Sloppy science, selective reporting and the replicability crisis

Lex Bouter

20-25 minutes followed by discussion.
The next two slides will outline the ideas from which my presentation departs.
Breaches of Research Integrity can be harmful

- **Primary**
  - **truth** (validity of knowledge)
  - **trust** in science (credibility)

- **Secondary**
  - **waste** of resources and unethical to participants
  - **harm** to society, nature, individuals/patients

The secondary damage is due to the primary harm done to the validity of knowledge.
Research misconduct is typically thought to concern one of the three ‘deadly sins’ against research integrity: fabrication, falsification or plagiarism.

The fabrication and falsification (leaving out and changing data points) of data is clearly wrong.

Plagiarism differs from fabrication and falsification in the sense that it typically does not compromise the validity of knowledge, but it undermines the trust in science and in scientists.

Most forms of plagiarism and especially self-plagiarism are a QRP rather than RM.
These figures come from the highly cited meta-analysis of Daniele Fanelli.

A recent survey shows higher numbers for research misbehavior in economics (4% and 94%) → more honest reporting and/or more misbehavior? – My economic friends have no consensus on that.

Self-reported → likely underestimation of true prevalences.

Difficult to translate to % papers or research with both categories of misconduct

Self report of fabrication and falsification implies intention to deceive.

Self report of QRP concern uncertain intention to deceive.


Salient findings

- **Selective reporting, selective citation**, flaws in **research quality** and insufficient **supervision** lead the rankings
- **Fabrication + Falsification** rank high on impact on validity but only in the middle range on frequency *X* impact on validity
- **Plagiarism** items rank low both on impact on truth and on frequency *X* impact on validity – slightly higher on trust rankings

Putative causes of research misbehavior

SYSTEM of science

local research CLIMATE

level of INDIVIDUAL
Putative causes of research misbehavior

SYSTEM of science
- Organizational injustice of science system
- Likelihood of detection by reviewers

level of INDIVIDUAL
- Scientific norm subscription
- Perceived work pressure
- Dependence on external funding
Putative causes of research misbehavior

local research CLIMATE
- Scientific norm adherence in the group and discipline
- Perceived competition
- Received mentoring for survival (+) and RCR (-)
- Perceived social support at work
- Organizational injustice in group and institution
- Likelihood of detection by colleagues in the group
This slide shows – as a simplified summary of what has been explained – how things can go wrong.

In most disciplines the proportion of papers reporting positive results increases over time. Positive results are published and cited more often, and also get more media attention. This will probably increase the likelihood of getting grants and tenure. We have also some evidence that conflicts of interest and sponsor interests may lead to sloppy science or worse. QRP and RM can easily lead to (false) positive results.

This mechanism will lead to publication bias, selective reporting and selective citation. These phenomena will distort the truth in the published record and can explain the large replication difficulties some fields (e.g. preclinical research) experience.

There is strong evidence for some of the relations suggested in this slide, but no or only little evidence for most of them. We really need more solid empirical research to clarify how these things work. Gaining this knowledge is important for effectively fostering RCR and preventing QRP and RM.
This wonderful article comes from the faculty where Diederik Stapel was dean: never waste a good crisis.

The idea of Researcher Degrees of Freedom indicates that sloppy science offers a lot of room to get the findings you want.

Later in this presentation I will show that more transparency can limit the Researcher Degrees of Freedom.
There are thus many incentives for selective reporting.

Selective publication comes in two forms.

Non-publication and selective reporting of findings may be the single most important source of research waste.

And it is the Achilles’ heel of systematic reviews, because these rely on the published reports of research projects.

Furthermore, selective reporting is unethical in the sense that the efforts of patients participating in the study are wasted.
Let me just give one example, although it’s a quite spectacular one

This is the title of a alarming article in Nature a few years ago

The authors tried to replicate 53 widely cited high impact preclinical studies on potential new cancer treatments – suprise, surprise, they were all positive

If needed they even went into the original labs and tried to replicate the study there together with the original PIs

Selective reporting of animal studies is a huge problem, leading to an embarrassing lack of replicability.

