Questionable research practices: Perceptions, prevalence and potential causes

Jesper W. Schneider
Centre for Studies in Research and Research Policy, Aarhus University
jws@ps.au.dk
Outline

• Introduction to the concepts of:
  • “responsible conduct of research” and “questionable research practices”
• Describe the recent history of “questionable research practices”, their link to the so-called “reproducibility crisis” and the current focus upon them
• Discuss the perceived effects of QRPs on the validity of knowledge claims and the “trust” among researches
• I will then critically discuss what we know about QRPs, where they are, how prevalent they are, what their causes may be and how we perceive them
• In the final part of the talk I will present our PRINT research project where we are investigating perceptions, prevalence and causes of QRPs across all fields of research
• As we are in the middle of our analyses I am only able to show some tentative findings, but we do have some important insights already
Research integrity

- Good research practices are based on fundamental principles of research integrity
- They guide researchers in their work and their engagement with the practical, ethical and intellectual challenges inherent in research

**Principles:**
- **Reliability** in ensuring the quality of research, reflected in the design, the methodology, the analysis and the use of resources
- **Honesty** in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair, full and unbiased way
- **Respect** for colleagues, research participants, society, ecosystems, cultural heritage and the environment
- **Accountability** for the research from idea to publication, for its management and organization, for training, supervision and mentoring, and for its wider impacts
Research integrity

We have a national Danish Code of Conduct as well
All universities are expected to implement it
Many funders promote it

It draws heavily on the so-called
Singapore Statement on Research Integrity

What about the UK?
A national code?
Do the funders require compliance?
Research integrity

• These notions of integrity originate with committee work from the National Academy of Sciences and the establishment of the Office of Research Integrity in 1992 in the US

• Demand for such an office from Congress due to several spectacular cases of research misconduct
Research integrity

• US Committee’s synthesis yielded six core values that, in their view, shape the norms of research and the practices that uphold integrity
  • Objectivity
  • Honesty
  • Openness
  • Accountability
  • Fairness
  • Stewardship

Thinking is governed by conditions in the natural, engineering and medical sciences
Research behavior: “Shades of grey”
Responsible conduct of research

Responsible conduct of research (RCR)
- ideal behaviour

Research misconduct (RM)
- Worst behaviour

'Shades of gray' in research behaviour
Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Research misconduct (RM)
- Worst behaviour

The practice of efficient production of relevant, valid, transparent, reliable and “reproducible” scientific knowledge

The (ideal) responsible researcher is conceptualized as: objective, meticulous, sceptical, rational, and not subject to external incentives such as prestige or social pressure.
Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Research misconduct (RM)
- Worst behaviour

- Fabrication
- Falsification
- Plagiarism

The practice of efficient production of relevant, valid, transparent, reliable and “reproducible” scientific knowledge

The (ideal) responsible researcher is conceptualized as: objective, meticulous, sceptical, rational, and not subject to external incentives such as prestige or social pressure

Misconduct is formally defined as three types of condemned behaviours: fabrication, falsification, and plagiarism (FFP)
Adopted by the U.S Federal government in 2000

©jws@ps.au.dk, ncrm2019
Responsible conduct of research

Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Research misconduct (RM)
- Worst behaviour

Represents the ideal standard
institutions and individuals
endeavor to meet

Encompasses practices
everyone agrees
should be avoided
Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Research misconduct (RM)
- Worst behaviour

Shades of gray' in research behaviour

Represents the ideal standard institutions and individuals endeavor to meet

Encompasses practices everyone agrees should be avoided

Research behavior viewed from the perspective of moral principles

Research behavior viewed from the perspective of professional standards
Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Research misconduct (RM)
- Worst behaviour

Represents the ideal standard institutions and individuals endeavor to meet

Encompasses practices everyone agrees should be avoided

'Shades of gray' in research behaviour
Responsible conduct of research

<table>
<thead>
<tr>
<th>Responsible conduct of research (RCR)</th>
<th>Questionable Research Practices (QRP)</th>
<th>Research misconduct (RCR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ideal behaviour</td>
<td>- the grey zone, a large zone of</td>
<td>- Worst behaviour</td>
</tr>
<tr>
<td>- good research practice with high</td>
<td>deviations from RCR that do not</td>
<td></td>
</tr>
<tr>
<td>integrity</td>
<td>fall under the established definition of RM</td>
<td></td>
</tr>
</tbody>
</table>

“QRPs are actions that violate traditional values of the research enterprise and that may be detrimental to the research process”

“... there is at present [1992] neither broad agreement as to the seriousness of these actions nor any consensus on standards for behavior in such matters”

“QRPs do not directly damage the integrity of the research process and thus do not meet the panel's criteria for inclusion in the definition of misconduct in science”
Responsible conduct of research

Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Questionable Research Practices (QRP)
- the grey zone, a large zone of deviations from RCR that do not fall under the established definition of RM

Research misconduct (RCR)
- Worst behaviour

"... this panel has identified some of these practices [QRPs] as not questionable at all but as clear violations of the fundamental tenets of research".

