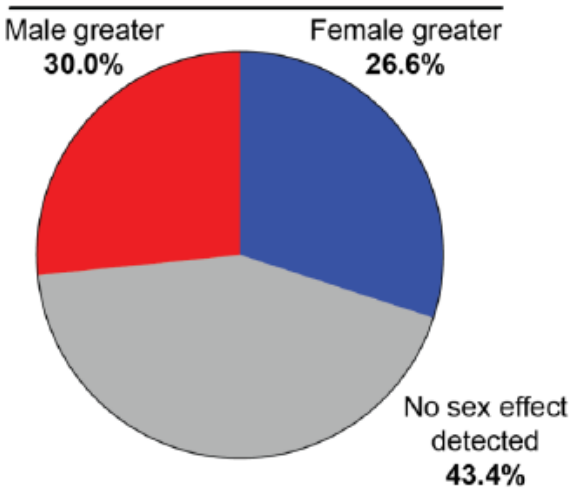
- 10 institutes
- 14,250 wildtype mice
- 40,192 mutant mice
- 2186 mutant lines
- up to 234 traits.

Sex matters

In control data

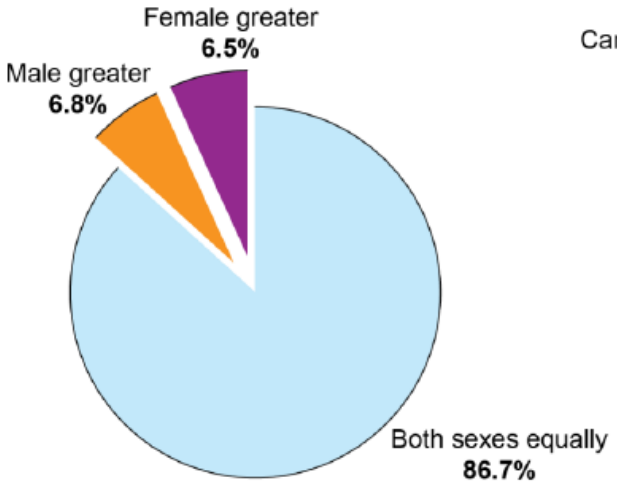


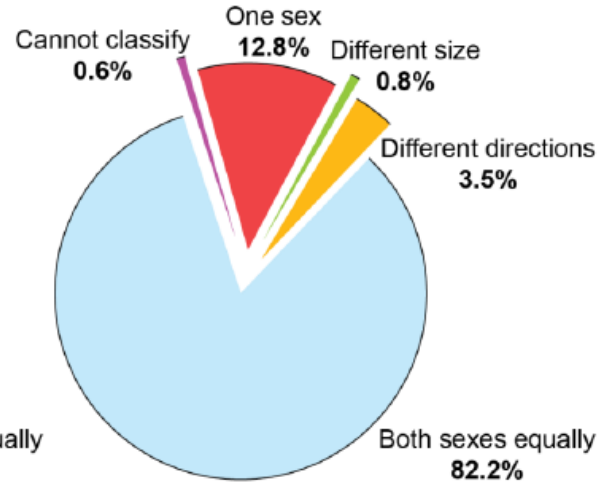
Categorical
N=545



Continuous
N=903

As a modifier of treatment effect?





Categorical

No ds = 266,952
No ds sig = 1,220

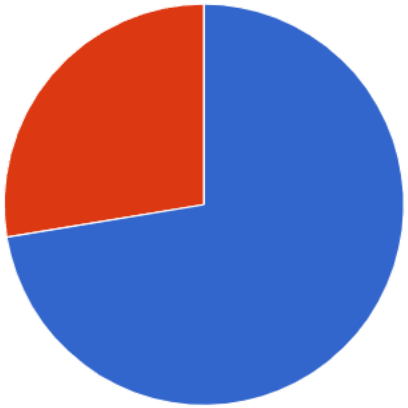
Continuous

No ds = 110, 586
No ds sig = 7 929

Karp et al 2017 Nature Communications

Emerging evidence that our knowledge base is biased

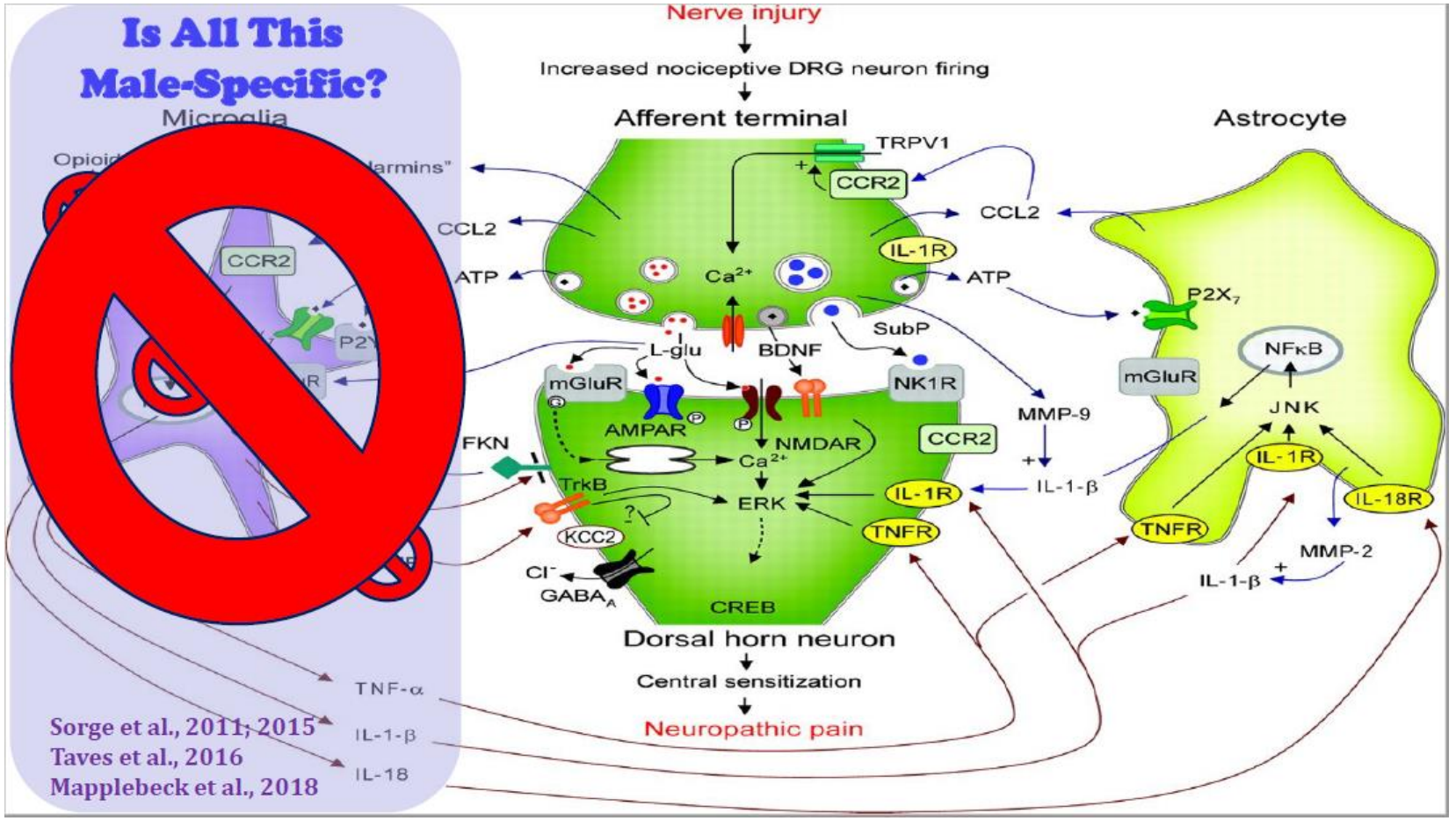
27.6% female only



72.4% male only

Pain processing

N=127



Mogil (2020) Nature Reviews Neuroscience

Sex matters but it isn't perceived as a doable problem

Sociological exploration

- **Generalizability** Important to embrace variation to understand biological differences
- **Avoiding complexity** To make progress in science reduce complexity
- **Practicality** Tension between the above. Impractical

Gompers, Annika. Genderscilab, 2018. www.genderscilab.org/blog/three-years-in-sex-as-a-biological-variable-policy-in-practice-and-an-invitation-to-collaborate

UK MRC survey

95% researchers saw benefit

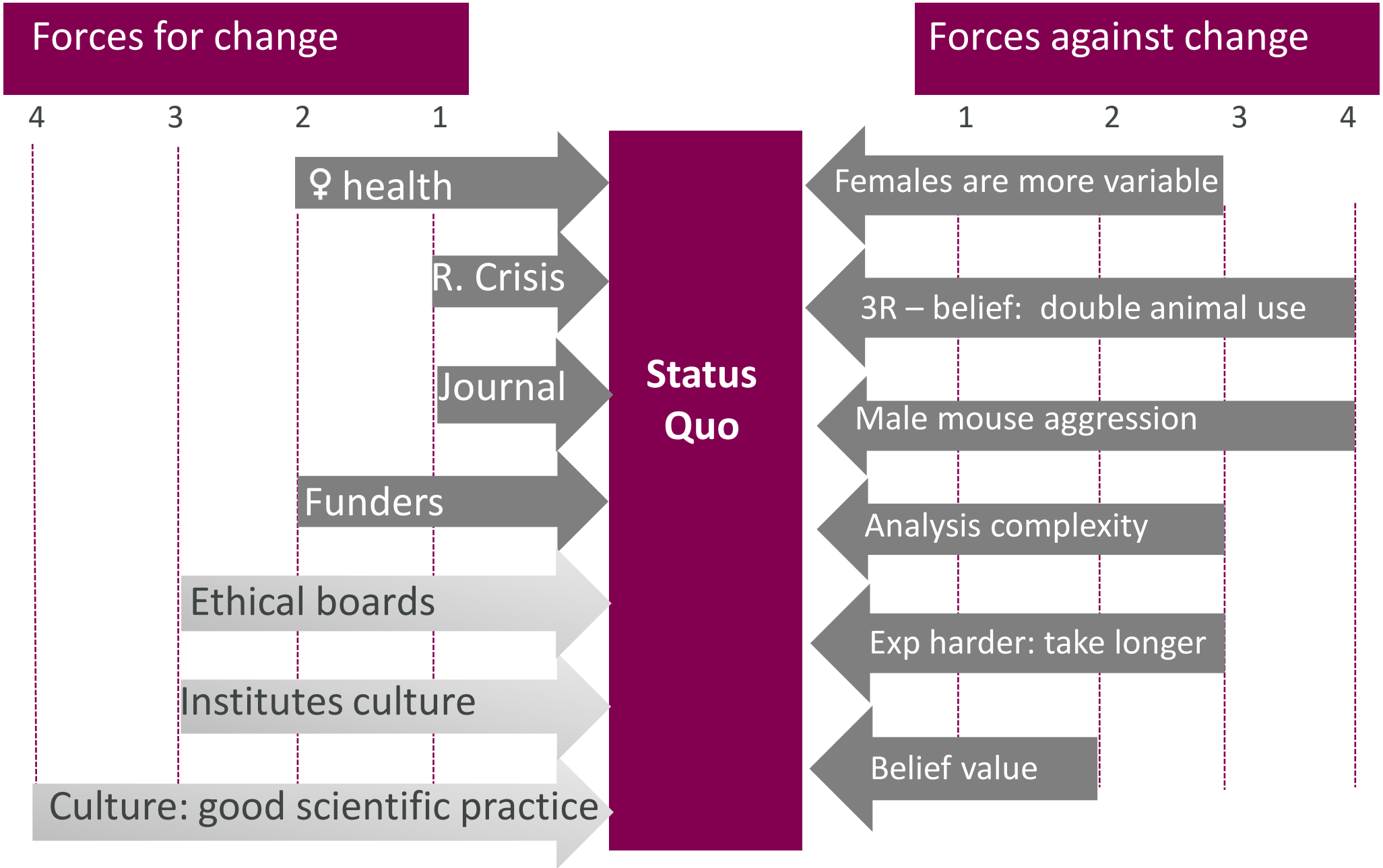
- Translatability
- Reproducibility
- Detecting sex specific effects

