Academic integrity
Where it may inadvertently be lacking

Martin C. Michel

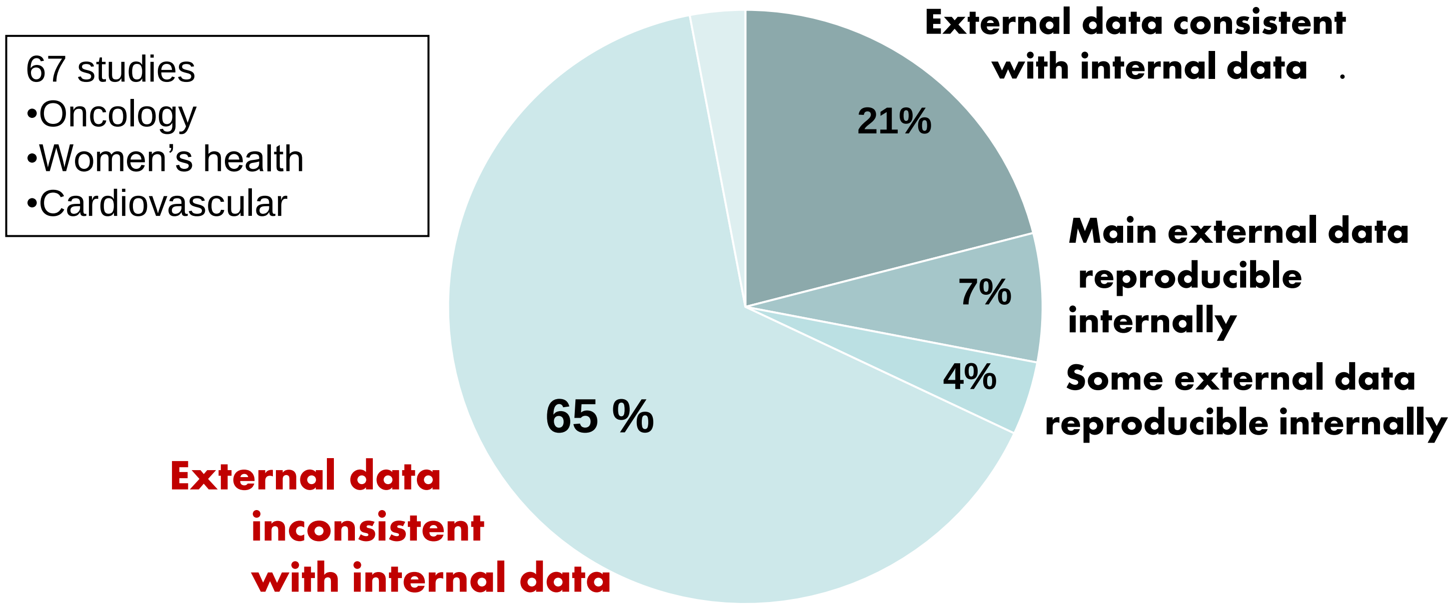
Dept. Pharmacology, Johannes Gutenberg University Mainz

Partnership for the Assessment and Accreditation of Scientific Practice

Work in the author lab funded by IMI as part of EQIPD consortium and by the US NIH

Innovative Medicines Initiative 2 Joint Undertaking (grant agreement no. 777364); this Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation program and EFPIA

Can you trust published data?



Poor replicability meanwhile found in several other studies in experimental medicine and many other disciplines including social sciences

Prinz et al. (2011) Nature Rev Drug Discov 10: 712-713

What are the root causes?

- Outright fraud minor component (intentionally falsified data)
- Lack of detail in Methods section makes replication difficult

But the big-3 apparently are:

- Biases in design, conduct, analysis and reporting of studies
- Low statistical power (sample sizes too small)
- Poor understanding of statistical concepts

Let’s focus on 2 examples were “good intentions” can get uncomfortably close to fraud

Randomness principle

- A p-value reports the probability of seeing an effect as large as observed, or larger, if the two samples had been selected randomly from populations with the same mean/median
  - Does not tell whether an observed effect is true or of relevant magnitude

- Only meaningful if all factors other than primary variable are randomly distributed among groups
  - P-values cannot be interpreted at face value if major bias exists (violation of randomness principle)

Motulsky (2014) Naunyn-Schmiedeberg's Arch Pharmacol 387: 1017-1023

Violation of randomness

- Violations (biases) can occur unconsciously or be investigator-induced
- Unconscious biases
  - Sampling error
  - Selection bias
  - Other biases
- Investigator-induced violations are also referred to as p-hacking
  - Includes HARKing