"This committee believes that many of the practices that up to now have been considered questionable research practices, as well as damaging behaviors by research institutions, sponsors, or journals, should be considered detrimental research practices (DRPs)"
Responsible conduct of research

Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Questionable Research Practices (QRP)
- the grey zone, a large zone of deviations from RCR that do not fall under the established definition of Research misconduct (RCR)
- Worst behaviour

Demarcation between QRP, DRP and RM is difficult to establish – there is no consensus

Detrimental Research Practices (DRP)
- Fabrication
- Falsification
- Plagiarism
Responsible conduct of research

Responsible conduct of research (RCR)
- ideal behaviour
- good research practice with high integrity

Questionable Research Practices (QRP)
- the grey zone, a large zone of deviations from RCR that do not fall under the established definition of RM

Research misconduct (RCM)
- Worst behaviour

Detrimental Research Practices (DRP)
- Fabrication
- Falsification
- Plagiarism

Wansink: “Massively misleading statistical analysis that falls short of falsification (DRM)”
- Eventually treated as misconduct

Demarcation between QRP, DRP and RM is difficult to establish – there is no consensus
So what happened?
A sensational misconduct case
Attention to a “problem”

Use of "questionable research practice" in TI/AB

Publication bias is not caused by "file drawing"

Most failed studies are indeed published, masqueraded as a success by use of “researcher-degrees-of-freedom and QRPs such as p-hacking!"
Attention to an old “problem”

A problem with the use of statistical significance tests in small N (human) experimental studies

©jws@ps.au.dk, ncrm2019
General claims of a “reproducibility crisis”?


- 36 out of 100 studies replicated
- Mean effect size in replications was half the original size

"Statistical crisis in science"

The Statistical Crisis in Science

Data-dependent analysis—a “garden of forking paths”—explains why many statistically significant comparisons don’t hold up.

Andrew Gelman and Eric Loken

"Researcher-degrees-of-freedom"

"Selection on significance"

"Decline effect"
"Statistical crisis in science"

The Statistical Crisis in Science

The ASA’s Statement on p-Values: Context, Process, and Purpose

In February 2014, George Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College, posed these questions to an ASA discussion forum:

Q: Why do so many colleges and grad schools teach \( p = 0.05 \)?
A: Because that’s still what the scientific community and journal editors use.
Q: Why do so many people still use \( p = 0.05 \)?
A: Because that’s what they were taught in college or grad school.

Cobb’s concern was a long-worrisome circularity in the sociology of science based on the use of bright lines such as \( p < 0.05 \): “We teach it because it’s what we do; we do it because it’s what”

2014) and a statement on risk-limiting post-election audits (American Statistical Association 2010). However, these were truly policy-related statements. The VAM statement addressed a key educational policy issue, acknowledging the complexity of the issues involved, citing limitations of VAMs as effective performance models, and urging that they be developed and interpreted with the involvement of statisticians. The statement on election auditing was also in response to a major but specific policy issue (close elections in 2008), and said that statistically based election audits should become a routine part of election processes.

"Selection on significance"

“Decline effect”
So clearly something has happened the last 6-8 years!

- Integrity attention has shifted towards questionable research practices
Attention has shifted towards questionable research practices

• Clearly, focus has been on practices related to quantitative data analyses
  • (e.g. confirmatory hypothesis testing studies, significance tests)
• Likewise, the “attention” is clearly restricted to certain areas
  • Presumably relying on similar knowledge production models

• So, what is the concern?
  • **QRPs lead to bias**
    • Biases in a paper or in whole literatures add up making their knowledge claims unreliable or even deceitful
    • Knowledge of prevalence becomes interesting as it may indicate the extent of bias
So the name of the game is bias

- But bias should be understood more broadly and not just in relation to specific quantitative data analysis techniques!
But there are actually many more ...  

100+ examples of deviations from responsible conduct of research that threaten the relevance, validity and efficiency of research, the trust between scientists, and the trust in science

- **Study design** (items that concern the phase before the start of data collection)
- **Data collection** (items that concern the phase of data collection)
- **Data-analysis** (items that concern the phase of data-analysis)
- **Reporting** (items that concern reporting of results of the study)
- **Collaboration** (items that concern obligations towards colleagues and science as a whole)

... Somewhat biased towards the medical fields
Harms to “truth” and “trust”

• The research enterprise is more than published knowledge claims

• For example, it is also a highly competitive social system, with considerable stratification in its reward structures

Breaches can be harmful to

• Primary
  • truth (validity of knowledge)
  • trust between researchers (fairness)
  • trust in science (credibility)

• Secondary
  • waste of resources and unethical to participants
  • harm to society, nature, individuals/patients
So what do we know about QRPs?

Where are they?
How prevalent are they?
What may their causes be?
How do we perceive them?
Where are they?

Judging from the literature discussing QRP issues, they are mainly a concern in the social, behavioural and medical fields.

Don’t they have authorship, citing or review problems in the “natural sciences”?
Ioannidis is over-selling his claim because it only relates to a certain knowledge production model and as it turned out, while his concern was very real his formal argument was flawed.
Where are they? 

