Science under Pressure

Essays on Research Integrity

Dies Natalis 2019

Tilburg University

Understanding Society

Science under Pressure

Essays on Research Integrity

on occasion of the Dies Natalis 2019 of Tilburg University

Dies Natalis 2019

Table of contents

Introduction

You have before you a set of essays written on occasion of the 92nd Dies Natalis of Tilburg University, celebrated on November 21, 2019. The five essays share the common theme of the Dies, which was "Science under Pressure". Without further focusing, Science under Pressure can refer to a multitude of topics ranging from the repercussions the Van Rijn report has on the financial situation of the Dutch universities to the influence fake science disseminated through the Internet has on public opinion, but the Dies' theme referred to scientific integrity and the transition science is going through at present. Many routines, traditions and rules for the planning, execution and reporting of research are the subject of discussion and it is difficult to say where the practice of research will be in, say, five to ten years from now.

In the previous decade, a number of serious breaches of scientific integrity have plagued the practice of Dutch and international science. In the Netherlands, these integrity breaches have had the effect of a wake-up call on the universities and other knowledge organizations implying that like any other game, the game of science needs a set of rules for its participants. Competition for limited journal space and grant money, premiums for completed PhD dissertations, jobs, prizes and status not only promote creativity and productivity but also drive some individuals toward the boundary where integrity ends and breaches of integrity start. This calls for direction.

The recent version of the Netherlands Code of Conduct for Research Integrity was presented in 2018 by the Royal Netherlands Academy of Arts and Sciences (KNAW), the Netherlands Federation of University Medical Centres (NFU), the Netherlands Organisation for Scientific Research (NWO), the Associated Applied Research Institutes (TO2), the Netherlands Association of Universities of Applied Sciences (VH), and the Association of Universities in the Netherlands (VSNU). The Code allows these organizations to show that integrity is an inalienable part of their research practice and to provide their researchers with an open environment in which they can share concerns about dilemmas and discuss errors made, and take full responsibility. A vital part of the Code is the organizations' duties of care and the individual researcher's accountability for her professional behavior.

Questionable research practices present a less visible problem that is at least as serious as the often rather eccentric integrity breaches that nevertheless prove highly damaging for science' credibility and reputation. An important difference with integrity breaches is that questionable research practices happen more frequently but usually without any bad intentions; they simply happen due to unfortunate habits (e.g., to keep one's data to oneself), goals (e.g., to publish as many articles as possible), and misunderstandings (e.g., applying statistical models one badly understood). Both in the Netherlands and in the international science arena an intense discussion about research integrity and improved research practices continues to take place. Many proposals have been done and numerous discussions are viral that revolve around the questions of how to promote better research integrity and how to abandon undesirable research practices, irrespective of whether they are the result of intention.

The five essays in this booklet touch upon topics that address the culture of doing research and each argues for change. In his Dies opening speech, Klaas Sijtsma discusses limitations on academic freedom and accountability of researchers for providing open data and seeking statistical assistance. Michèle Nuijten argues that open data should be the default in data management but also leaves room for exceptions. Han van der Maas presents an alternative to classical peer review leaving the initiative to publish with the author, and argues that this will reduce the avalanche of often low-quality articles. Anne de Vries criticizes drawbacks of the financial rewarding of completing PhD research and discusses alternatives. Finally, in his Dies keynote, Lex Bouter asks attention for the role research institutes can and should play in promoting research integrity.

This volume of essays covers a couple of important topics but not all. For example, the preregistration of research time-stamps research entries so that it can be seen whether later changes of research questions, research design, statistical methods and the data-analysis plan are potentially data-driven, which may invite a high risk of non-replicable results. The issue of quantity (publish as much as you can or perish) versus quality (publish only important work) is only touched upon. These and other topics are not discussed or only mentioned in passing, but the topics that are discussed are highly valuable and topical. I hope that this volume provides food for thought and inspiring discussions, and adds if only modestly, to realizing the best research practices.

Klaas Sijtsma
Rector Magnificus of Tilburg University
November 2019

Klaas Sijtsma

Rector Magnificus of Tilburg University

Science under Pressure Creates a New Research Culture Promoting Open Data and Better Research Skills[1]

[1] A Dutch-language version of this contribution will appear in *De Psycholoog*.

