Some reflections

Lex Bouter
Just two recent Nature headlines.
The topic draws attention, and rightly so.
Conclusion

1. Replication is crucial in science
2. Scientists are tempted to cut corners which contributes substantially to the ‘replication crisis’
3. We know how to make research more reproducible
4. Multiple stakeholders need to take action
Every baby knows the scientific method!

1. Make an observation.
2. Form a hypothesis.
3. Perform the experiment.
4. Analyze the data.
5. Report your findings.
6. Invite others to reproduce the results.
“Only when certain events recur in accordance with rules or regularities, as in the case of repeatable experiments, can our observations be tested—in principle—by anyone.... Only by such repetition can we convince ourselves that we are not dealing with a mere isolated ‘coincidence,’ but with events which, on account of their regularity and reproducibility, are in principle inter-subjectively testable.”

Karl Popper. The Logic of Scientific Discovery. London: Hutchison. 1959, p. 45
Bradford Hill is a prominent biostatistician who introduced the RCT after the 2nd World War and introduced the still influential criteria for causality that played a decisive role in the debate on smoking and lung cancer.

In 1959 he already explained why also negative results need to be published.
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 effectively help to get (false) positive results.

Negative findings are so unpopular that often these are not reported at all. This mechanism will lead to publication bias, selective reporting and selective citation. Especially small studies with positive outcomes will predominantly be chance findings. 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 and conclusions you want.

Please note: we’re talking about hypothesis testing research (confirmatory research), NOT about exploratory research. In the latter domain ‘anything goes’ as long as it’s clearly stated that exploration is at issue.
A manifesto for reproducible science

Marcus R. Munafò\textsuperscript{1,2*}, Brian A. Nosek\textsuperscript{3,4}, Dorothy V. M. Bishop\textsuperscript{5}, Katherine S. Button\textsuperscript{6}, Christopher D. Chambers\textsuperscript{7}, Nathalie Percie du Sert\textsuperscript{8}, Uri Simonsohn\textsuperscript{9}, Eric-Jan Wagenmakers\textsuperscript{10}, Jennifer J. Ware\textsuperscript{11} and John P. A. Ioannidis\textsuperscript{12,13,14}

http://www.nature.com/articles/s41562-016-0021
<table>
<thead>
<tr>
<th>Theme</th>
<th>Proposal</th>
<th>Examples of initiatives/potential solutions (extent of current adoption)</th>
<th>Stakeholder(s)</th>
</tr>
</thead>
</table>
| Methods                       | Protecting against cognitive biases | All of the initiatives listed below (** to ****)  
Blinding (***)  
Rigorous training in statistics and research methods for future researchers (*)  
Rigorous continuing education in statistics and methods for researchers (*) | J, F            |
| Independent methodological support | Involvement of methodologists in research (**)  
Independent oversight (*) |                                           | F              |
| Collaboration and team science | Multi-site studies/distributed data collection (*)  
Team-science consultlia (*) |                                           | I, F           |
| Reporting and dissemination   | Promoting study pre-registration | Registered Reports (*)  
Open Science Framework (*)  
Use of reporting checklists (**)  
Protocol checklists (*) | J, F            |
|                               | Protecting against conflicts of interest | Disclosure of conflicts of interest (****)  
Exclusion/limitation of financial and non-financial conflicts of interest (*) | J              |
<table>
<thead>
<tr>
<th>Theme</th>
<th>Proposal</th>
<th>Examples of initiatives/potential solutions</th>
<th>Stakeholder(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproducibility</td>
<td>Encouraging transparency and open science</td>
<td>Open data, materials, software and so on (* to **)</td>
<td>J, F, R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre-registration (**** for clinical trials, * for other studies)</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>Diversifying peer review</td>
<td>Preprints (* in biomedical/behavioural sciences, **** in physical sciences)</td>
<td>J</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pre- and post-publication peer review, for example, Publons, PubMed Commons (*)</td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td>Rewarding open and reproducible practices</td>
<td>Badges (*)</td>
<td>J, I, F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Registered Reports (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transparency and Openness Promotion guidelines (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Funding replication studies (*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Open science practices in hiring and promotion (*)</td>
<td></td>
</tr>
</tbody>
</table>

Estimated extent of current adoption: *, <5%; **, 5-30%; ***, 30-60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.
MAKING REPLICATION MAINSTREAM

Rolf A. Zwaan, Alexander Etz, Richard E. Lucas, & M. Brent Donnellan

concerns about replication

1. Context is too variable
2. Value of direct replications is limited
3. Direct replications are not feasible in some domains
4. Replications are a distraction
5. Replications affect reputations
6. No standard method to evaluate replication results

https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/making-replication-mainstream/2E3D8805BF34927A76B963C7BBE36AC7

https://www.researchgate.net/profile/Rolf_Zwaan/publication/320658906_MAKING_REPLICATION_MAINSTREAM/links/59fd9a2fa6fdcca1f299c9f0/MAKING-REPLICATION-MAINSTREAM.pdf
The current low levels of reproducibility are wasteful in the sense that resources were wasted on the production of these false leads in the scientific literature. It’s also unethical when animals or humans have been burdened for unpublished studies or for published false positive findings.

In theory the solution is easy and takes the form of ensuring that all research findings are published and the whole process is transparent, meaning that all steps can be checked and reconstructed. Studies need to be preregistered and a full protocol must be uploaded in a repository before the start of data collection. Similarly a data-analysis plan, syntaxes, data sets and full results need to be uploaded. Amendments and changes are possible but should always leave traces, thus enabling users to identify actions that were potentially data-driven. While ideally these elements of transparency are publicly accessible, there are many situations where delayed, conditional or incomplete access is indicated. But that does not detract from the principle of full transparency: even the process and outcomes of highly classified research for the defence industry should if necessary be made available for a thorough check by an investigation committee that is bound by confidentiality.

Bouter LM. Open data is not enough to realize full transparency. J Clin Epidemiol 2016; 70: 256-7.

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. Funding agencies should demand that research institutions fulfil the duties outlined above and that the study is executed with full transparency and according to the study proposal that was granted. Funding agencies can easily take measures that scientist and their institutions don’t like, as applicants are eager for grants and willing to sign demanding contracts. Increasingly

Conclusion

1. Replication is crucial in science
2. Scientists are tempted to cut corners which contributes substantially to the ‘replication crisis’
3. We know how to make research more reproducible
4. Multiple stakeholders need to take action