... it depends!

Clearly, qualitative or non-empirical knowledge production models are not susceptible to a range of QRPs related to significance testing.
Other QRPs?

- Biomedical and health sciences
- Social sciences
- Life sciences
- Humanities
- Earth sciences
- Mathematics and computer science
- Physical sciences and engineering
The prevalence of QRPs
How prevalent are they?  

- Considerable variation in prevalence estimates:
  - Across studies
  - Within studies

- Estimates are dependent upon
  - Item wording
  - Framing (anchor, priming)
  - Instrument
  - Social desirability

- Primarily estimates from surveys in psychology and ecology

- Restricted set of QRPs related to quantitative data analysis

... it depends!

Figure 2. Prevalence indices (shaded bars) derived from admission rates of respondents committing questionable research practices at least once (gray bars) and repetition frequency (white bars), compared to the original John et al. (2012) data (black bars). Modified item wordings appear in italics.

The potential causes of QRPs
What are the causes or mechanisms behind QRPs?

- The usual suspects when asking researchers are “incentive structures”

<table>
<thead>
<tr>
<th>Item</th>
<th>Response categories</th>
<th>Item statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competition for publications in “high-impact journals”</td>
<td>Does not apply</td>
<td>M=3.88 SD=0.38</td>
</tr>
<tr>
<td>Competition for permanent positions (e.g., professorships)</td>
<td>Does not apply</td>
<td>M=3.74 SD=0.54</td>
</tr>
<tr>
<td>Competition for external funding</td>
<td>Does not apply</td>
<td>M=3.57 SD=0.64</td>
</tr>
<tr>
<td>Incentives within the publication system (e.g., originality over replication)</td>
<td>Does not apply</td>
<td>M=3.56 SD=0.69</td>
</tr>
<tr>
<td>Socialization processes and implicit norms within scientific training</td>
<td>Does not apply</td>
<td>M=3.49 SD=0.70</td>
</tr>
<tr>
<td>The feeling of being at a disadvantage when one meets standards that others circumvent</td>
<td>Does not apply</td>
<td>M=3.20 SD=0.85</td>
</tr>
<tr>
<td>Lack of awareness of the implications of these practices for science</td>
<td>Does not apply</td>
<td>M=2.91 SD=0.85</td>
</tr>
<tr>
<td>Lack of role models for good scientific practice within the community</td>
<td>Does not apply</td>
<td>M=2.74 SD=0.99</td>
</tr>
<tr>
<td>Competition within work groups</td>
<td>Does not apply</td>
<td>M=2.54 SD=0.94</td>
</tr>
<tr>
<td>Lack of effective sanctions of scientific misconduct</td>
<td>Does not apply</td>
<td>M=2.53 SD=0.98</td>
</tr>
<tr>
<td>Lack of methodological knowledge among social psychologists</td>
<td>Does not apply</td>
<td>M=2.20 SD=0.94</td>
</tr>
<tr>
<td>Lack of knowledge of research ethics among social psychologists</td>
<td>Does not apply</td>
<td>M=1.96 SD=0.91</td>
</tr>
<tr>
<td>Personality traits of those employing problematic practices</td>
<td>Does not apply</td>
<td>M=1.72 SD=0.79</td>
</tr>
</tbody>
</table>

Notes. Table shows valid percent with smallest N=83 due to individual missing data with items sorted by means (downward). Scale ranges from 1 (=does not apply) to 4 (=does apply).

Divergent interests

- Incentive structures of “science” are epistemic
  - E.g. double-blind design, replication, peer review etc.
  - They are typically designed to aid scientific goals

- Incentive structures in academia are economic
  - E.g. tenure, h-index, JIF etc.
  - They are not designed to meet scientific goals
    - They emerge as by-products of human goals and economic competition within and among social groups

- Consequence: many academic incentives = sloppy, bad, questionable research (or worse)?

<table>
<thead>
<tr>
<th>To be a good scientist ...</th>
<th>To be a good academic ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be skeptical of your results</td>
<td>&quot;Sell&quot; your results</td>
</tr>
<tr>
<td>Interpret conclusions carefully</td>
<td>Highlight/exaggerate importance</td>
</tr>
<tr>
<td>Publish negative results</td>
<td>Publish &quot;strategically&quot;</td>
</tr>
<tr>
<td>Ignore social prestige</td>
<td>Use impact factors to make writing decisions</td>
</tr>
<tr>
<td>Challenge authority</td>
<td>Cite authority. Make friends</td>
</tr>
<tr>
<td>Replicate. Replicate</td>
<td>Replicate ... if you must</td>
</tr>
<tr>
<td>Novel exciting results are less likely to be true. Double-check them</td>
<td>Publish novel exciting results before you get scooped</td>
</tr>
<tr>
<td>P=0.06</td>
<td>P=0.04</td>
</tr>
<tr>
<td>Help other replicate and find flaws in your work</td>
<td>Have a &quot;territory&quot;</td>
</tr>
<tr>
<td>Be clear</td>
<td>Sound smart</td>
</tr>
<tr>
<td>Log-term goals for important truths</td>
<td>Short-term results with high impact</td>
</tr>
<tr>
<td>Build a consensus</td>
<td>Hot topics are higher impact</td>
</tr>
</tbody>
</table>
What are the causes or mechanisms behind QRPs?

