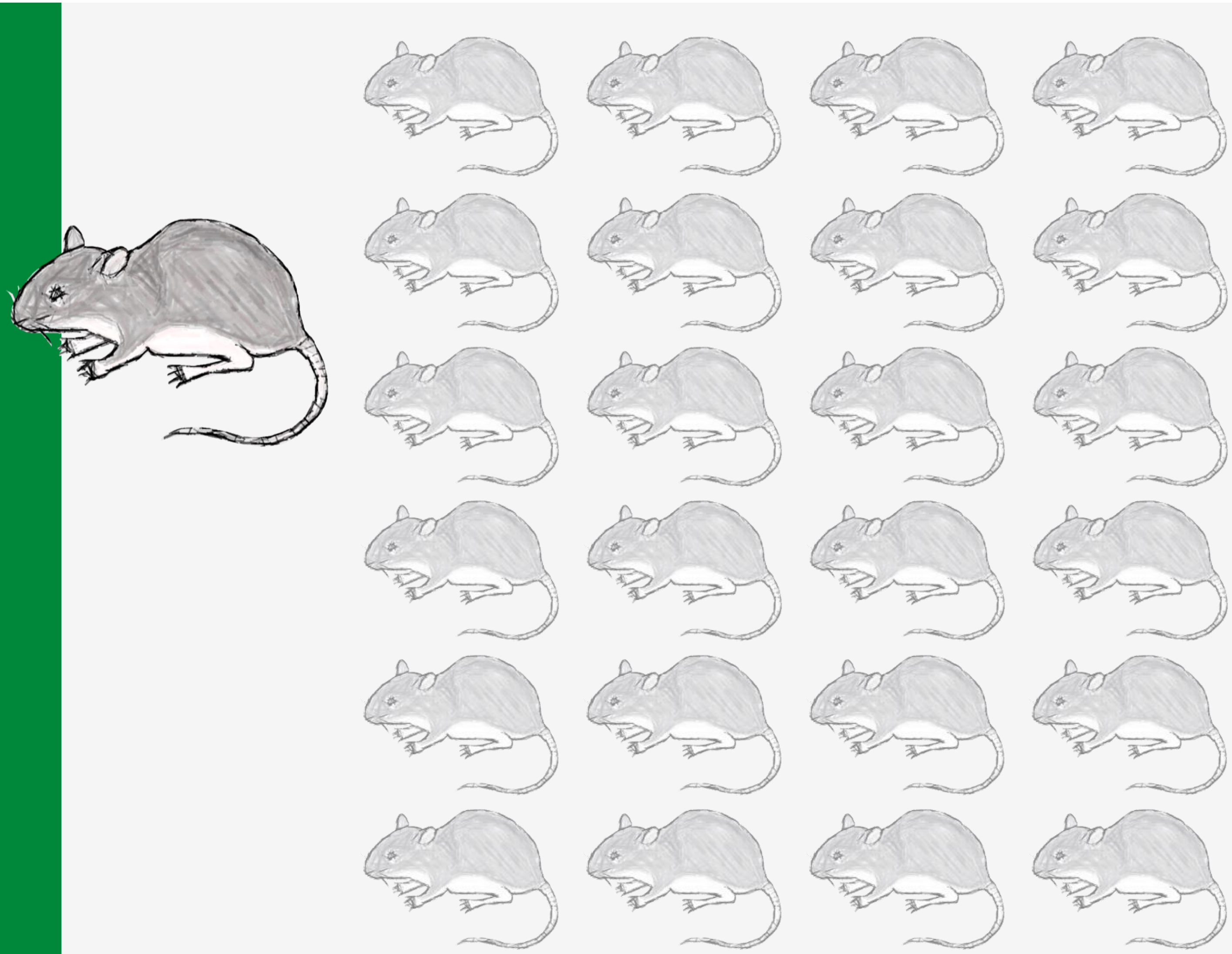
LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUTE FOR MEDICAL INFORMATION PROCESSING, BIOMETRY, AND EPIDEMIOLOGY (IBE)

Uncertain or biased input to sample size and power calculations in preclinical animal studies