But there were barriers/concerns

- Cost of experiments
- Complexity of research design
- Compliance with 3Rs

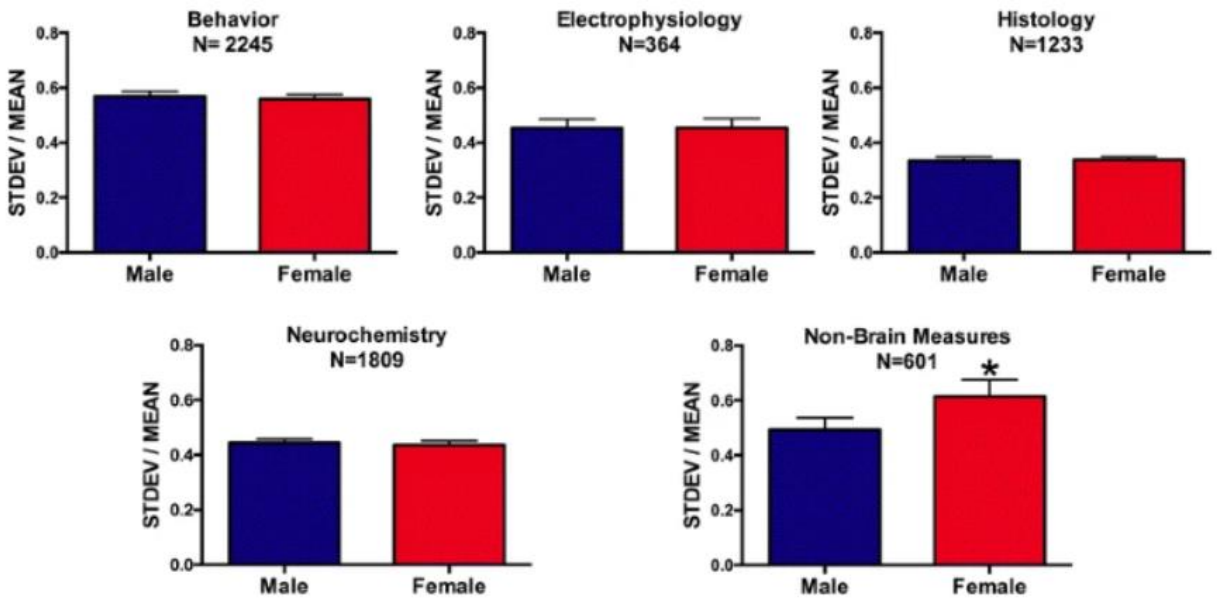
www.ukri.org/wp-content/uploads/2022/03/MRC-090322-SexInExperimentalDesign-SummaryReport.pdf

Lewin's Force field analysis



Misconception: hormonal cycles: females more variable

Rats Becker 2016 BSD



"Female rats were not more variable at any stage of the estrous cycle than male rats."

Mice Prendergast 2014 NNBR

- meta-analysis 293 published articles
- behavioral, physiological, morphological, and molecular traits
- CV distribution = no differences
- At trait level – for three types of traits males were more variable than females

"Randomly cycling female mice were no more variable than males on any trait."

Inclusion isn't at odds with the 3R mindset

- Breeding – produces both males and female animals
  - Research suggest that could be an overproduction of **25 million or more mice and rats** worldwide (Nunamaker & Turner 2023.)
- Reduction in N across experiments – more efficient to include both sexes

	Standard	**Contemporary**
Reduction	Methods which minimise the number of animals used per experiment	Appropriately designed and analysed animal experiments that are robust and reproducible, and truly add to the knowledge base

https://www.nc3rs.org.uk/the-3rs

Misconception: It will DOUBLE my animal usage

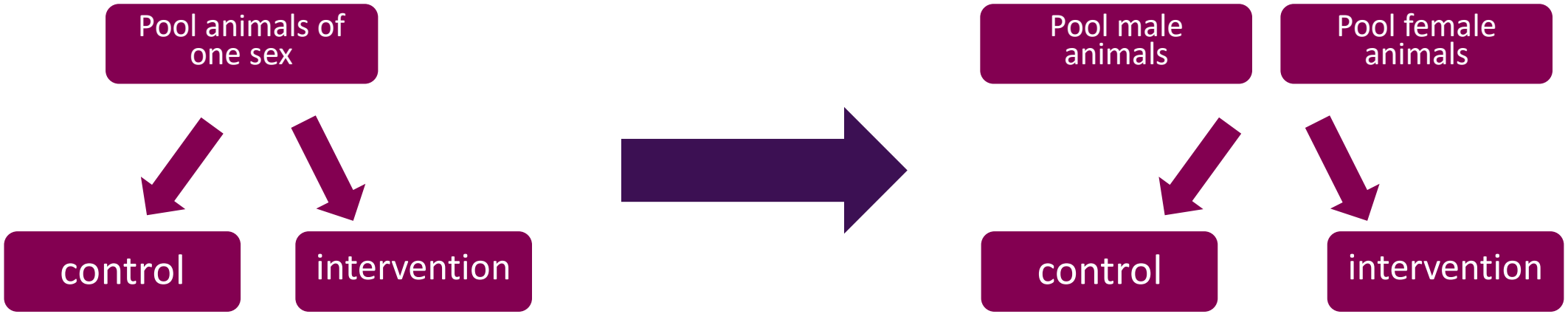
"Keep doing what you are already doing but change half the animals in your study to female"

*McCarthy **2015** Schizophrenia Bulletin*

Evidence:

- **Statistical simulations show that scientists need not increase overall sample size by default when including both sexes in in vivo studies** (Phillips PLoS Biology 2023)
- **Inclusion of females does not increase variability in rodent research studies** (Beery Curr Opin Behav Science 2018)
- **Benefits of a factorial design focusing on inclusion of female and male animals in one experiment** (Buch Journal of Molecular Medicine 2019)

Moving from complete randomised to factorial design



outcome ~ treatment

X pool

X disaggregate

✓ factorial analysis

outcome ~ treatment + sex + treatment*sex

Fear of change

"To date, sex hasn't explained variation in my model"

- Lack of data regarding sex differences does not indicate there are none
- The goal isn't to identify sex differences but to estimate generalisable effects and be able to detect very large differences when they do occur

"My prior work has only considered in one sex"

- Unfortunately, it carries lots of risk.
- " To change is difficult. Not to change is fatal" William Pollard

Funding bodies are driving change

- Movement from recommending to a requirement with active questions in funding process

Body	Year
NIH	2016 – required incorporation both *in vivo* and *in vitro*
Canadian Institute of Health Research	2010 – questions in grant application
Irish Research Council	2013 – questions in grant application
European Commission	2020 – required incorporation both *in vivo* and *in vitro*
MRC	2022 – inclusion of both sexes the default for *in vivo* and *in vitro*
CRUK	*2023 – inclusion of both sexes the default for animal, tissues or cells*

- WT funded: MESSAGE (Medical Science Sex and Gender Equity) project
  - Co-develop a sex and gender policy framework for funders and regulators in the UK

Take home: Expectations are changing. Sex-inclusive designs are becoming the default.

Expectation?

Requirement

- Specify the sex for human or animal tissues and cells used in experiments
- Inclusion of both sexes as default for studies involving animals and human and animal tissues and cells
- Justification for exclusion
- Analysis should account for sex

Vision

- Not to study sex differences but rather estimate a **generalisable effect**
- Experiments are powered to detect the effect of interest across the two sexes
- If the effect is very different between the sexes then this will become apparent

Take home: Goal of sex inclusive research is to estimate an effect that represents both sexes.

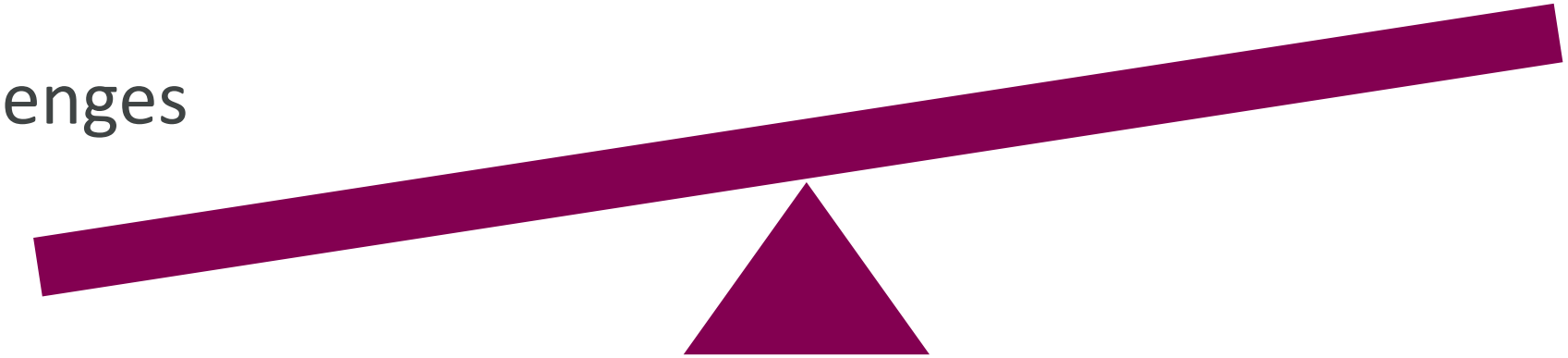
Exception?

- Where sex cannot be determined
- Pure molecular studies such as P-P interactions
- Sex-specific conditions or phenomena e.g. ovarian cancer
- Acutely scare resources (e.g. rare disease)
- If you can provide strong justification.

Single sex justification

Viable experiment

Challenges



The justification could be appropriate following exploration for that study of logistical, ethical, or cost implications relative to the benefit of using both sexes of animals in a research proposal.

SIRF: Sex Inclusive Research Framework

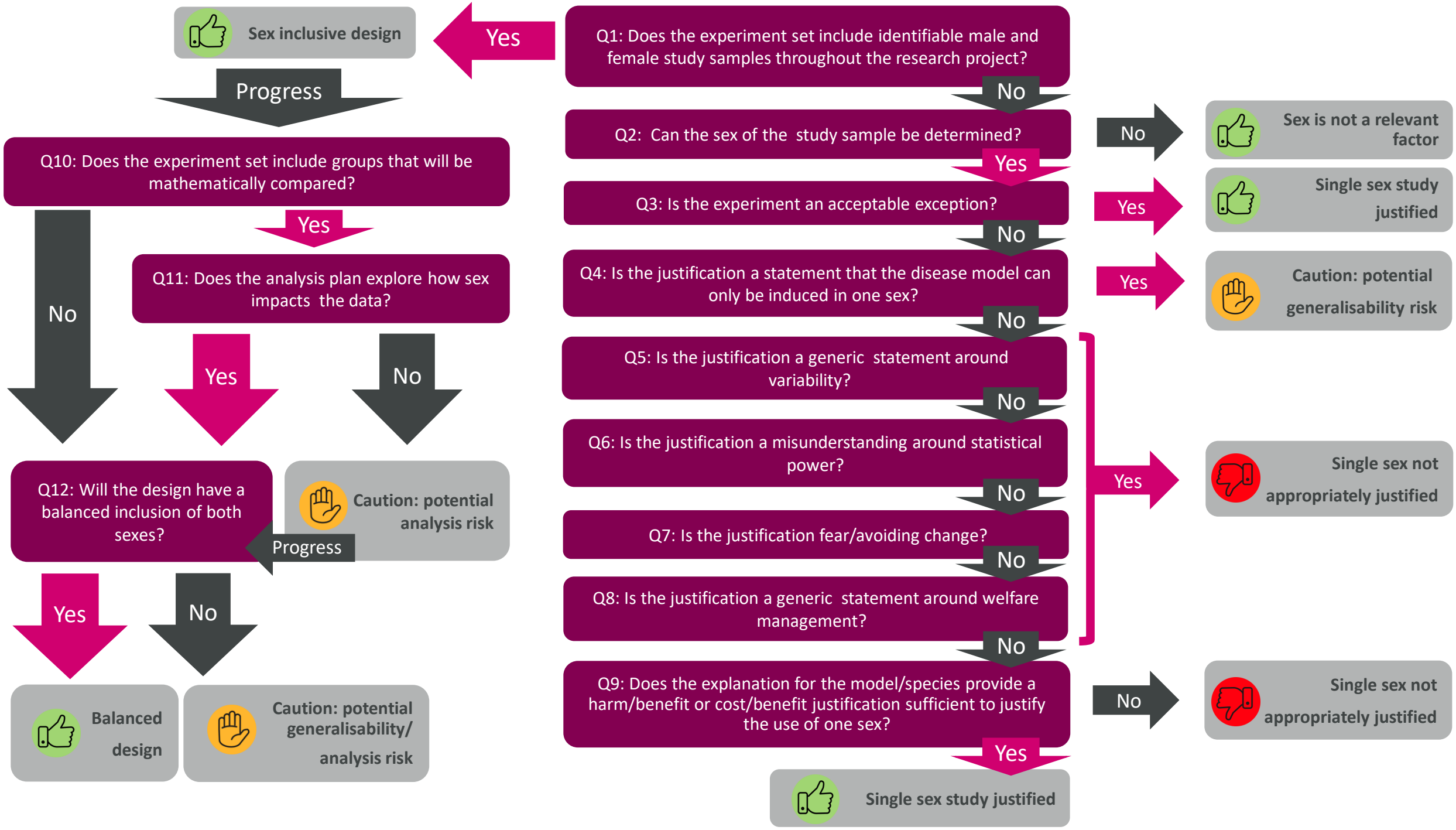
Why?