- But the situation is probably much more complex
  - Mechanisms are probably both social, cultural and individual
  - They interact and have varying effects

**CULTURE**
- Group think
- Wrong role models
- Insufficient mentoring

**SYSTEM**
- Publication pressure
- Hyper competition
- Low risk – high rewards

**INDIVIDUAL**
- Stress
- Moral attitudes
- Personality traits
So what?

.... obviously we need more research!
“Practices, Perceptions, and Patterns of Research Integrity” (PRINT)

- Nationally (Danish) funded research project (1 mill Euros)

- Support national integrity initiatives

- 2 year project, funding ends June 30th

- Still, ongoing and I can only give you an idea of our aims and some tentative empirical findings

- Analyses, reporting and follow-up studies will continue the next couple of years

Collaborators:
Nick Allum, Essex, UK
Jens Peter Andersen, Aarhus, DK
Pernille Bak Pedersen, Aarhus, DK
Michael Bang Petersen, Aarhus, DK
Asger Dalsgaard Pedersen, Aarhus, DK
Nicole Foeger, ENRIO, AU
Allan Rye Lyngs, Aarhus, DK
Ana Marušić, Split, CR
Niels Mejlggaard, Aarhus, DK
Philippe Mongeon, Aarhus, DK
Mads P. Sørensen, Aarhus, DK
Bobby Zachariae, Aarhus, DK

http://print-cfa.dk/
Main research objectives of PRINT

1. To examine, define and typologize Questionable Research Practices (QRPs)
2. To examine prevalence and perceptions of them within and across main fields of research (knowledge production models)
3. To elucidate the most predominant mechanisms potentially influencing QRPs, and examine how they relate to individuals, institutions, norms, and standards
So what is our motivation?

• We need better conceptual clarifications, preferably a typology (or typologies)
• We need to examine QRPs across fields and link them to knowledge production models
• We clearly need contextualization
• We need to examine varying perceptions
• We need to examine prevalence, both empirically and methodologically
• We need to examine potential associations between psychological and social factors on the one hand, an perceptions of QRPs on the other
The three work packages and their aims

**WP1**
Systematically review and elicit contextualized knowledge and perceptions on QRPs through desk research and focus-group interviews

**WP2**
Examine prevalence and perceptions of QRPs, as well as mechanisms leading to QRPs, in an international survey linked to publication data

**WP3**
Various “meta-research” analyses of publication data to unobtrusively examine markers of QRP and potential institutional and social factors generating or amplifying such practices
Some tentative empirical examples
Perceptions of QRPs
Perceptions of QRPs

• Examined in 22 focus-group interviews as well as in the survey
• Focus-groups were composed according to main field and knowledge production mode
• Various discussions and card sorting exercises dealing with perceptions, severity, prevalence and causes
  • 8 pre-defined QRPs + a number of “own-examples” QRPs
Perceptions of QRPs

• Interestingly, 7 groups from the humanities and non-quantitative areas of the social sciences suggested: “unoriginality” as a QRP

• Why is “unoriginality” so important within these groups?

“Boring”, “uninteresting”, “doesn’t bring anything new”, “reuse of arguments”
“Leads to ‘repetitiveness’ and to research output that “sounds like a broken record”, “same points, just wrapped in differently”
Also reveals a “lack of curiosity” and an “avoidance of risk taking”
“QRPs are actions that violate traditional values of the research enterprise and that may be detrimental to the research process”

“Originality” is typically linked to epistemic values and research quality
Not a behaviour or a practice

Should we consider “unoriginality” as a QRP (in some areas)?

What does it tell us about perceptions of QRPs?
Surveying perceptions and prevalence of QRPs
## Survey

<table>
<thead>
<tr>
<th>Structure</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>consent</td>
</tr>
<tr>
<td>Specific field of research</td>
<td>own formulation</td>
</tr>
<tr>
<td>Preferred research approach</td>
<td>Knowledge production model (4 categories)</td>
</tr>
<tr>
<td>Trust in research findings</td>
<td>7-point scale</td>
</tr>
<tr>
<td>Statements about potential QRPCs</td>
<td>9 QRP statements</td>
</tr>
<tr>
<td></td>
<td>Prevalence in field, 7-point scale</td>
</tr>
<tr>
<td></td>
<td>Own practice, 7 point scale</td>
</tr>
<tr>
<td></td>
<td>Severity in relation to “truth”, 7-point scale</td>
</tr>
<tr>
<td>List experiment (10th QRP statement)</td>
<td>Social desirability instrument</td>
</tr>
<tr>
<td>Statements concerning research conditions</td>
<td>6 q's on &quot;pressure&quot;, 7-point scale</td>
</tr>
<tr>
<td></td>
<td>4 q's on &quot;peer review&quot;, 7-point scale</td>
</tr>
<tr>
<td></td>
<td>6 q's on &quot;organisation/reward&quot;, 7-point scale</td>
</tr>
<tr>
<td>Personality traits</td>
<td>10 TIPI q's, 7-point scale</td>
</tr>
<tr>
<td>Demographics</td>
<td>gender</td>
</tr>
<tr>
<td></td>
<td>PhD age</td>
</tr>
<tr>
<td></td>
<td>RI-training</td>
</tr>
<tr>
<td></td>
<td>Familiarity with DCoC</td>
</tr>
</tbody>
</table>