Motulsky (2014) Naunyn-Schmiedeberg‘s Arch Pharmacol 387: 1017-1023

P-hacking

- Various design choices may be fine if pre-specified
- Post-hoc changes in design, conduct, analysis and reporting introduce bias and violate randomness principle
- This makes resulting p-values difficult to interpret, irrelevant or even misleading
  - Bias for finding an effect even if it is not there
  - Even if effect true, trend for exaggerated effect size estimate

Motulsky (2014) Naunyn-Schmiedeberg‘s Arch Pharmacol 387: 1017-1023

P-hacking examples

- After n = 6 yielded p = 0.055, add 2 additional experiments
  - The new n = 8 is biased by the trend in n = 6 and no longer a random sample
  - Variation: Stop adding experiments as soon as $p < 0.05$
- Post-hoc decision to log-normalize data
  - Log normalization can be justified or even required when raw data are skewed and only get closer to a normal distribution on a log scale
- Post-hoc change of denominator
  - From fmol/mg protein to fmol/g wet weight
- Switch to a different statistical test
  - paired vs. unpaired test
- “Outlier” removal (attrition bias)

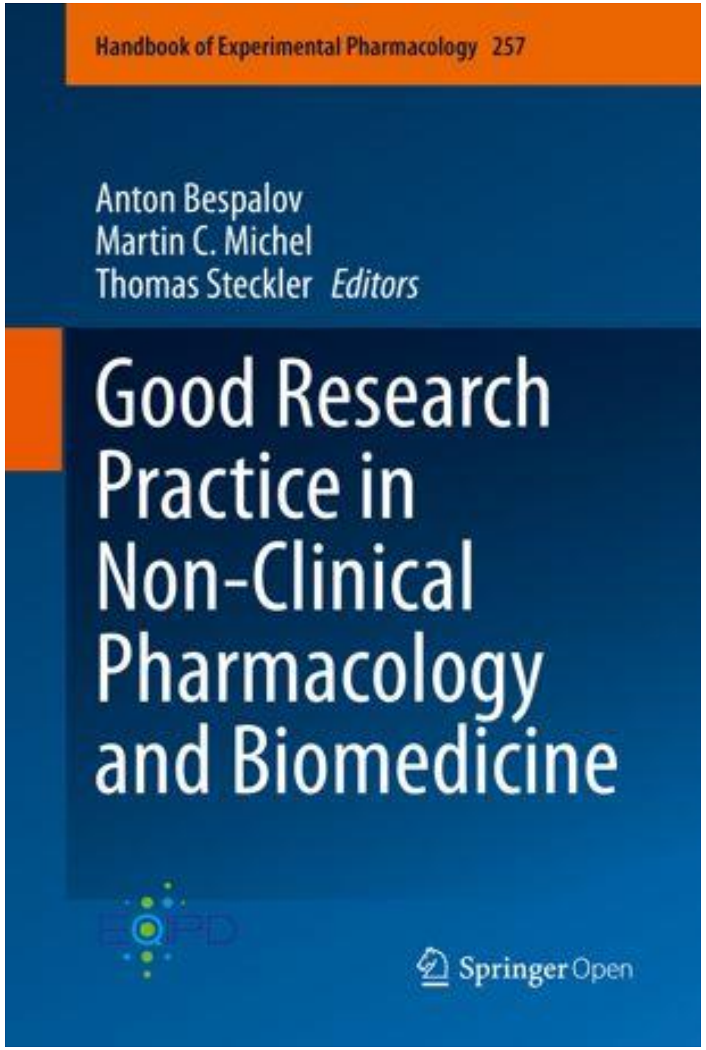
HARKing

Hypothesizing **A**fter **R**esults are **K**nown

- Redefining study aim after results have been seen
- Introduces bias into reporting
  - Focus on “positive” results in reporting
  - Making possible chance finding appear as study aim
- Ignores impact of low prior probability on False Discovery Rate
  - Increases probability for false positives
- HARKing becomes fraud if paper claims that a HARKed hypothesis had been pre-existing

Munafo et al. (2017) Nature Hum Behav 1: 0021

Open access book on reproducibility



Conclusions

- Biases at any level of study design, conduct, analysis and reporting violate the randomness principle
  - This makes p-values difficult to interpret, possibly misleading
- Pre-specification of all critical elements of study design, conduct and analysis, and randomization and blinding are key defenses against unintentional bias
- P-Hacking and HARKing are intentionally introduced biases
  - Often lead to misleading finding/conclusions or at least exaggerated effect sizes
  - Get uncomfortably close to fraud

DON‘T

JUST DO IT.