- Regulatory bodies assessing whether a research proposal is appropriate
- Need transparency in the decision-making process
- We need educational resource to help researchers consider whether sex inclusion is a possibility.
- Frequently, barriers mentioned are misconceptions

What?

- Decision tree of 12 questions and associated supporting information
- Delivers 1 or more classifications
  - **Green**: Proposal is appropriate
  - **Amber**: Caution is required (I.e., the proposed design/analysis carries some risk)
  - **Red**: Justification for single sex study is not sufficient

https://openinnovation.astrazeneca.com/preclinical-research/sex-inclusive-research-framework.html


Sex inclusive design
Yes
Q1: Does the experiment set include identifiable male and female study samples throughout the research project?
No
Q2: Can the sex of the study sample be determined?
No
Sex is not a relevant factor
Yes
Q3: Is the experiment an acceptable exception?
Yes
Single sex study justified
No
Q4: Is the justification a statement that the disease model can only be induced in one sex?
Yes
Caution: potential generalisability risk
No
Q5: Is the justification a generic statement around variability?
No
Q6: Is the justification a misunderstanding around statistical power?
No
Q7: Is the justification fear/avoiding change?
No
Q8: Is the justification a generic statement around welfare management?
No
Yes
Single sex not appropriately justified
Q9: Does the explanation for the model/species provide a harm/benefit or cost/benefit justification sufficient to justify the use of one sex?
No
Single sex not appropriately justified
Yes
Single sex study justified
Progress
Q10: Does the experiment set include groups that will be mathematically compared?
Yes
Q11: Does the analysis plan explore how sex impacts the data?
Yes
No
No
Q12: Will the design have a balanced inclusion of both sexes?
Caution: potential analysis risk
Progress
Yes
No
Balanced design
Caution: potential generalisability/ analysis risk

Examples “Caution is required”

Arises?

- Unbalanced inclusive designs
  - **Generalisability/analysis risks**
- Inclusive designs that do not consider sex in the analysis
  - **Analysis risks**
- Studies for disease which effects both sexes but the model can only be induced in one
  - **Generalisability risk**

Example scenario

'In all experiments, male and female littermates will be pooled together and analyzed as one group"

- Q1 – inclusive? Yes
- Q10 – Groups compared? Yes
- Q11 – analysis considers sex? No

Sex inclusive design

Caution: potential generalisability/ analysis risk

Decision to proceed depends on reflection on the risk

Examples  "Single sex not appropriately justified"

Arises?

Misconceptions

- "Females are more variable"
- "Including both sexes will increase the variation in my data"
- "Including both sexes will double the sample size needed"

Fear/Avoiding change

- "My previous data is all in one sex"
- "Sex hasn't been shown to date to matter"

Example scenario

We plan to use male mice, as female mice tend to have twice the levels of circulating CORT as males, and these levels may shift in response to stage of the estrus cycle.

Q1 – inclusive? No

Q2 – Can the sex be determined? Yes

Q3 – acceptable exception? No

Q4 – disease model induction issue? No

Q5 – generic statement around variability Yes

Single sex not appropriately justified

Examples “proposal is appropriate outcomes”

Exception?

Female mice implanted with patient derived ovarian cancer tumours

Q1 – inclusive? No

Q2 - Can the sex be determined? Yes

Q3 – acceptable exception? Yes

Single sex study justified

Harm &/or Cost evaluation versus benefit

Th9 transfer experiments will be done in male mice because Foxp3Sf donor Th9 cells are obtained from male mice and could not be transferred to female recipients due to risk of rejection.