©jws@ps.au.dk, ncrm2019
International survey: WP2

• Eight (all) Danish universities, ten foreign universities in four countries
  • Danish part: 1 invitation + 4 reminders
    • 3402 “full” responses = 22% rate (2000 more have engaged with survey)
  • International part: 1 invitation + 3 reminders
    • 1308 “full” responses = 4% rate (600 more have engaged with survey)
      • Austria 3.5%, UK 5%, Croatia 4.2% and US 2%

<table>
<thead>
<tr>
<th>Sampling according to knowledge production mode</th>
<th>Theoretical</th>
<th>Qualitative</th>
<th>Quantitative</th>
<th>Significance test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>630</td>
<td>798</td>
<td>160</td>
<td>1824</td>
<td>3412</td>
</tr>
<tr>
<td>Austria, UK, Croatia, US</td>
<td>324</td>
<td>270</td>
<td>48</td>
<td>656</td>
<td>1298</td>
</tr>
<tr>
<td>Total</td>
<td>954</td>
<td>1068</td>
<td>208</td>
<td>2480</td>
<td>4710</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sampling according to main fields and gender</th>
<th>Female</th>
<th>Male</th>
<th>Non-binary</th>
<th>Prefer not to answer</th>
<th>total</th>
<th>prop female</th>
</tr>
</thead>
<tbody>
<tr>
<td>arts &amp; humanities</td>
<td>266</td>
<td>306</td>
<td>6</td>
<td>28</td>
<td>606</td>
<td>0.44</td>
</tr>
<tr>
<td>medical &amp; health</td>
<td>495</td>
<td>676</td>
<td>17</td>
<td>30</td>
<td>1218</td>
<td>0.41</td>
</tr>
<tr>
<td>natural &amp; tech</td>
<td>489</td>
<td>1349</td>
<td>22</td>
<td>71</td>
<td>1931</td>
<td>0.25</td>
</tr>
<tr>
<td>social science</td>
<td>367</td>
<td>538</td>
<td>10</td>
<td>31</td>
<td>946</td>
<td>0.39</td>
</tr>
<tr>
<td>total</td>
<td>1617</td>
<td>2869</td>
<td>55</td>
<td>160</td>
<td>4701</td>
<td>0.34</td>
</tr>
</tbody>
</table>

©jws@ps.au.dk, wcri2019
Perceptions of trust
Trust in knowledge claims

Trust in research findings
In recent years there has been some debate about to which extent we can trust scholarly claims and findings.

In general, to what extent do you trust the findings and claims in the recent literature in your field (statskundskab)?

Danish respondents

International respondents

©jws@ps.au.dk, wcri2019
Trust in knowledge claims

Strong belief that somewhat contradicts the “crisis” narrative!
Prevalence of QRPs
What’s different

• Martinson, Anderson & de Vries (2005)
• John, Lowenstein & Prelec (2012)
• Fiedler & Schwarz (2015)
• Agnoli et al (2017)
• Fraser et al (2018)
• Hjellbrekke et al (2018)
• Artino, Driessen & Maggio (2019)
• Fox, Honeycutt & Jussim (2018)