The issue is that positive chance findings are difficult to reproduce.

There are many more examples of serious replication problems of preclinical studies.

### Table 1. Occurrence of non-reproducibility

<table>
<thead>
<tr>
<th>Field</th>
<th>Approach</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preclinical studies, general biology&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Researchers from Bayer HealthCare attempted to reproduce 67 original studies on potential drug targets in in-house projects</td>
<td>In 20% to 25% was relevant published data completely in line with in-house findings</td>
</tr>
<tr>
<td>Preclinical studies, oncology&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Amaen team attempted to replicate the findings of 53 'landmark' studies</td>
<td>Scientific findings were confirmed in 11% of cases</td>
</tr>
<tr>
<td>Preclinical studies, genetics&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Reproduce 18 microarray gene expression studies from raw data</td>
<td>11% of studies could be reproduced</td>
</tr>
<tr>
<td>Animal studies, neurology&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Retest of 9 potential drugs that had been reported to slow down disease in mouse model for ALS</td>
<td>0% was found to be beneficial</td>
</tr>
<tr>
<td>RCTs, clinical medicine&lt;sup&gt;5&lt;/sup&gt;</td>
<td>A retrospective analysis of high quality randomized clinical trials</td>
<td>85% of main findings could be reproduced</td>
</tr>
<tr>
<td>Experimental psychology&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Direct replications of 13 effects across 36 independent samples</td>
<td>77% of effects were reproduced consistently</td>
</tr>
<tr>
<td>Experimental psychology&lt;sup&gt;7&lt;/sup&gt;</td>
<td>The Open Science Collaboration attempted to independently replicate selected results from 100 studies in psychology</td>
<td>36% of the replications had significant results, compared to 97% of the original studies. The mean effect sizes were cut down by half</td>
</tr>
</tbody>
</table>

Uit: concept KNAW advies Replication studies
<table>
<thead>
<tr>
<th>Area</th>
<th>Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study methods</td>
<td>Weak experimental design/failure to control for biases</td>
</tr>
<tr>
<td></td>
<td>Low statistical power to detect effects</td>
</tr>
<tr>
<td></td>
<td>Technical/human errors in executing the study and poor quality control</td>
</tr>
<tr>
<td></td>
<td>Unknown variables</td>
</tr>
<tr>
<td></td>
<td>Analytical flexibility leading to p-hacking/data-dredging</td>
</tr>
<tr>
<td></td>
<td>Inappropriate statistical analyses</td>
</tr>
<tr>
<td></td>
<td>Failure to repeat experiments and assess reproducibility</td>
</tr>
<tr>
<td>Study reporting</td>
<td>Omitting negative or non-significant results (non-reporting and selective reporting) leading to publication bias</td>
</tr>
<tr>
<td></td>
<td>Lack of data and method sharing</td>
</tr>
<tr>
<td></td>
<td>Presenting <em>post hoc</em> hypotheses as tested hypotheses (HARKing)</td>
</tr>
<tr>
<td></td>
<td>Outcome switching</td>
</tr>
<tr>
<td></td>
<td>Lack of adequate peer review</td>
</tr>
<tr>
<td>Incentive system</td>
<td>Rewarding many and ‘high-impact’ publications</td>
</tr>
<tr>
<td></td>
<td>Rewarding positive and novel/’break-through’ findings</td>
</tr>
<tr>
<td></td>
<td>Highly competitive funding</td>
</tr>
<tr>
<td></td>
<td>Not rewarding open and reproducible practices</td>
</tr>
<tr>
<td></td>
<td>Belief that a rigorous research process hampers discovery</td>
</tr>
</tbody>
</table>

Uit: concept KNAP advies Replication studies
transparency

sincerity

clarity

fairness

honesty

openness

truth

accuracy

believability

sentiment

uprightness
Let me present my dream for the future.