Q1 – inclusive? No

Q2 – can the sex be determined? Yes

Q3 – acceptable exception? No

Q4 – disease model induction issue? No

Q5:8 – misconceptions/fear of change? No

Q9: Cost &/or harm versus benefit? Yes

Single sex study justified

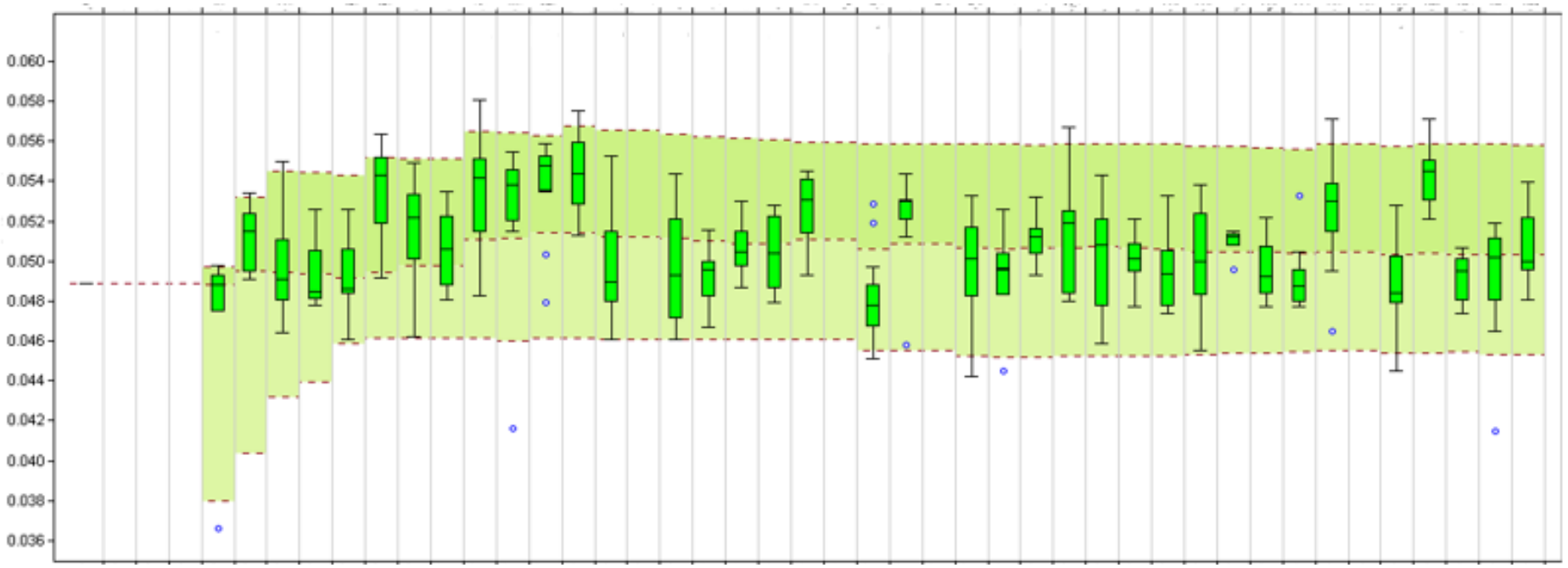
5

Batch impact on reproducibility

Temporal variation in highly standardised pipelines

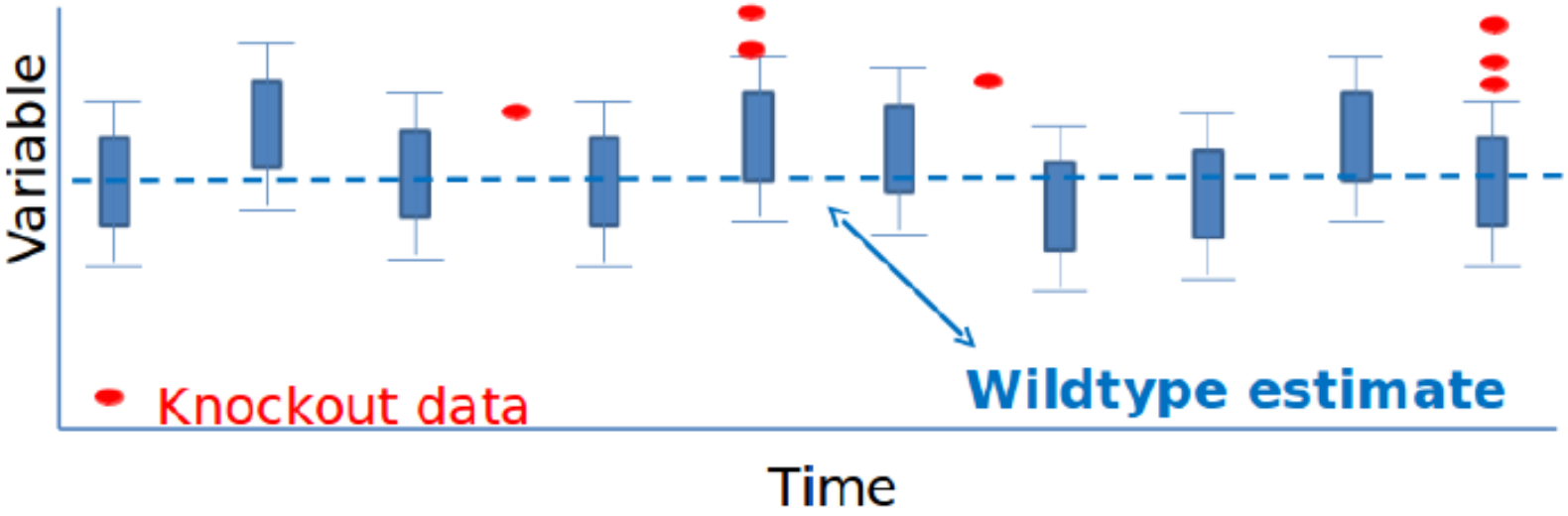
Bone Mineral Density

male wildtypes



Batch

Analysis had to account for batch variation

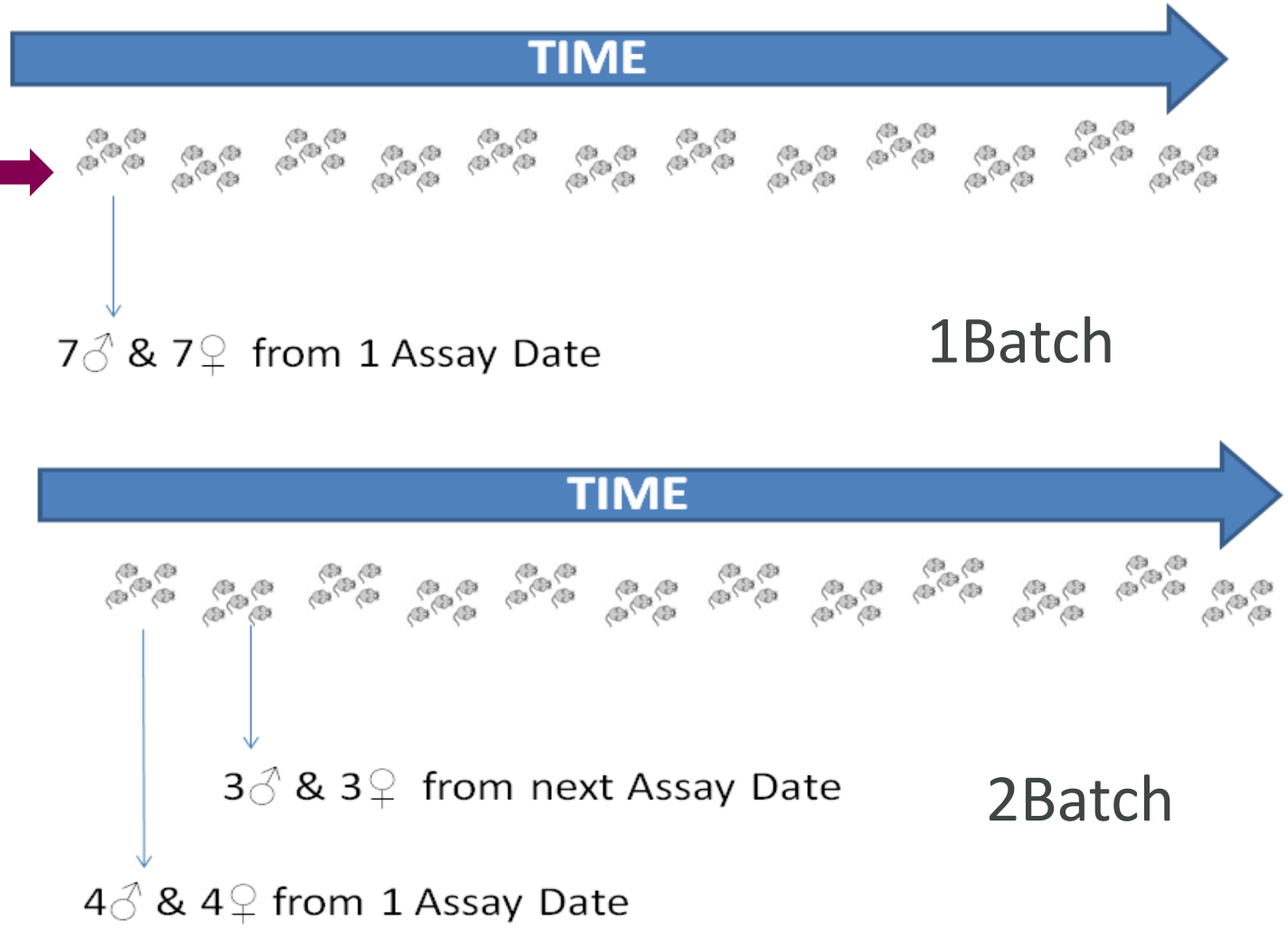


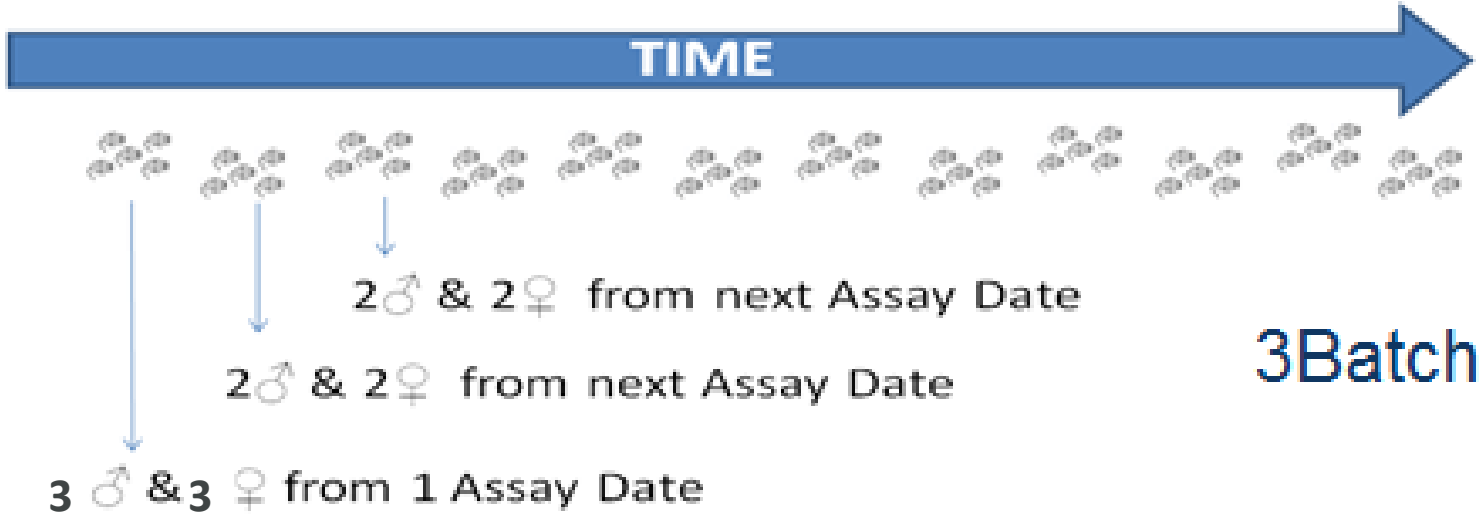
Assume batch is randomly, independent and normal distributed

Karp et al 2012 PLOS ONE

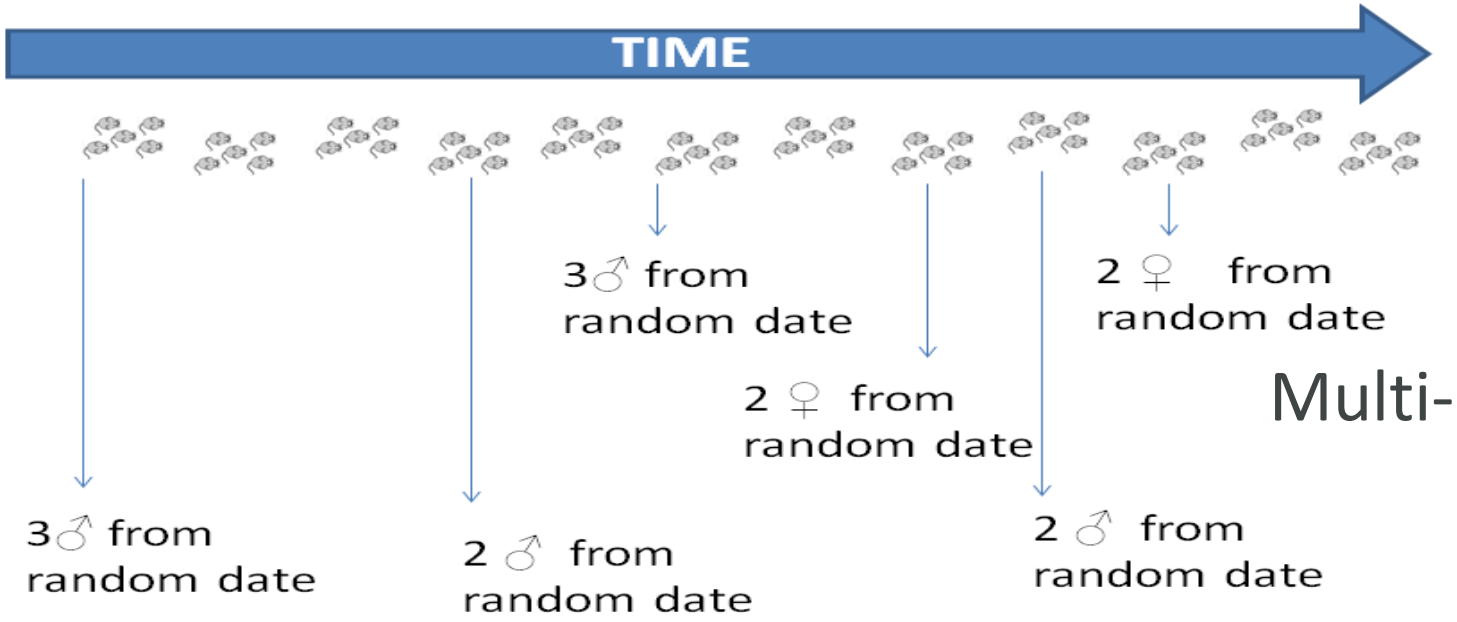
Different institutes had different workflows