• Much broader set of QRP statements, beyond the canonical 10 items
• Statements are framed towards “the recent literature in the field” or “own recent publications” and not persons
• Different prevalence measures
• In principle all fields of research are included
<table>
<thead>
<tr>
<th>Id</th>
<th>Type</th>
<th>QRP statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>qrp1</td>
<td>Authorship</td>
<td>Included authors on a paper who had not contributed sufficiently to the work to merit authorship</td>
</tr>
<tr>
<td>qrp2</td>
<td>Authorship</td>
<td>Failed to offer authorship to collaborators who had contributed sufficiently to the work to merit authorship</td>
</tr>
<tr>
<td>qrp3</td>
<td>Lack of transparency</td>
<td>In a publication, failed to disclose relevant personal, financial, political or intellectual conflicts of interests</td>
</tr>
<tr>
<td>qrp4a</td>
<td>Data dredging</td>
<td>In significance testing studies, continued to collect more data until you obtained a targeted result, more data until you get the desired result.</td>
</tr>
<tr>
<td>qrp4b</td>
<td>Data dredging</td>
<td>In significance testing studies, continued to re-analyse data until you obtained a targeted result.</td>
</tr>
<tr>
<td>qrp5</td>
<td>Recycling (Lack of contribution)</td>
<td>Published a research paper despite knowing that it is redundant and does not contribute to the existing knowledge base. Please note, reviews, translations and other types of research publications are not necessarily redundant if they fill a scholarly need in the scientific communication.</td>
</tr>
<tr>
<td>qrp6</td>
<td>Selective citing</td>
<td>Cited literature deemed relevant for your study without actually having read it</td>
</tr>
<tr>
<td>qrp7</td>
<td>Selective reporting</td>
<td>Claimed to have used a particular qualitative analytical approach appropriately, for example “grounded theory” or “triangulation”, when this was not the case.</td>
</tr>
<tr>
<td>qrp8</td>
<td>Lack of transparency</td>
<td>Avoided to share data, research protocols, information on experimental setup, instrumentation, coding, or other information about a study requested by colleagues to evade transparency.</td>
</tr>
<tr>
<td>qrp9</td>
<td>Biased reviewing</td>
<td>Agreed to review a manuscript despite knowing that you have inadequate expertise to provide a competent review.</td>
</tr>
<tr>
<td>qrp10</td>
<td>Biased reviewing</td>
<td>When reviewing a manuscript, not invested the effort necessary to conduct a thorough review.</td>
</tr>
<tr>
<td>qrp11</td>
<td>Biased reviewing</td>
<td>Submitted a biased review report that evaluated the manuscript unfairly.</td>
</tr>
<tr>
<td>qrp12</td>
<td>Recycling</td>
<td>In a publication, deliberately reused all or parts of previously published data without disclosure.</td>
</tr>
<tr>
<td>qrp13</td>
<td>Recycling</td>
<td>Deliberately divided the results of a study over more publications than needed with the intention to increase the number of publications.</td>
</tr>
<tr>
<td>qrp14</td>
<td>Selective citing</td>
<td>Selectively cited irrelevant or unnecessary publications to please reviewers or editors.</td>
</tr>
<tr>
<td>qrp15</td>
<td>Selective citing</td>
<td>Deliberately cited your own publications more than warranted by their relevance, to promote the visibility of your work or improve your citation metrics.</td>
</tr>
<tr>
<td>qrp16</td>
<td>Selective citing</td>
<td>Willfully disregarded citing relevant publications that contradicts your own beliefs, theories, hypotheses, methods or findings.</td>
</tr>
<tr>
<td>qrp17</td>
<td>Data dredging</td>
<td>When analysing data, selectively focused on parts of the data or source material that support your preconceptions or hypotheses and disregarded parts that do not.</td>
</tr>
<tr>
<td>qrp18</td>
<td>Selective reporting</td>
<td>Deliberately refrained from reporting findings that could weaken or contradict your theories, hypotheses or findings.</td>
</tr>
<tr>
<td>qrp19</td>
<td>Selective reporting</td>
<td>Willfully presented findings as more “clear-cut” than justified by the data. For example, neglecting to disclose contradictory results, or thoroughly discuss study limitations, or deliberately overlooking counter arguments.</td>
</tr>
<tr>
<td>qrp20</td>
<td>Selective reporting</td>
<td>Without disclosure, formulated or changed your hypothesis after having seen the results, thereby presenting an unexpected finding as having been predicted from the start in the form of a research hypothesis.</td>
</tr>
<tr>
<td>qrp21</td>
<td>Selective reporting</td>
<td>Without disclosure, presenting an unexpected finding as having been predicted from the start.</td>
</tr>
<tr>
<td>qrp22</td>
<td>Selective reporting</td>
<td>Presented statistically significant main findings, without distinguishing between their “statistical significance” and their potential practical or theoretical importance. For example claiming a “significant” finding solely based on the p-value and not the effect size.</td>
</tr>
<tr>
<td>qrp23</td>
<td>Selective reporting</td>
<td>Reported main findings that turned out not to be “statistically significant”, as evidence for no difference, no effect or no association between study groups or variables.</td>
</tr>
<tr>
<td>qrp24</td>
<td>Plagiarism</td>
<td>Deliberately used another researcher's unpublished idea without giving credit. For example, publishing an idea voiced by a colleague at an informal meeting without giving her/him credit.</td>
</tr>
</tbody>
</table>
Estimating prevalence in the recent literature in one’s specific field of research

1 of 9
Consider the following research practice:

**Presenting statistically significant main findings, without distinguishing between their ‘statistical significance’ and their potential practical or theoretical importance.**

For example, claiming a ‘significant’ finding solely based on the p-value and not the effect size. Notice, importance refers to clinical, economic, biological, psychological, sociological etc.

To what extent do you believe this practice is used in the recent publications reporting statistical significance tests in your specific field of research?

- In no recent publications
- In all recent publications
- Unable to answer

To what extent is this practice used in your recent sole or co-authored publications reporting statistical significance tests?

- In no recent publications
- In all recent publications
- Such tests were not used
- Unable to answer

To what extent do you believe this practice is damaging the trust in research findings and claims in your specific field of research?

