60TH ONLINE SEMINAR

Preregistration: The Panacea for Trustworthy and Useful Science?

Ulrich Dirnagl @dirnagl
WWW, September 21, 2021
Preregistration: The panacea for trustworthy and useful science?

(Helmholtz Open Science Online-Seminar 21.9.2021)
In pre-registration, researchers describe their hypotheses, methods, and analyses before a piece of research is conducted, in a way that can be externally verified. Registration prioritizes theory, analysis and methods over results.
[Houston]  We have a problem...
Inflation bias: *Almost every statistical test is significant (p≤0.05)*

<table>
<thead>
<tr>
<th>Study</th>
<th>ES (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature 2017</td>
<td>91.47 (86.40, 95.69)</td>
<td>25.46</td>
</tr>
<tr>
<td>Subtotal (I^2 = 46.2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science 2017</td>
<td>87.54 (79.88, 93.94)</td>
<td>22.81</td>
</tr>
<tr>
<td>Subtotal (I^2 = 76.4%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNAS 2017</td>
<td>96.77 (93.48, 99.11)</td>
<td>31.66</td>
</tr>
<tr>
<td>Subtotal (I^2 = 47.2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature 1997</td>
<td>98.53 (93.78, 100.00)</td>
<td>10.90</td>
</tr>
<tr>
<td>Subtotal (I^2 = 21.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science 1997</td>
<td>100.00 (99.96, 100.00)</td>
<td>5.89</td>
</tr>
<tr>
<td>Subtotal (I^2 = 7.5%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNAS 1997</td>
<td>90.19 (75.17, 99.45)</td>
<td>3.27</td>
</tr>
<tr>
<td>Subtotal (I^2 = 7.6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall (I^2 = 59.61%)</td>
<td>94.26 (91.79, 96.42)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Proportion of significant P values and 95% CIs

*Citation:* Cristea IA, Ioannidis JPA (2018) *P* values in display items are ubiquitous and almost invariably significant: A survey of top science journals. PLoS ONE 13(5): e0197440. [https://doi.org/10.1371/journal.pone.0197440](https://doi.org/10.1371/journal.pone.0197440)
"Only ten publications (2%) [of 525] reported no significant effects on infarct volume and only six (1.2%) did not report at least one significant finding."
90% of researchers surveyed by Nature think they are experiencing a 'reproducibility crisis'

**THE CAUSE**
The survey asked scientists what led to problems in reproducibility. More than 60% of respondents said that each of two factors — pressure to publish and selective reporting — always or often contributed. More than half pointed to insufficient replication in the lab, poor oversight or low statistical power.

**WHAT CAN BE DONE?**
Respondents were asked to rate 11 different approaches to improving reproducibility in science, and all got ringing endorsements. Nearly 90% — more than 1,000 people — ticked “More robust experimental design” “better statistics” and “better mentorship”.

*By Monya Baker*
The Adventures of Sherlock Holmes

The struggle of super minds in the crime of the century!

Caryl, Nigel, Ida, Alan
Rathbone, Bruce, Lupino, Marshall
Inflation bias: p-Hacking

Researchers try out several statistical analyses and/or data eligibility specifications and then selectively report those that produce significant results.

E.g. by
- recording many response variables and deciding which to report post analysis
- conducting analyses midway through experiments to decide whether to continue collecting data
- deciding whether to include or drop outliers postanalyses
- excluding, combining, or splitting treatment groups postanalysis
- including or excluding covariates postanalysis
- stopping data exploration if an analysis yields a significant p-value
- Performing multiple statistical tests without prespecification and reporting only the significant one(s)

"If you torture the data long enough, it will confess to anything"
Inflation bias: *Hypothesizing after the results are known (HARKING)*

An Agenda for Purely Confirmatory Research

Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, Han L. J. van der Maas, and Rogier A. Kievit
University of Amsterdam, The Netherlands

Perspectives on Psychological Science
7(5) 632–638
©The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1756554812463078
A hiking trip through the garden of forking paths...