Control animals collected in multiple batches



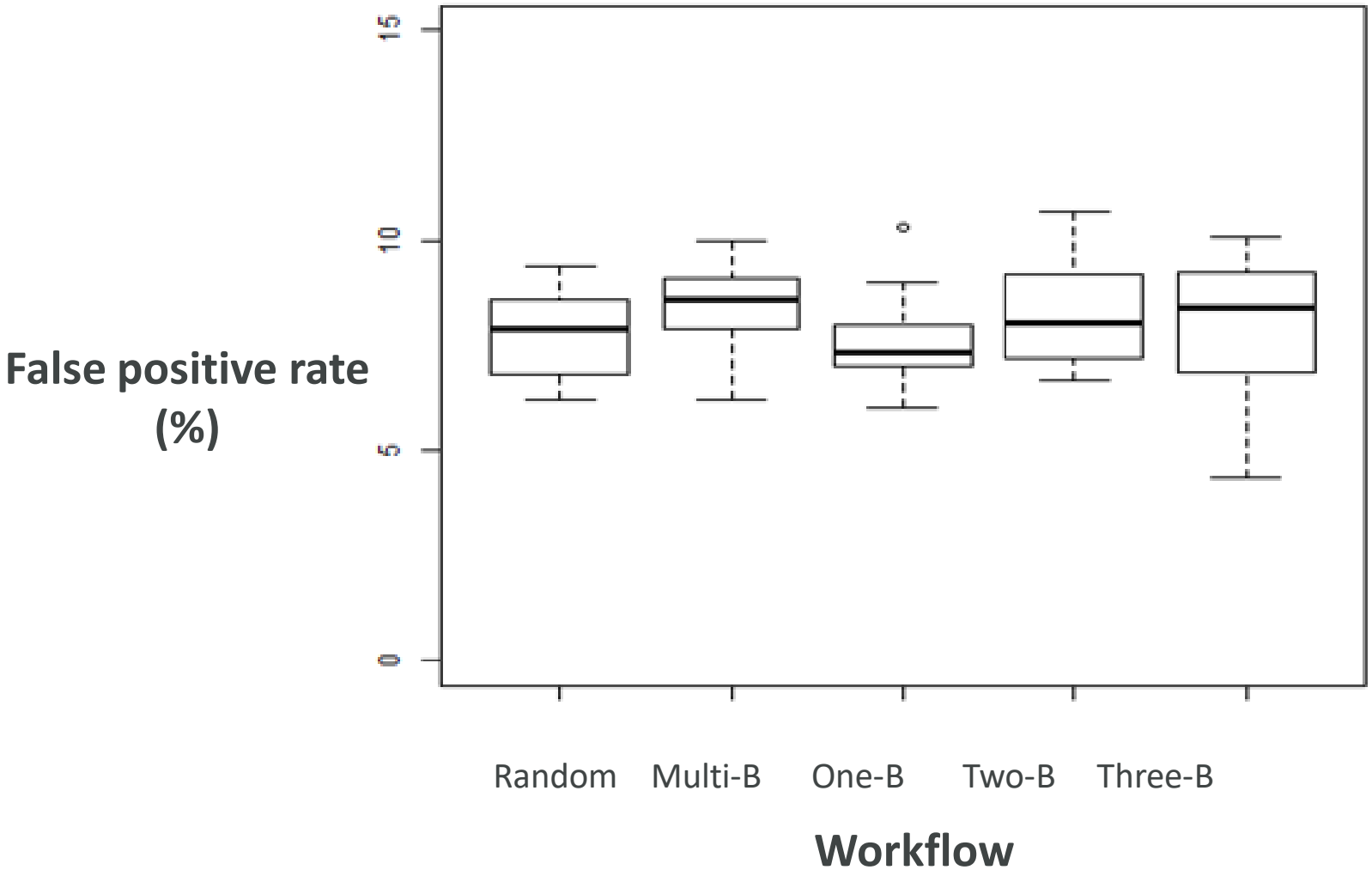


3Batch



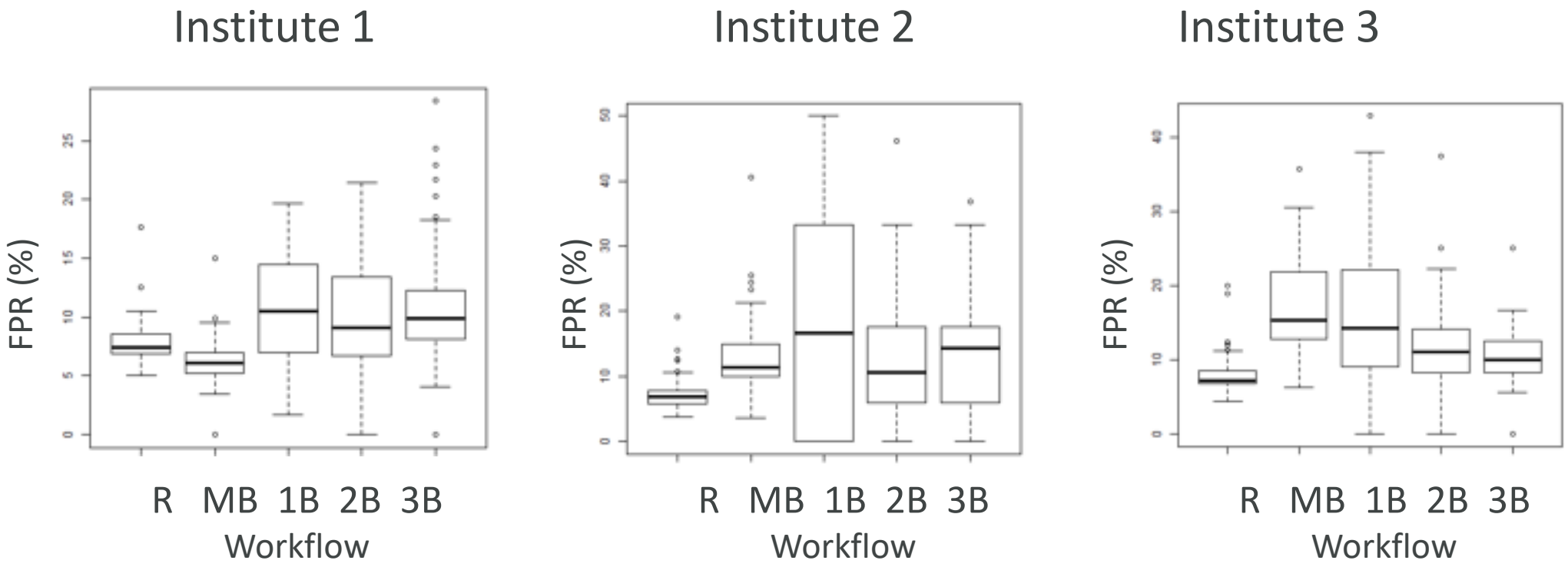
Multi-batch

Performance of the test when no difference - simulated data



Karp et al 2014 PLOS ONE

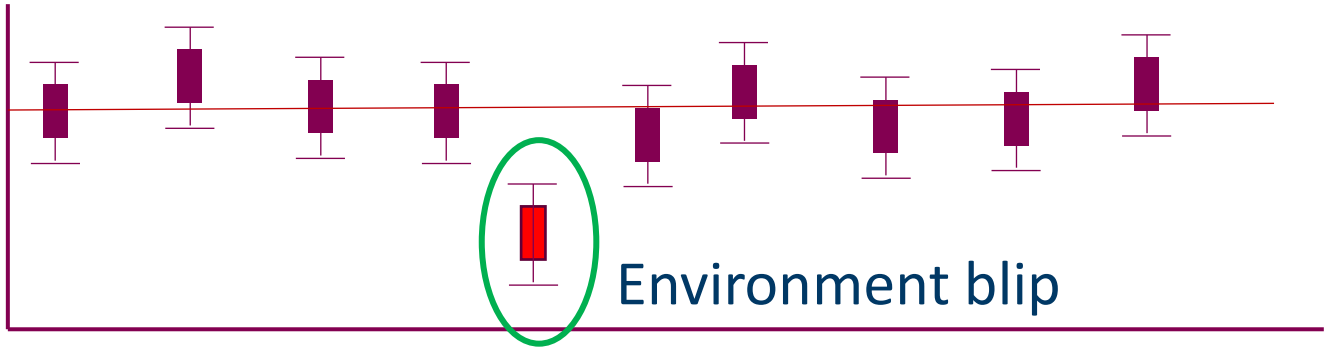
With real data, the FPR depended on workflow



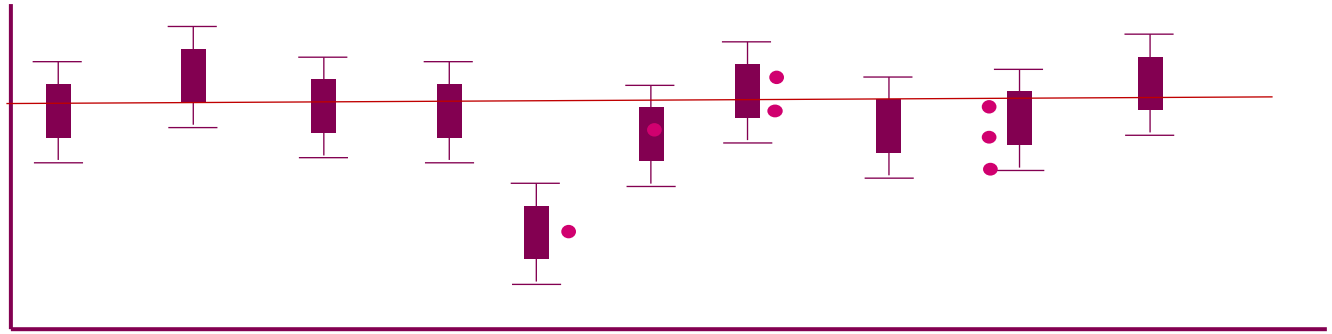
Karp et al 2014 PLOS ONE

Why?

Model assumes batch is normal distributed and makes global estimates



model assume Δ all genotype

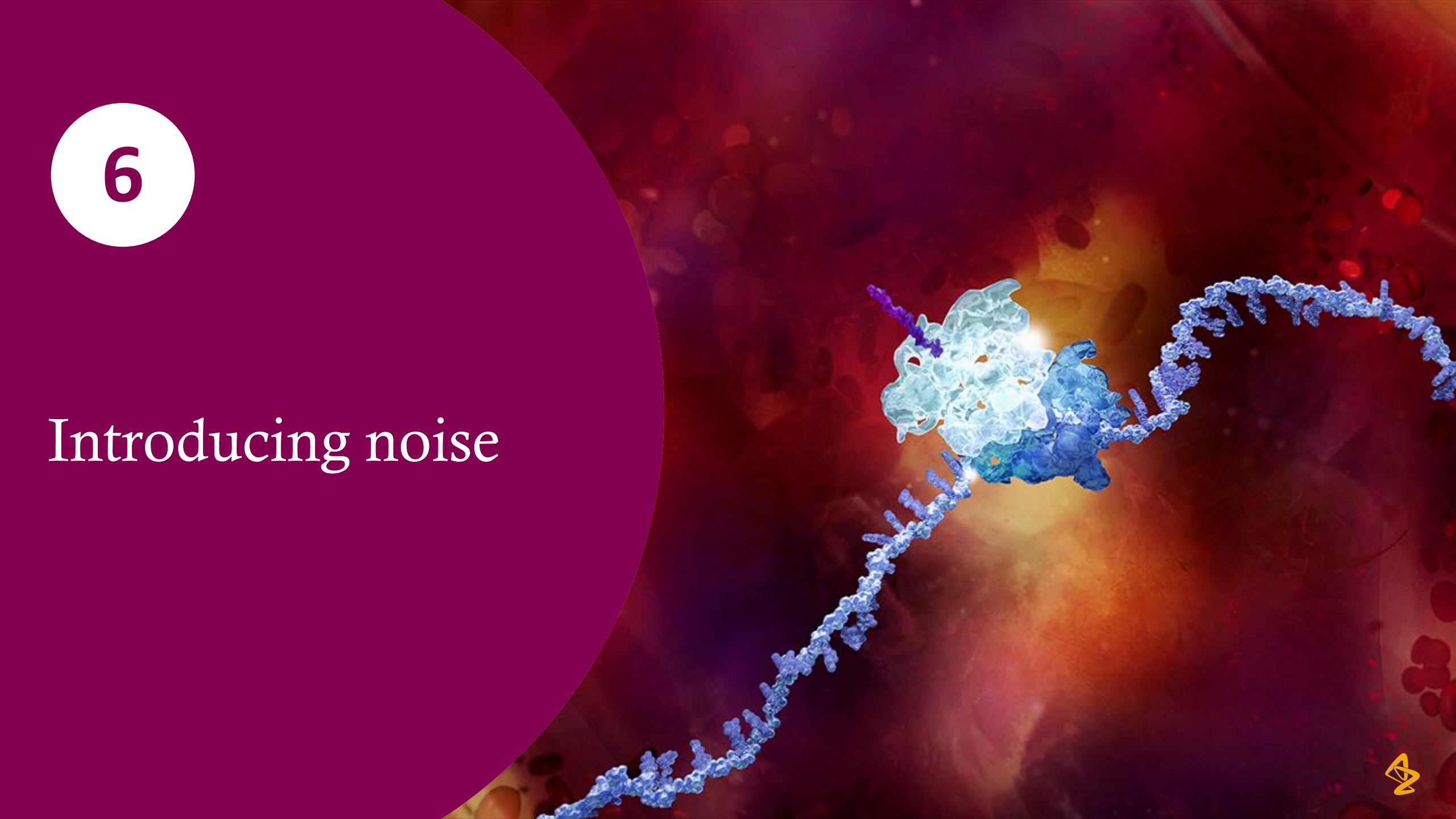


Even within a highly standardised pipeline environmental differences are impacting the observed phenotype

The environmental differences are not the things we think to standardise or capture as meta data