- To a very small extent
- To a very great extent
- Unable to answer

Optional: Please feel free to elaborate on the above questions...
Estimating prevalence in the recent literature in one’s specific field of research

Danish respondents

International respondents

©jws@ps.au.dk, wcri2019
Estimating prevalence in the recent literature in one’s specific field of research

1 of 9

Consider the following research practice:

*Presenting statistically significant main findings, without distinguishing between their 'statistical significance' and their potential practical or theoretical importance.*

For example, claiming a 'significant' finding solely based on the p-value and not the effect size. Notice, importance refers to clinical, economic, biological, psychological, sociological etc.

To what extent do you believe this practice is used in the recent publications reporting statistical significance tests in your specific field of research?

- In no recent publications
- In all recent publications
- Unable to answer

To what extent is this practice used in your recent sole or co-authored publications reporting statistical significance tests?

- In no recent publications
- In all recent publications
- Such tests were not used
- Unable to answer

To what extent do you believe this practice is damaging the trust in research findings and claims in your specific field of research?

- To a very small extent
- To a very great extent
- Unable to answer

Optional: Please feel free to elaborate on the above questions...
Reporting own practice in one’s recent publications

Danish respondents

International respondents

©jws@ps.au.dk, wcri2019
Estimating prevalence in the recent literature in one’s specific field of research
Strong symmetry in responses

Higher estimates of prevalence in one’s field $\approx$ higher self-reporting prevalence and vice versa
Aggregate results (dichotomizing self-reporting)

To what extent is this practice used in your recent sole or co-authored publications?

According to previous studies, this would produce a prevalence rate of ≈ **0.37** (Danish) and **0.40** (International) based on dichotomizing the ordinal response. Median number of “positive” QRP questions for individual respondents are 3 (average 3.1 (Danish) and 3.4 (International)).
Overall discrepancy between how respondents perceive the recent literature in their field and their own recent practices (Danish respondents)

According to previous studies, this would produce a prevalence rate of 0.37 (self-admission) and 0.88 (field-perception), based on dichotomizing the ordinal response.
Seven QRPs deemed most damaging for “trust” (dichotomized)

Aggregate prevalence for the seven QRPs deemed most damaging for “trust” ≈ 0.32 (Danish) and 0.36 (international).
Challenges estimating the prevalence of QRPs
Challenges estimating the prevalence of QRPs

• Despite the striking symmetry, what have we actually estimated?
• To what extent do we see anchoring effects where self-reporting is influenced by previous estimates of field prevalence?
• To what extent do the wording of QRP statements influence the prevalence estimates? Our wording is more general.
• To what extent do the framing of “in most recent publications” influence the prevalence estimates? Should it be more specific for respondents to recollect?
• To what extent do other psychological factors well-known to impact respondents in surveys influence our estimates, e.g. social desirability?
• We clearly need to scrutinize the responses and the response patterns
  • Follow-up survey
  • List experiment
  • Linking respondents’ answers to their recent publications. Publications are then scrutinized for QRP markers. Subsequent comparison of the self-reported prevalence with unobtrusively obtained aggregates of QRPs (e.g. p-curves)
Challenges estimating the prevalence of QRPs

• Despite the striking symmetry, what have we actually estimated?
• To what extent do we see anchoring effects where self-reporting is influenced by previous estimates of field prevalence?
• To what extent do the wording of QRP statements influence the prevalence estimates? Our wording is more general.
• To what extent do the framing of “in most recent publications” influence the prevalence estimates? Should it be more specific for respondents to recollect
• To what extent do other psychological factors well-known to impact respondents in surveys influence our estimates, e.g. social desirability?
• We clearly need to scrutinize the responses and the response patterns
  • Follow-up survey
  • List experiment
  • Linking respondents’ answers to their recent publications. Publications are then scrutinized for QRP markers. Subsequent comparison of the self-reported prevalence with unobtrusively obtained aggregates of QRPs (e.g. p-curves)

Indeed, is it meaningful to estimate prevalence in surveys?
The potential causes of QRPs
Some preliminary findings in relation to personality traits’ potential association with QRP responses

- Personality: TIPI Big Five Inventory (Gosling et al., 2003)
- Count of QRPs related to authorship, selective analysis, undisclosed recycling, selective citing, selective reporting and reviewing (total of 9)
- “Controls” include: Gender, main field, PhD year and country
Some preliminary findings in relation to personality traits’ potential association with QRP responses

- QRPCs are especially prominent among those
  - low in Openness,
  - low in Agreeableness
  - low in Conscientiousness

- Associations with Agreeableness and Conscientiousness are in line with findings in other domains: Sloppiness and selfishness are associated with higher prevalence of QRPCs

- Association with Openness is specific for QRPCs: The psychological costs of being wrong?