... i.e. an exploratory biomedical research project

Andrew Gelman† and Eric Loken†

Undisclosed flexibility in data collection and analysis – researcher’s degree of freedom

Given the small samples and sensible but unexplored experimental approaches, other data patterns could easily occur by chance, which would naturally lead to different data analyses and inferences supporting alternative research hypotheses

“What we are saying is that the evidence in these research papers is not as strong as stated” (Gelman and Loken)

Preregistration of study protocols

- Limits unwarranted and/or undisclosed researcher’s degrees of freedom
- Prevents ‘outcome switching’
- Prevents HARKING
- Provides scooping protection
- Reduces publication bias
- Distinguishes between exploratory/discovery and knowledge claiming / confirmatory research
- ...
Since the launch of the clinicaltrials.gov registry in 2000, which forced researchers to preregister their methods and outcome measures, the percentage of large heart-disease clinical trials reporting significant positive results plummeted from 57% to a mere 8%.
Clinical trials need to be registered to be published in major journals (Consort)
Summary results need to be deposited within 12 month after study completion (WHO, EU)

https://clinicaltrials.gov/
https://www.clinicaltrialsregister.eu/
https://www.drks.de/
(Pre) Registration of non-clinical studies

All purpose registries (not reviewed)

https://osf.io/
https://aspredicted.org/

Animal study registries (ASR) (not reviewed)

German Centre for the Protection of Laboratory Animals
https://www.animalstudyregistry.org

Preclinicaltrials.eu https://preclinicaltrials.eu/

Timestamp servers / Blockchain (not reviewed)

e.g. https://github.com/decred/dcrtimegui

Registered reports (Elife, PlosBiol, F1000Res etc.) (reviewed!)

BIH QUEST
Center for Responsible Research
Registered reports

https://www.cos.io/rr
Preregistration: Main stream?

Trust in science would be improved by study pre-registration

Chris Chambers, Marcus Munafo and more than 80 signatories

https://www.theguardian.com/science/blog/2013/jun/05/trust-in-science-study-pre-registration
Ernst Lubitsch's
TROUBLE IN PARADISE
Miriam Hopkins with Kay Francis
Herbert Marshall
Preregistration: Science in chains?

Loss of exploration, creativity, flexibility which are the fundament of science.

Pre-registration would put science in chains

The pre-registration of study designs must be resisted, says Sophie Scott

‘After all, even Newton sometimes employed dubious methodologies. His celebrated physical laws were supported by data, but history tends to overlook his equally enthusiastic pursuit of alchemy, which swam in a sea of null results.’

https://www.timeshighereducation.com/comment/opinion/pre-registration-would-put-science-in-chains/2005954.article
Preregistration: Performative but not effective?

Beware performative reproducibility

Well-meant changes to improve science could become empty gestures unless underlying values change.

https://www.nature.com/articles/d41586-021-01824-z

COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time

Ben Goldacre1*, Henry Drysdale1, Aaron Dale1, Ioan Milosevic1, Eirion Slade1, Philip Hartley1, Cicely Marston2, Anna Powell-Smith1, Carl Heneghan1 and Kamal R. Mahtani1

Abstract

**Background:** Discrepancies between pre-specified and reported outcomes are an important source of bias in trials. Despite legislation, guidelines and public commitments on correct reporting from journals, outcome misreporting continues to be prevalent. We aimed to document the extent of misreporting, establish whether it was possible to publish correction letters on all misreported trials as they were published, and monitor responses from editors and trialists to understand why outcome misreporting persists despite public commitments to address it.

**Methods:** We identified five high-impact journals endorsing Consolidated Standards of Reporting Trials (CONSORT) (New England Journal of Medicine, The Lancet, Journal of the American Medical Association, British Medical Journal, and Annals of Internal Medicine) and assessed all trials over a six-week period to identify every correctly and incorrectly reported outcome, comparing published reports against published protocols or registry entries, using CONSORT as the gold standard. A correction letter describing all discrepancies was submitted to the journal for all misreported trials, and detailed coding sheets were shared publicly. The proportion of letters published and delay to publication were assessed over 12 months of follow-up. Correspondence received from journals and authors was documented and themes were extracted.

**Results:** Sixty-seven trials were assessed in total. Outcome reporting was poor overall and there was wide variation between journals in pre-specified primary outcomes (mean: 76% correctly reported; journal range: 25–96%), secondary outcomes (mean: 55%, range: 31–72%), and number of undeclared additional outcomes per trial (mean: 54, range: 29–83). Fifty-eight trials had discrepancies requiring a correction letter (87%, journal range 67–100%). Twenty-three letters were published (40%) with extensive variation between journals (range: 0–100%). Where letters were published, there were delays (median 99 days, range 0–257 days). Twenty-nine studies had a pre-trial protocol publicly available (43%, range 0–86%). Qualitative analysis demonstrated extensive misunderstandings among journal editors about correct outcome reporting and CONSORT. Some journals did not engage positively when provided correspondence that identified misreporting; we identified possible breaches of ethics and publishing guidelines.

(Continued on next page)

Preregistration: Inference fallacies?

The illusion of transparency
'Under present systems of preregistration, there is still substantial room for selective reporting and researchers’ degrees of freedom'

The illusion of robustness
'The fact that a particular finding was anticipated under a preregistered protocol does not mean that one can reasonably expect to observe it again and again; it only means that one is likely to observe it under the specific conditions of the preregistered protocol.’