Juliane Wilcke & Anne-Laure Boulesteix

NCS conference, Wiesbaden, 26 September 2024

squeak overview

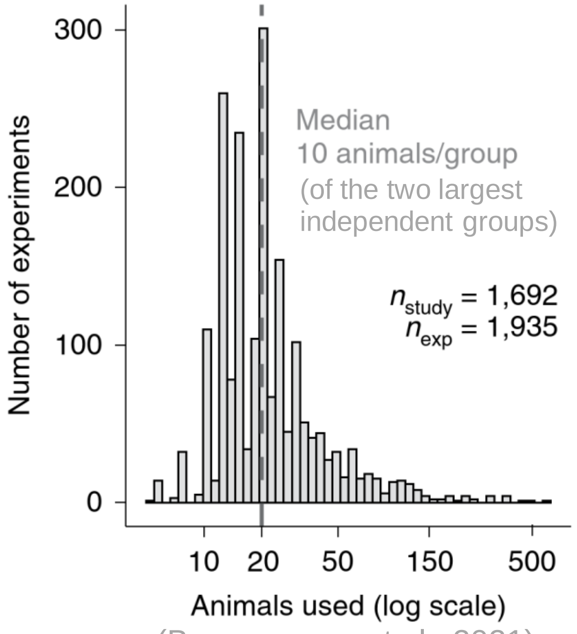
- Sample sizes in preclinical animal studies
- Determining *sample size* in preclinical animal studies – and alternatives
- Inputs to determinations of sample size or power
- Improvements of such determinations:
  Taking uncertainties and biases into account
- Using, studying and further developing the improvements

This work was supported by the Federal Ministry of Education and Research (BMBF) as part of its funding guideline **Confirmatory preclinical studies – quality in health research**, grant 01KC2005B E, and by the German Research Foundation (DFG), grant BO3139/7.

Introduction: Sample sizes in preclinical animal studies

- When planning a preclinical animal study, it is crucial to consider **the number of animals** and its consequences. Why?
  - Ethics approval
  - Other organisational aspects (housing, personnel, finances, duration)
  - Potentially misleading results! (low precision, low probability of finding true effects, low positive predictive value, overestimation of effect sizes vs. biologically or clinically irrelevant significant effects) [cf. ARRIVE 2.0]
- Samples in preclinical animal studies tend to be **small**. Why?
  - Exploratory research, limited resources, conventions, limits by authorities, lower variability between genetically almost identical animals, design optimisation for reduction
  - Be aware of lower external validity and pseudoreplicates E

→ Do not hinder knowledge acquisition *and*
avoid wasting resources, including animal lives,
by using samples that are too small or too large!



(Bonapersona et al., 2021)

Determining sample size in preclinical animal studies – and alternatives

Case A: Sample size can in fact be chosen (within practical and regulatory constraints) to obtain either

- **adequate statistical power** for the hypothesis test of interest or
- **adequate precision** in detecting the effect of interest.

Case B: Sample size is predetermined.

- Check the **expected power/precision or the minimum detectable effect size**.

→ Is it worth conducting the experiment??

Prospective power of papers assuming the median *published effect size*: (Bonapersona et al., 2021)

- Most papers (mode): 19% power
- 93.5% of papers have power < 50%

If not, consider changing the design of the experiment or the analysis plan. E

Inputs to determinations of sample size or power in preclinical animal research

- All of these sample size and power/precision calculations **require information on the effect of interest** (e.g., true effect size, variability, other relevant parameters) before data collection.
- Which **type of effect size** best fits the research question (cf. Lakens, 2022) or situation?
  - *Expected effect size* (→ power to allow detection of this or any larger effect size)
  - *Smallest effect size* of biological/clinical interest (for superiority/inferiority)
    [or largest effect size for non-inferiority/non-superiority and equivalence]
- What also needs to be known, but is **not the focus here**:
  - Statistical analysis plan (in line with the research question, hypotheses about primary outcome incl. distributions, study design, incl. the appropriate experimental unit)
  - Sample size or power calculation software and *method*
  - Alpha level (incl. sidedness of test, correction for multiple comparisons)
  - Desired power/precision or fixed sample size
  - Anticipated attrition rate / reserve animals

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Inputs to determinations of sample size or power in preclinical animal research

- There are **different types of sources** of this information,
  - which carry different amounts of **uncertainty**
  - and are prone to different **biases**:

Expert judgement:
- How reliable is it?
- Inflation of effect sizes in small samples, selection bias, publication bias

Published findings, including meta-analyses and field-specific effect size distributions**:**
- Level of evidence? Number of previous studies and how combined?
- Inflation of effect sizes in small samples, selection bias Ⓔ, publication bias

Heuristics (e.g., Cohen, 1988)**:**
- How valid are they? *Specific to content and method*
- Inflation of effect sizes in small samples, selection bias, publication bias

Own or other pilot data (should be preprocessed and analysed as similarly as possible/sensible to the planned trial)**:**
- Small or very small *n* in pilot data Ⓔ
- Inflation of effect sizes in small samples, selection bias

- In addition, it needs to be taken into account how relevant/applicable the information is to the study at hand (e.g., differences in design Ⓔ, disease model, measures, protocol, population, analyses, lab, batch, experimenter, etc. – even more so if transition from in vitro experiments).