©jws@ps.au.dk, wcrl2019
Provisional findings in relation to research conditions’ potential association with QRP responses

Factor analysis & scale construction

<table>
<thead>
<tr>
<th>Standardised factor loadings (ML estimation)</th>
<th>Local culture</th>
<th>Personal pressure</th>
<th>Publishing culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership at my department role models for RI.</td>
<td>0.91</td>
<td>0.02</td>
<td>-0.13</td>
</tr>
<tr>
<td>Most prominent researchers at my department are role models for RI.</td>
<td>0.75</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>Rewards at my department encourage most rigorous research.</td>
<td>0.69</td>
<td>-0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Peer culture within my department is a safeguard against QRP.</td>
<td>0.57</td>
<td>-0.06</td>
<td>0.14</td>
</tr>
<tr>
<td>I feel pressure to publish in a prestigious outlet.</td>
<td>0.04</td>
<td>0.66</td>
<td>0.04</td>
</tr>
<tr>
<td>I feel strong pressure to attract external funding.</td>
<td>-0.06</td>
<td>0.63</td>
<td>0.04</td>
</tr>
<tr>
<td>Bibliometric indicators are important for my career.</td>
<td>0.04</td>
<td>0.62</td>
<td>0.04</td>
</tr>
<tr>
<td>When applying for funding I feel pressure to oversell impact.</td>
<td>-0.10</td>
<td>0.54</td>
<td>-0.15</td>
</tr>
<tr>
<td>QRP are rewarded in the leading publication outlets within my field</td>
<td>-0.03</td>
<td>-0.01</td>
<td>0.69</td>
</tr>
<tr>
<td>Peer review and editorial provide sufficient safeguard against QRP</td>
<td>0.03</td>
<td>0.03</td>
<td>0.59</td>
</tr>
<tr>
<td>The most prestigious outlets in my field have highest standard of RI</td>
<td>0.07</td>
<td>0.10</td>
<td>0.57</td>
</tr>
<tr>
<td>Peer review and editorial encourage me to oversell my results</td>
<td>-0.03</td>
<td>-0.27</td>
<td>0.39</td>
</tr>
<tr>
<td>Variance explained</td>
<td>22%</td>
<td>14%</td>
<td>8%</td>
</tr>
<tr>
<td>Scale reliability (Cronbach’s alpha)</td>
<td>.82</td>
<td>.70</td>
<td>.63</td>
</tr>
</tbody>
</table>

Mean perceptions (factors) by field

- **Local culture**
- **Publishing culture**
- **Personal pressure**

Only Danish respondents
Provisional findings in relation to research conditions’ potential association with QRP responses

Note: QRP score is percentage of QRPs that respondent reports ever doing out of 9 offered.

• Substantial commonality of perceptions of research context and environment across fields
• More pressure associated with a slightly higher reported QRPs
• Confidence in publishing and department and local leadership associated with fewer reported QRPs

©jws@ps.au.dk, wcri2019
... and there is much more to come!
Rounding up
Hitherto, focus of QRP studies have been too narrow

- Much focus on p-hacking, HARKing and similar practices
  - Fine, but that does “only” address certain dimensions of QRPs: “data dredging” and “selective reporting”
  - And a specific quantitative/statistical knowledge production model
- Clearly, the challenge of QRPs go beyond this focus
- Likewise we should not make general claims about “science” of “scholarly activities” when our data are very restricted!
Not necessarily consensus about perceptions

- HARKing are seen as detrimental in confirmatory hypothesis testing studies.
- However, other traditions do not necessarily see such practices as particularly questionable.

Social science perception of severity (Danish respondents)

Quantitative

21: “Without disclosure, formulated or changed your hypothesis after having seen the results, thereby presenting an unexpected finding as having been predicted from the start in the form of a research hypothesis”

Qualitative

Humanities perception of severity (Danish respondents)

Quantitative

22: “Without disclosure, presenting an unexpected finding as having been predicted from the start”

Qualitative

©jws@ps.au.dk, wcri2019
Conceptual challenges

- Questionable (negotiable) versus detrimental (clear-cut)
- Can we “objectively” or “unanimously” determine that a practice is “questionable” (by itself contradictory) or “detrimental?

- This is the legal situation in Denmark where our national legislation states that a national committee should investigate and rule in misconduct cases
  - If FFP then misconduct, if not then the case is returned to the university and their task is now to examine whether it is QRP
  - The challenge: Nowhere is it stipulated when it is QRP, or what counts as QRPs
  - The latter is to be determined by “experts” based on norms, traditions and values within (arbitrarily defined) fields
Conceptual challenges

• The conceptual approach and essential categorization so far in this area is driven by (post)positivist ideals (psychological and medical fields)

• Nothing wrong *per se* with such ideals, except that they don’t work for everybody, hence they are points of view

• Notions of QRPs should be linked to knowledge production models/modes

• Knowledge production models, their relevance criteria and implicit social norms are based upon (different) philosophies (ontological, epistemological and methodological)

• This should be respected!
So where are the “solutions”?

• They are underway
• Some are being probed or tested

• But we still need a better mapping and understanding of the problems and challenges of QRPs across fields and knowledge production models in order to respect diversity and come up with contextualized solutions to promote good research practice!
Thank you for your attention!

jws@ps.au.dk

This work is supported by the PRINT project (Practices, Perceptions, and Patterns of Research Integrity) funded by the Danish Agency for Science and Higher Education (Ministry of Higher Education and Science) under grant No 6183-00001B.