Reproducible but pointless science
‘To elevate the quality of [...] science, it is not sufficient to encourage the preregistration of [...] studies as a means to enhance their reproducibility.’

Does preregistration interfere with claiming intellectual property?
Citation: Dirnagl U (2020) Preregistration of exploratory research: Learning from the golden age of discovery. PLoS Biol 18(3): e3000690. https://doi.org/10.1371/journal.pbio.3000690

PERSPECTIVE
Preregistration of exploratory research: Learning from the golden age of discovery

Ulrich Dirnagl
* QUEST Center for Transforming Biomedical Research, Berlin Institute of Health, Berlin, Germany

https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000690
Exploration/Discovery vs. Confirmatory (knowledge claiming) research

Exploration: Generates hypotheses and does not lead to a formal knowledge claim.

Hypothesis testing / Confirmatory / Knowledge claiming experiment: A clear, predefined hypothesis, including a clear predefined primary outcome measure to test the hypothesis and a predefined and appropriate statistical test. The proposed sample size should be stated, along with a justification based on the statistical power to detect a biologically important effect.

A given study can involve hypothesis-testing and exploratory parts, for instance by defining one primary endpoint (hypothesis-testing), with all other measured endpoints being exploratory.

There is a one-way street between confirmatory and exploratory experiments: if you find interesting results which contradict your hypothesis, a confirmatory experiment can turn into an exploratory experiment. However, an exploratory experiment can never become confirmatory.
Exploration/ vs. confirmation

Exploration: Generates hypotheses and does not lead to a formal knowledge claim.

Hypothesis testing / Confirmatory / Knowledge claiming experiment: A clear, predefined hypothesis, including a clear predefined primary outcome measure to test the hypothesis and a predefined and appropriate statistical test. The proposed sample size should be stated, along with a justification based on the statistical power to detect a biologically important effect.

A given study can involve hypothesis-testing and exploratory parts, for instance by defining one primary endpoint (hypothesis-testing), with all other measured endpoints being exploratory.

There is a one-way street between confirmatory and exploratory experiments: if you find interesting results which contradict your hypothesis, a confirmatory experiment can turn into an exploratory experiment. However, an exploratory experiment can never become confirmatory.
Conclusions

Preregistration and Registered report **can (but not necessarily or automatically)**

- limit unwarranted/undisclosed 'researcher’s degrees of freedom’
- prevent 'outcome switching’
- prevent HARKING
- provide scooping protection
- reduce publication bias
- do not put 'science in chains’
- …

Open issues: RRs overburdening peer review? Preregistration of exploratory research? Inference fallacies?
To infinity, and beyond!

When carmakers hack brains

You just see the "youtube video". Horribly cut sequences of busy young scientists in high tech laboratories waving lab coats, nervously fiddling guys are soldering electronic circuits and slamming into cool boxes, we are really in a video game. But through an animated brain of a sighted mouse cells. And in between all this, we stage at the California Academy of Sciences, car and rocket manufacturers like Elon Musk announce their latest scientific find. The symptoms of the human brain as artificial intelligence (AI). This is the time plan for now, which does not involve much manipulation to Mars, but the roadmap to a revolutionary near Mars lane interface (BMI), designed and manufactured by his company SpaceX. You may have guessed, this has caused a tremendous media hype all over the world. The world is the press and on the net too: "Mars at the end of the street", but the announcement is a breakthrough that there must be something to it. The news could be called "But can we be dangerous for mankind? Do we need a new wave of stuff like this?"

"No, I don't think so. Because this is a huge hype. And maybe a lot of fake news. But it could also be done. It is possible."

Yet, it will take a lot longer until we can use BMI to extend our thoughts and sense new content to.
Contact

- Mail – open-science@helmholtz.de
- Website – https://os.helmholtz.de
- Open Science Newsletter
- Mailing list (only for members of Helmholtz) – Helmholtz Open Science Professionals
- Twitter – @helmholtz_os
- LinkedIn – Helmholtz Open Science Office
Thank you for your interest!

Prof. Dr. Ulrich Dirnagl
https://orcid.org/0000-0003-0755-6119

Lea Maria Ferguson, Helmholtz Open Science Office
https://orcid.org/0000-0002-7060-3670

Antonia C. Schrader, Helmholtz Open Science Office
https://orcid.org/0000-0001-7080-634X

Nina Leonie Weisweiler, Helmholtz Open Science Office
https://orcid.org/0000-0001-6967-9443

All texts in this presentation, except citations, are licenced under Attribution 4.0 International (CC BY 4.0):
https://creativecommons.org/licenses/by/4.0/deed.de