Improvements of such determinations: Taking uncertainties and biases into account

In each case(!):

- Which **uncertainties**, how likely, how large, and how relevant for the sample size or power calculation?
  - Inflate the expected **nuisance variability** Ⓔ
  - May be incorporated in the form of **prior distributions**
  - Can be usefully explored in **sensitivity analyses**
- How to combine the different uncertainties for a single study (e.g., additively as a first approximation or rather ‘holistically’)?
- Further options:
  - Combining different sources of information if available
  - If not all data collected at once Ⓔ: re-examining sample size through interim analyses

Improvements of such determinations: Taking uncertainties and biases into account

In each case(!):

- Which **biases**, how likely, how large, and how relevant for the sample size or power calculation?
- Inflation of effect sizes in small samples, selection bias, and publication bias all call for **shrinking the expected effect size** (less clear for the smallest effect size of interest), but how and by how much?
  - Rules of thumb (mostly for the design of replication studies in psychology):
    - 2.5 times the sample size of the original study in order to have ~80% power to reject $d_{33\%}$ (Simonsohn, 2015)
    - Aiming for the lower end of the 60% CI around the reported effect size (Perugini et al., 2014)
    - Dividing the published effect size by 2 (Schönbrodt & Bollmann, 2016); using ~2/3 of exploratorily observed effect sizes in animal trials (Piper et al., 2022)
  - Using the most conservative instead of the median effect size E
  - Adjusting the desired power, e.g., 50% power for the smallest effect size of interest in confirmatory studies (Danziger et al., preprint)

←

Conclusion and Outlook: Using, studying and further developing the improvements

- Ask about each study:
  - Is it worth conducting?
  - What are likely **biases** and **uncertainties** in the input parameters to the sample size or power calculation and how should we best deal with them?
    → Shrinkage of effect sizes among others
- Investigate the different mitigation strategies in the context of preclinical animal studies

- More generally, pay attention to the many *Researcher Degrees of Freedom in Power Analyses and Sample Size Planning* (title of CEN2023 talk by Nicole Ellenbach) and document the chosen options! (esp. confirmatory research should be preregistered)

→ For the quality of research as well as the lives of animals and ultimately patients: Preclinical sample sizes need to

- Closely match the research question and be well justified, and additionally
- Be well reported according to the ARRIVE 2.0 guidelines (Percie du Sert et al., 2020)

References

Baker, D., Lidster, K., Sottomayor, A., & Amor, S. (2014). Two years later: Journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biology*, *12*(1), e1001756. https://doi.org/journal.pbio.1001756

Bonapersona, V., Hoijtink, H., Sarabdjitsingh, R., & Joëls, M. (2021). Increasing the statistical power of animal experiments with historical control data. *Nature Neuroscience*, *24*(4), 470–477. https://doi.org/10.1038/s41593-020-00792-3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Danziger, M., Collazo, A., Dirnagl, U., & Toelch, U. (preprint). Balancing sensitivity and specificity in preclinical research. *bioRxiv* 2022.01.17.476585. https://doi.org/10.1101/2022.01.17.476585

Gosselin, R. D. (2021). Insufficient transparency of statistical reporting in preclinical research: A scoping review. *Scientific Reports*, *11*(1), 3335. https://doi.org/10.1038/s41598-021-83006-5

Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, *8*(1). https://doi.org/10.1525/collabra.33267

Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., ... & Sena, E. S. (2015). Risk of bias in reports of in vivo research: a focus for improvement. *PLoS biology*, *13*(10), e1002273. https://doi.org/10.1371/journal.pbio.1002273

Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., ... & Würbel, H. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLoS Biology*, *18*(7), e3000410. https://doi.org/10.1371/journal.pbio.3000410

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard Power as a Protection Against Imprecise Power Estimates. *Perspectives on Psychological Science*, *9*(3), 319-332. https://doi.org/10.1177/1745691614528519

Piper, S., Zocholl, D., Toelch, U., Roehle, R., & Konietschke, F. (2022). User Guide for Biometric Planning of Animal Trials. *Zenodo.* https://doi.org/10.5281/zenodo.7359565

Schönbrodt, P. D. F., & Bollmann, S. (2024). *Advanced power analysis* [workshop slides]. Department of Psychology, LMU Munich. https://osf.io/d76gc

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science*, *26*(5), 559-569. https://doi.org/10.1177/0956797614567341

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Thank you!! Any thoughts, reactions or questions?

juliane.wilcke@ibe.med.uni-muenchen.de