Society must be able to trust scientists and scientists should have good reasons to trust their colleagues.

To deserve trust research needs to be open, honest and transparent.

The record should be complete and verifiable.

Without a study protocol that was made publicly available before the start of the data collection, it is very hard to judge whether all planned research questions are answered in the published report.

Equally a data-analysis plan that was publicly deposited before the data were collected is necessary to judge whether the statistical analysis was not partly data-driven.

Bouter LM. Open data is not enough to realize full transparency. J Clin Epidemiol 2016;
70: 256-7.
I firmly believe that data collected among volunteering participants of (clinical) research belong to the public domain – and also that this is an important principle outside of clinical research.

Of course, some months of embargo can be reasonable, proper acknowledgements should be made, and maybe the original investigators should be offered the opportunity to participate in the secondary analyses.

Breaches of privacy and misuse of the data ought to be prevented, and also obstruction and intimidation by excessive open records requests.

Open record laws can have nasty side effects – more uniform disclosure standards are needed.

Data sharing clearly is the norm, but how to do this is less obvious – although some very promising projects are underway.

And it is obvious that for secondary analyses the same rules for transparency apply, starting with a predefined study protocol.
Halpern, Mann - Open records versus harassment - Science 2015; 348 479
This is a very informative series of short articles on most aspects of transparency in research.

The downside is misuse of transparency, which also deserves our attention. Legitimate tools of scholarly exchange can be weaponized. – five double edged tools (use and abuse) Ten red flags for authors and ten red flags for their critics

Lewandowsky & Bishop – Don’t let transparency damage science – Nature 2016; 520; 459-61
How can we promote transparency?

→ re-design reward system

- No exclusive focus on citations and high IF journals
- Reward publication of protocols and ‘negative’ results
- And reward data sharing and replication
- As well as dissemination and application of findings

See e.g.: Ioannidis and Khoury - Assessing value in biomedical research - JAMA 2014; 312 483-4.
Transparency could be enforced by a concerted action of granting agencies, Institutional Review Boards (IRBs) and scientific journals.

Demanding a timely public deposition of study protocol, syntax and outcome reports as a condition for the last payment, for permission to perform the study and for accepting the paper for publication, respectively, would obviously be strong incentives to behave transparent.

Especially the impact of demands for transparency by funding agencies may be substantial.

A manifesto for reproducible science

Marcus R. Munafò,1,2*, Brian A. Nosek,3,4, Dorothy V. M. Bishop,5, Katherine S. Button,6, Christopher D. Chambers,7, Nathalie Percie du Sert,8, Uri Simonsohn,9, Eric-Jan Wagenmakers10, Jennifer J. Ware11 and John P. A. Ioannidis12,13,14

http://www.nature.com/articles/s41562-016-0021
Registered Reports

- Peer review of Introduction, Methods and Proposed Analysis
- In principle acceptance for publication on the basis of relevance of study question and soundness of methods
- Study is conducted
- Peer review of full manuscript
- Acceptance for publication if data collection and analysis and reporting are OK – irrespective of the findings

Registered Reports

Core idea: decision on publication should be based on relevance of study question and soundness of methods, NOT on the nature of the results

Chambers et al - Instead of playing the game its time to change the rules - registered reports - AIMS Neuroscience 2014; 1 4-17
Chambers - Ten reasons why journals must review manuscripts before results are known - Addiction 2015; 110 10-11
https://cos.io/our-services/registered-reports
Conclusions

- **Selective reporting** threatens validity and efficiency
- Leading to a **replication crisis**
- More **transparency** is urgently needed
- For that we need to re-design the **reward system**
- And let multiple **stakeholders** take action

No magic bullet – complex etiology – multi-mode intervention needed.

We have little sound knowledge about how to prevent RM + QRP and how to foster RCR.