6

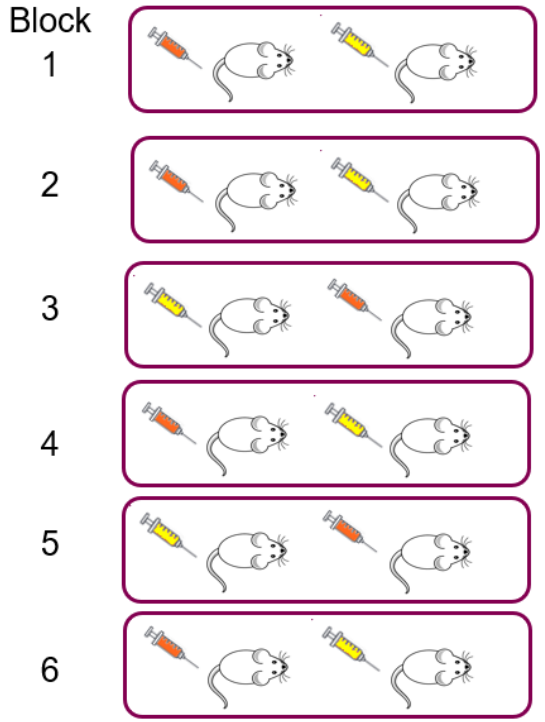
Introducing noise

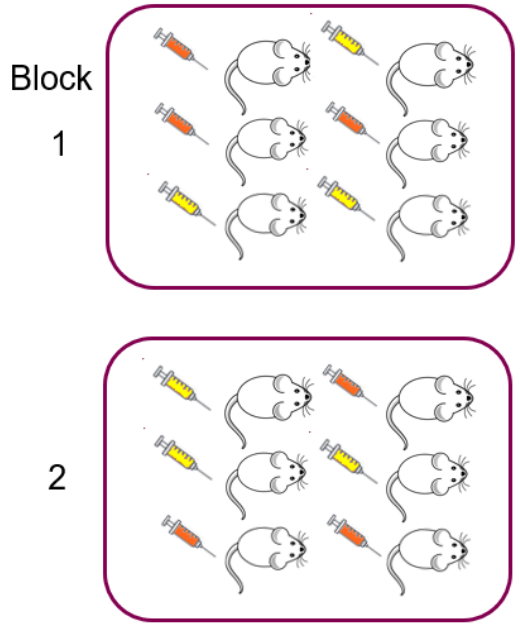


Planned variation

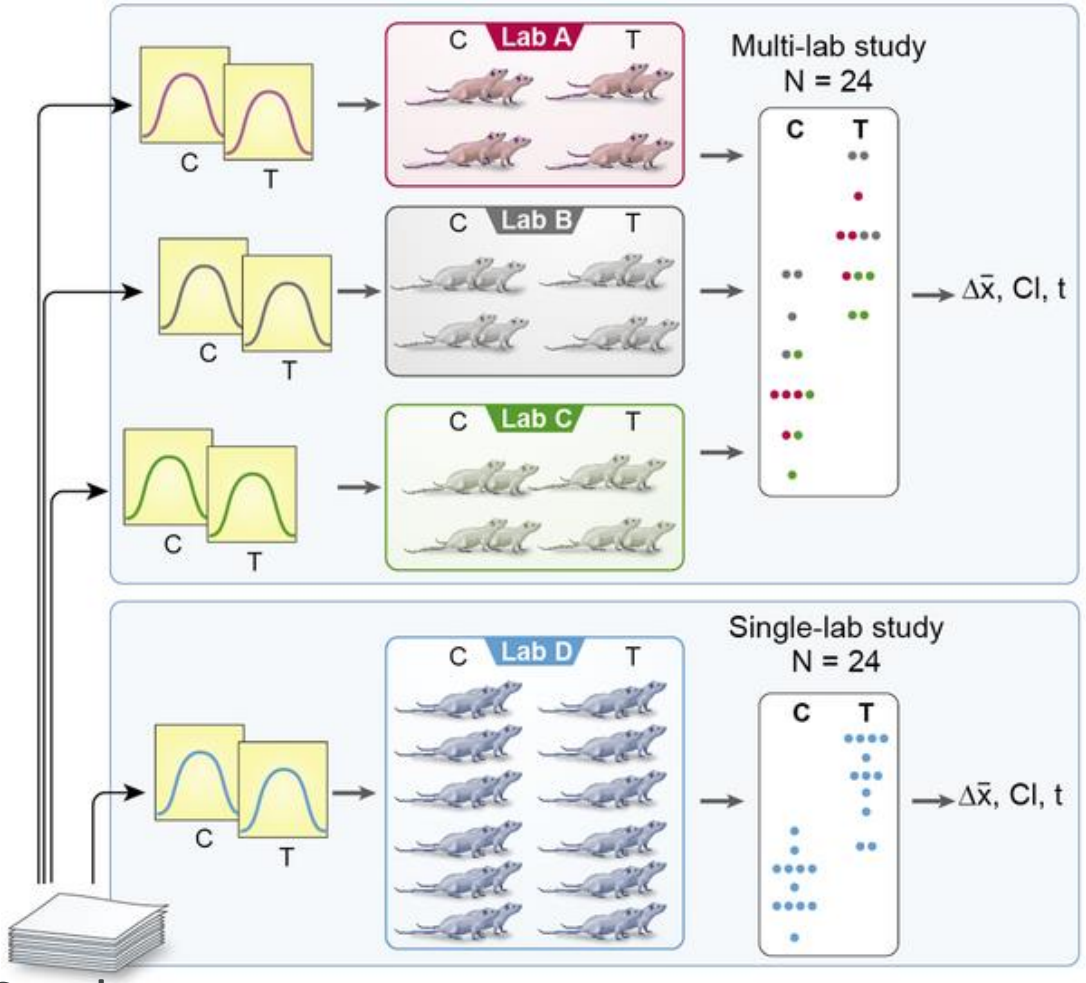
Blocking: standardisation and heterogenisation

- Within a block standardise.
- Vary between blocks.
- Can identify treatment effects that are consistent across blocks (generalise) and with replicate within block explore where the effect differs





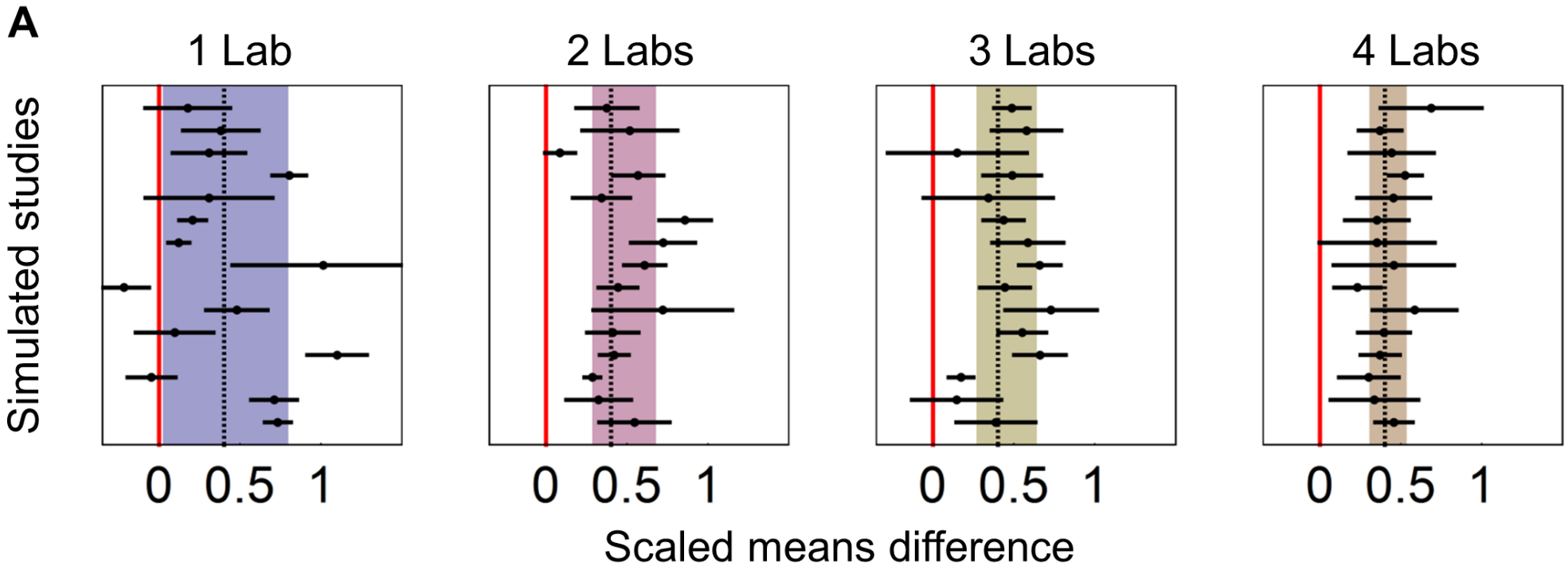
Simulations of a multi-lab design

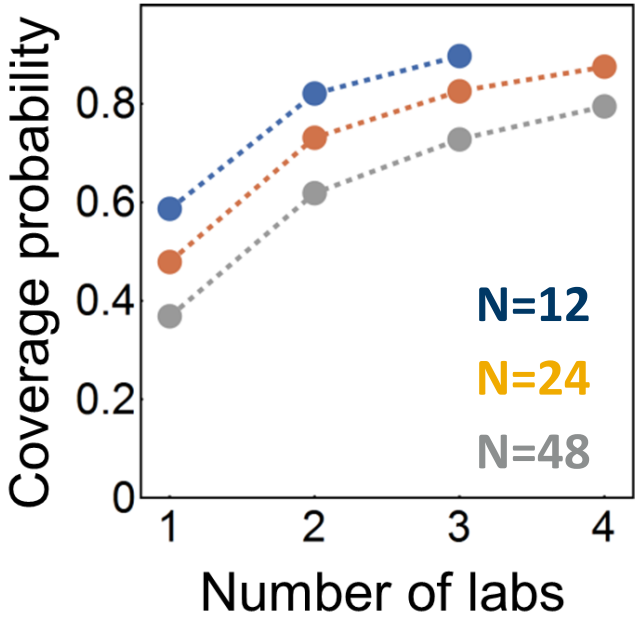


Stroke
Myocardial infarction
Breast cancer

- Calculated: coverage probability
- How often the ‘true’ effect is included within the 95% CI.

Voelkl *et al* 2018 PLOS Biology





Voelkl *et al* 2018 PLOS Biology

How can we practically introduce variation?

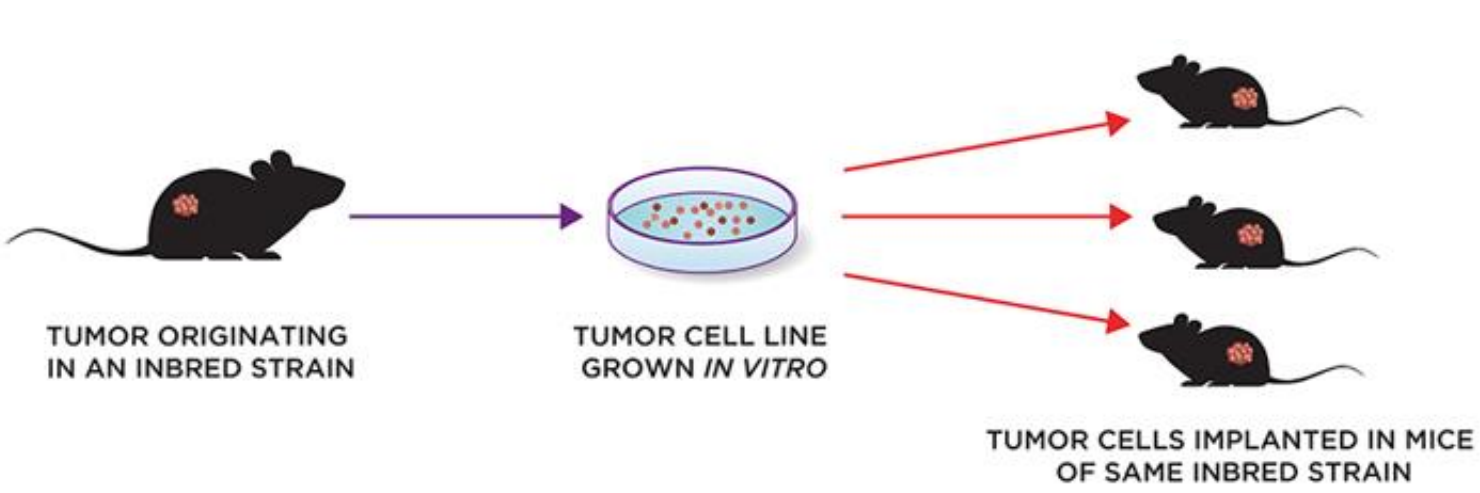
What factors have been consider?

- Time – am/pm
- Operator
- Environmental enrichment
- Genetic
- Sex
- Lab
- cage size
- microbiota

Many unknowns

- Which variable will matter when?
- How much variation do we need to introduce?

Case study 1: Syngeneic studies



vehicle

N=16

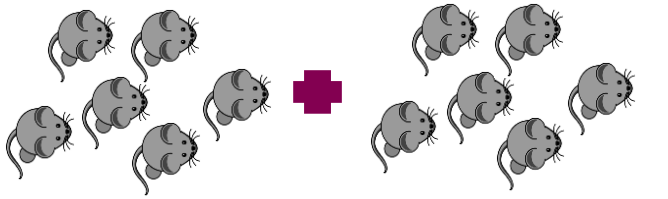
treatment

N=16

Monitor tumor volume

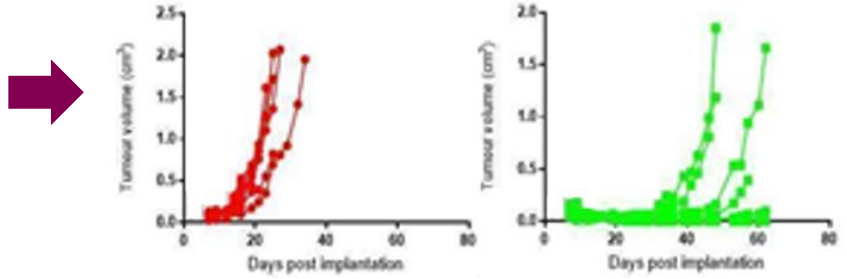
Batch as a block strategy

Batch 1

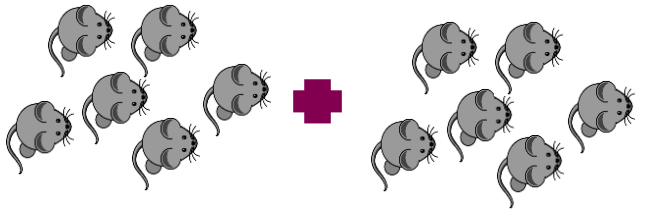


Vehicle

Treatment

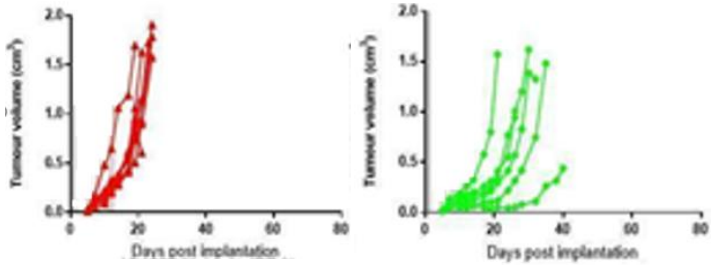


Batch 2

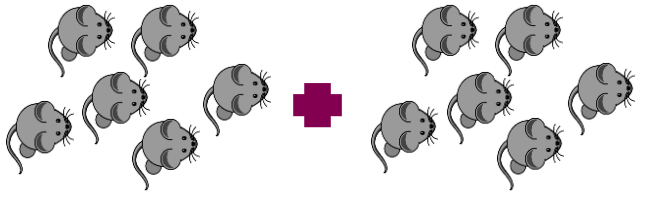


Vehicle

Treatment

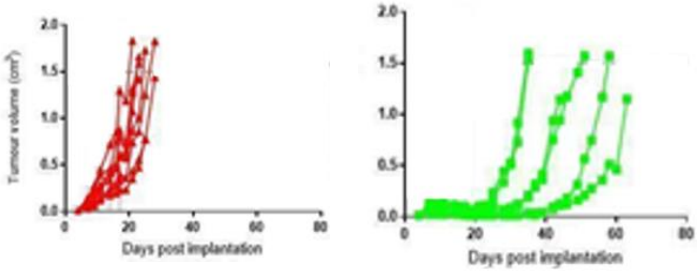


Batch 3



Vehicle

Treatment

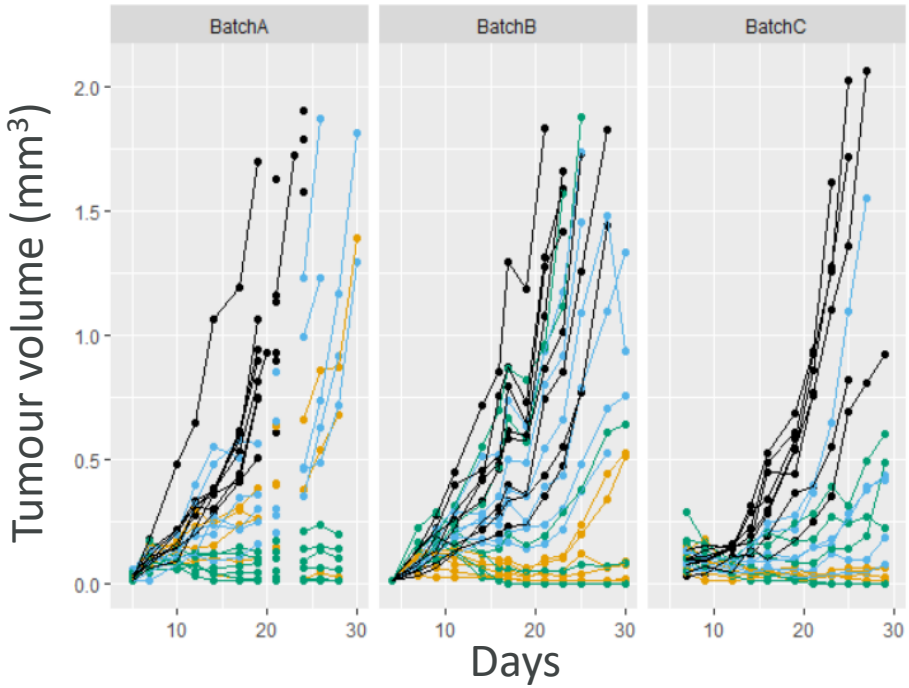


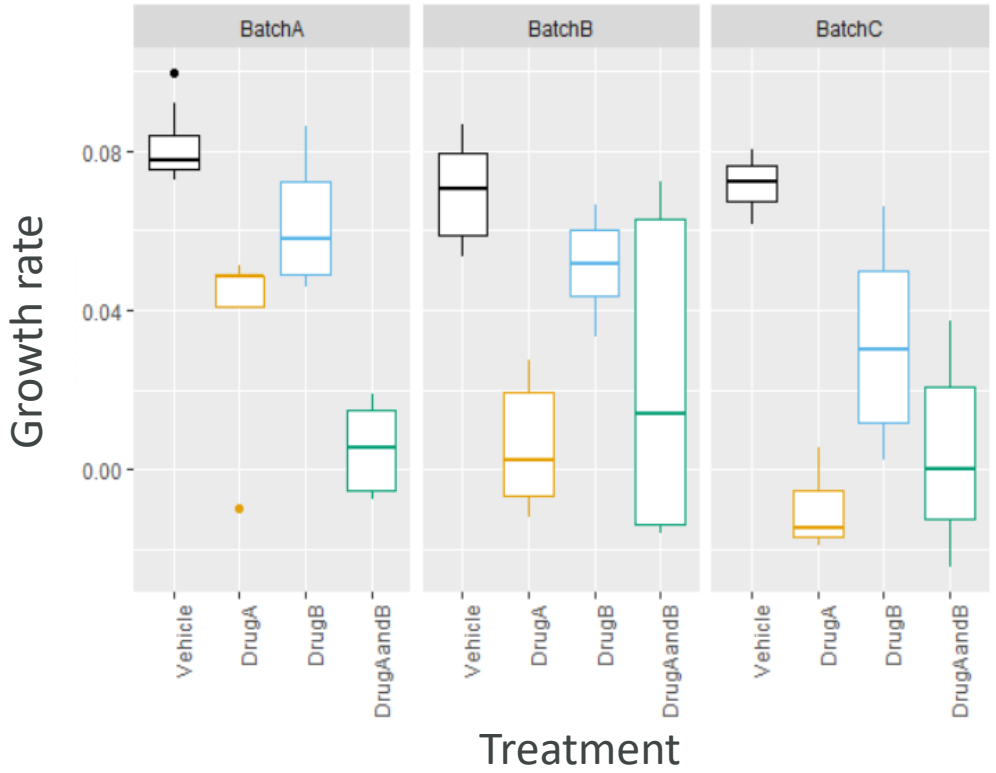
Rate of growth
Summary measure

Integrated data analysis

Estimate efficacy across three studies

Example multi-batch output



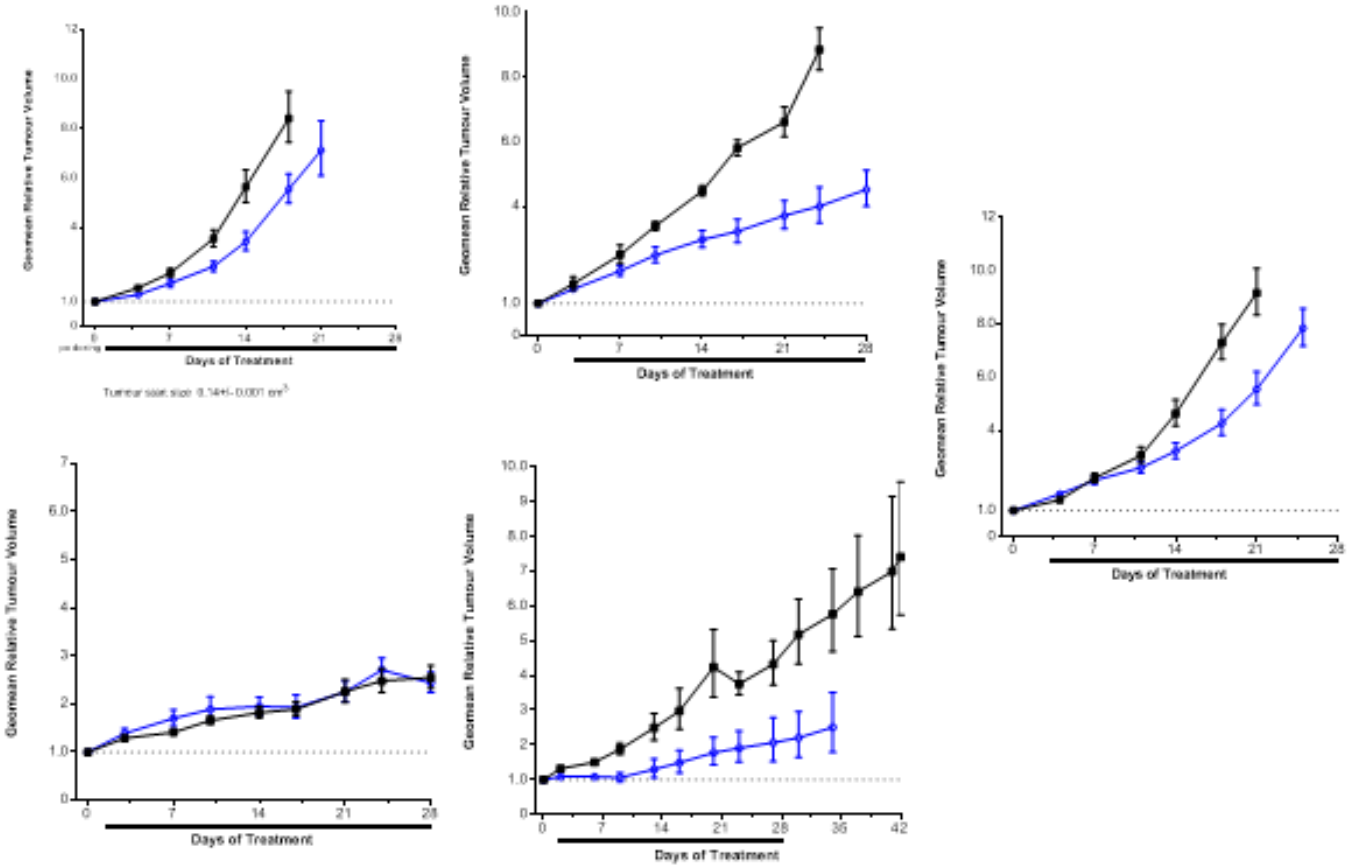


Compound	Estimate	SE	P value
DrugA	0.0640	0.0102	3.69e-10
DrugB	0.0260	0.0067	9.83e-5
DrugAandB	0.0635	0.0069	3.35e-20

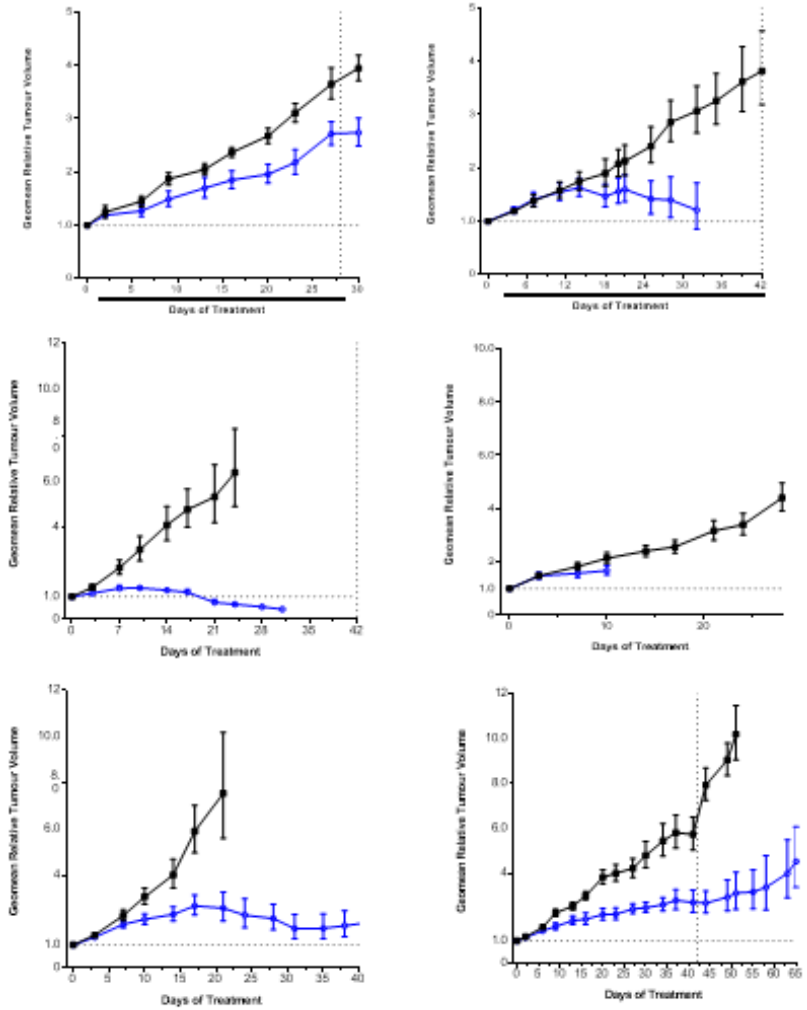
Karp *et al.* A multi-batch design to deliver robust estimates of efficacy and reduce animal use – a syngeneic tumour case study. *Sci Rep* **10,** 6178 (2020).

Case 2: Multi-model cancer studies

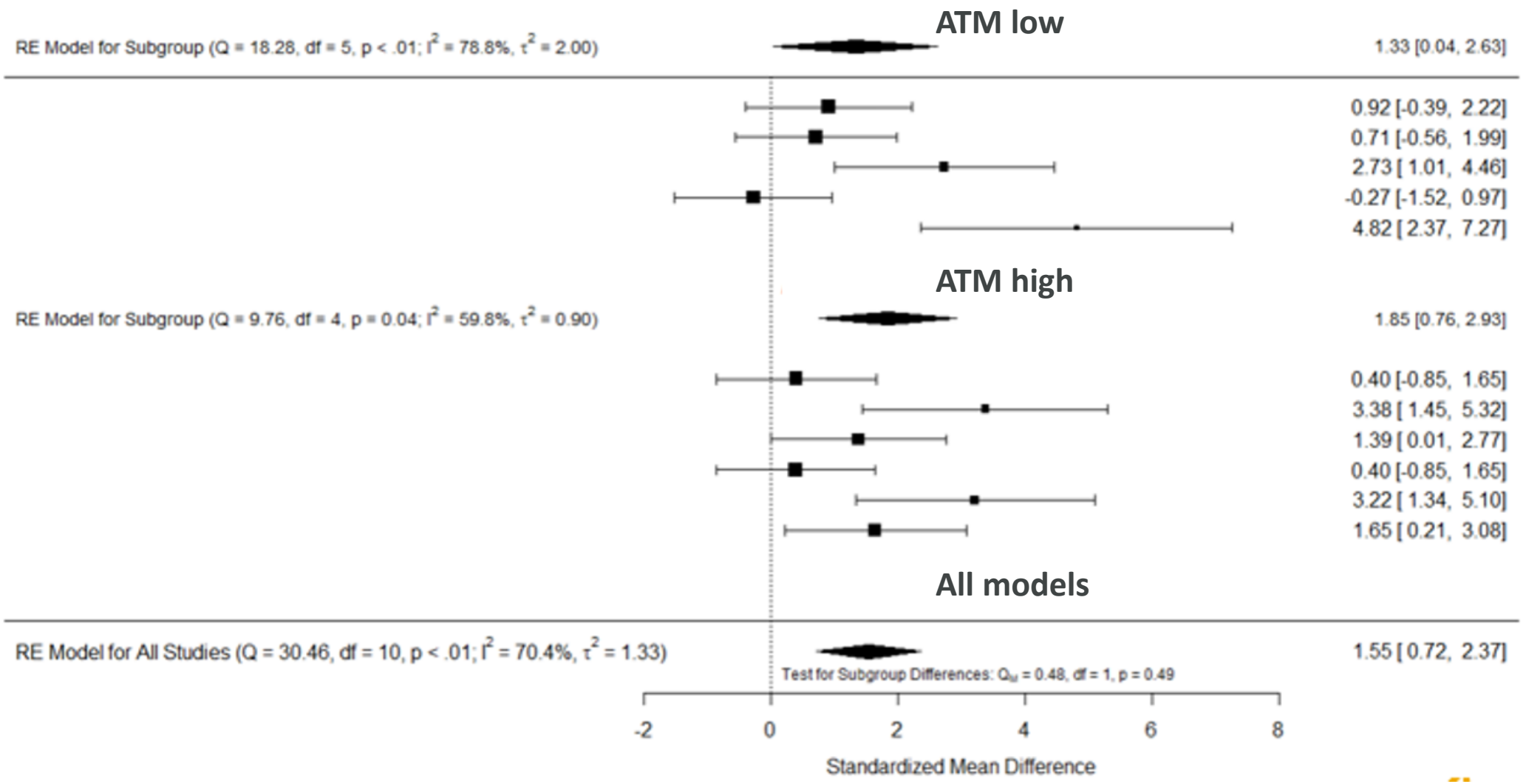
ATM high



ATM low



Meta-analysis of the growth rate



Conclusions

- Traditionally, there has been a call to standardise and study within a narrow window of testing space.
- We then extrapolate the findings.
  - This approach is questionable because of phenotypic plasticity and significant sex differences
  - Improving translation and reproducibility requires us to embrace variation.

- Sex is binary and is an easy first step to improve generalisability.
  - Scientists believe sex matters but frequently do not perceive it is doable. However, many of the barriers are misconceptions and fear of change.
- Even in a highly standardised environment there is unpredicted variation across all biological screens. Even within your own lab you can struggle to reproduce effects.
  - Block designs are the solution and embracing multiple batches will give confidence in reproducibility for your condition

Acknowledgement

AstraZeneca

Neil Reavey
Timo Haschler
Dominic Corkill

TEL AVIV UNIVERSITY

Ruth Heller
Shay Yaccoby
Yoav Benjamini

IMPC

EMBL-EBI

Jeremy Mason

UCL

Richard Mott
Judith Mank

SIRF working group

Name	Institute
Manuel Berdoy	University of Oxford
Jon Gledhill	Newcastle University
Lillian Hunt	Wellcome Trust
Maggy Jennings	RSPCA, Animals in Science Dept
Angela Kerton	The Learning Curve (Development) Ltd
Matt Leach	Comparative Biology Centre, Newcastle University
Jordi L. Tremoleda	Queen Mary University of London
Esther J. Pearl	The NC3Rs
Nathalie Percie du Sert	The NC3Rs
Benjamin Phillips	Data Sciences & Quantitative Biology, Discovery Sciences, R&D, AstraZeneca
Penny S Reynolds	University of Florida, USA
Kathy Ryder	Department of Health, Belfast
S. Clare Stanford	University College London
Sara Wells	The Mary Lyon Centre at MRC Harwell
Lucy Whitfield	OWL Vets Ltd

References

- Baker, M. "Reproducibility crisis." *nature* 533.26 (2016): 353-66.
- Becker, J.B, Prendergast BJ, Liang JW. Biol Sex Differ. (2016 Jul) 26;7:34. doi: 10.1186/s13293-016-0087-5.
- Beery AK. Current Op in Behav Sci. 23(2018): 143-149.
- Begley, C. Glenn, and Lee M. Ellis. "Raise standards for preclinical cancer research." *Nature* 483.7391 (2012): 531-533.
- Bwire, G.M. SN Compr Clin Med. 2020;2(7):874-876. doi: 10.1007/s42399-020-00341-w. Epub 2020 Jun 4.
- Doerre, A & Doblhammer, G. PLoS One. (2022) May 6;17(5):e0268119. doi: 10.1371/journal.pone.0268119.
- Crabbe, J.C *et al*. "Genetics of mouse behavior: interactions with laboratory environment." *Science* 284.5420 (1999): 1670-1672.
- Errington, Timothy M., et al. "Investigating the replicability of preclinical cancer biology." *Elife* 10 (2021): e71601.Garcia-Sifuentes Y, Maney DL. Elife. (2021) Nov 2;10:e70817. doi: 10.7554/eLife.70817.
- Gompers, A. Genderscilab, (2018). https://www.genderscilab.org/blog/three-years-in-sex-as-a-biological-variable-policy-in-practice-and-an-invitation-to-collaborate. Last accessed March 2024.
- Heidari, S. *et al*. Res Integr Peer Rev 1 (2016): 2.
- Karp NA, et al. Nature communications 8.1 (2017): 15475.
- Karp, NA., et al. "Impact of temporal variation on design and analysis of mouse knockout phenotyping studies." *PLoS One* 9.10 (2014): e111239.
- Karp NA, Reavey N. Br J Pharmacol. (2019) Nov;176(21):4107-4118. doi: 10.1111/bph.14539. Epub 2018 Dec 12.
- McCarthy, MM. Schizophrenia bulletin 41.5 (2015): 1016-1020. doi: 10.1093/schbul/sbv077.
- Mogil, JS. Nature Reviews Neuroscience 21.7 (2020): 353-365. doi:10.1038/s41583-020-0310-6
- MRC. Working Group on Sex in Experimental Design of Animal Research. Meeting Report 6 September 2021. Online. Available at: https://www.ukri.org/publications/sex-in-experimental-design-summary-report/. Last accessed March 2024.
- Phillips, B. *et al*. PLoS Biol. (2023) Jun 8;21(6):e3002129. doi: 10.1371/journal.pbio.3002129.
- Prinz, Florian. *et al*. "Believe it or not: how much can we rely on published data on potential drug targets?." *Nature reviews Drug discovery* 10.9 (2011): 712-712.
- Prendergast BJ, *et al*. Neurosci Biobehav Rev. (2014) Mar;40:1-5. doi: 10.1016/j.neubiorev.2014.01.001. Epub 2014 Jan 20.
- Shah, K. *et al*. American journal of physiology. Cell physiology vol. 306,1 (2014): C3-18. doi: 10.1152/ajpcell.00281.2013.
- Shansky RM. Science. (2019) May 31;364(6443):825-826. doi: 10.1126/science.aaw7570.
- World Health Organisation/ Gender and Health. Online. Available at: https://www.who.int/health-topics/gender#tab=tab_1. Last accessed March 2024.
- Woitowich NC, *et al*. Elife. (2020) Jun 9;9:e56344. doi: 10.7554/eLife.56344.
- Taylor, K.E. *et al*. Biology of sex differences 2 (2011): 1-7. doi: 10.1186/2042-6410-2-11
- Yoon, DY. *et al*. Surgery. (2014) Sep;156(3):508-16. doi: 10.1016/j.surg.2014.07.001.
- Voelkl, B. et al. "The standardization fallacy." *Nat Methods* 18.1 (2021): 5-7.
- Voelkl, B. et al. "Reproducibility of preclinical animal research improves with heterogeneity of study samples." *PLoS biology* 16.2 (2018): e2003693.