Science under Pressure can refer to many different issues, ranging from the financial repercussions the Van Rijn Report[2] will have on our university to the dissemination of fake-science concerning the alleged dangers of vaccination. However, I am referring to issues concerning research integrity. These difficult issues are at the heart of what we do and who we are, and put pressure on our credibility. In times in which opinions count as facts and facts as opinions, we depend on our credibility. I wish to emphasize that I do not mean to say that researchers are doing their scientific work without integrity and are ignoring ethical standards; absolutely not. But the academic community at large including Tilburg University is in the middle of a transition process toward a new and comprehensive research culture including the adaptation of a new code of conduct, which is the Netherlands Code of Conduct for Research Integrity 2018[3].

Science blossoms by pursuing unusual solutions to familiar problems and taking unexpected turns thus leaving a planned trajectory. The classical idea of academic freedom focuses on the individual scientist who can choose to study whatever she sees fit without interference by interest groups. This ideal, irrespective of whether it ever was realistic, is difficult to maintain today. Society expects science to come up with results useful for the general good; that is, science must have impact. Government, companies and charities, but also the Netherlands Organization for Scientific Research (NWO) and the European Research Council (ERC) want to have a say in what we study. This presents another call on us to realize impact.

Emphasis on team science

However, for impact to be valid and credible, researchers need an environment without restrictions on research methods, execution of research, and publication of results, so that they can do their scientific work without the interruption of interest groups that may compromise validity and credibility. The recent emphasis on team science[4] will further restrict the individual's freedom to some degree, meanwhile protecting individual researchers better from unreasonable expectations about their output, so typical of the past few decades. Clearly, freedom is not absolute and comes with practical limitations.

[2] Rapport Commissie Bekostiging Hoger Onderwijs en Onderzoek. *Wissels om. Naar een transparante en evenwichtige bekostiging, en meer samenwerking in hoger onderwijs en onderzoek.* Mei 2019. Downloaded from https://www.rijksoverheid.nl/documenten/rapporten/2019/07/19/adviesrapport-bekostiging-hoger-onderwijs

[3] *Netherlands Code of Conduct for Research Integrity 2018.* https://www.vsnu.nl/en_GB/news-items/nieuwsbericht/471-new-netherlands-code-of-conduct-for-research-integrity.html

[4] *Room for everyone's talent. Towards a new balance in the recognition and rewards of academics.* Downloaded from https://www.vsnu.nl/Erkennen-en-waarderen-van-wetenschappers.html

A downside to limitless academic freedom as we may have known it is that it focused so much on stimulating creativity that it forgot to hold researchers accountable for what they do. A couple of scandals at various universities including ours have presented a rough wake-up call. These scandals share a few characteristics. They all revolve around individuals having strong personalities, making their own rules, using the team they run for their own agenda, and requiring obedience at the cost of rejection from the tribe. For team science to be successful as a new model for scientific work, the modern team leader must embrace and disseminate generally accepted rules for good research, be able to give rather than take, and grant other, often younger team members the freedom to develop their own ideas and to publish them. This is quite a challenge in the highly competitive world of science and requires a lot of our leaders.

Academic freedom must not be abolished, but a generally accepted set of rules are necessary to play the game of science in an orderly fashion, like any game. I wish to emphasize that, in addition to team work, we need to grant individuals the opportunity to come up with brilliant ideas and to allow them to fail while trying—we don't want to miss an Einstein when one presents herself—meanwhile recognizing that team science offers many merits as well.

Today, it looks like all of academia is going through a phase of enlightenment that involves a growing awareness that we make mistakes and are no saints. Is making mistakes a bad thing? The Roman philosopher Seneca said it right, I think, when he stated: *Errare humanun est, sed in errare perseverare diabolicum*. Loosely translated: To err is only human, but to persist in one's errors is unforgivable. What are the errors we make and how can we prevent them from happening?

Much so-called meta-research focuses on errors in doing research and what we can do about them. Steneck (2006)[5] presented the next continuum running from right to wrong when practicing the noble art of science.

RCR ----------QRPs----------FFP

[5] Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics*, 12, 53–74.

On the left, RCR stands for Responsible Conduct of Research. This is textbook research taking place under ideal conditions, but somewhat remote from real research. On the right, FFP stands for Fabrication, Falsification, and Plagiarism. Fabrication is making up one's data with the intention to produce a particular outcome, such as an experimental effect. Falsification is spicing up one's research outcomes by leaving out unwelcome information, such as conditions suggesting an overall non-linear trend, while concluding that a linear trend is present. Plagiarism is theft of someone else's text or ideas while suggesting that they are the author's. Fabrication and falsification refer to fake science, but plagiarism may concern real and valid results. Researchers engaging in FFP are eccentrics, but when exposed they attract a disproportional amount of attention. This damages our credibility.

Research practices undermining our credibility

What worries me most, are the Questionable Research Practices (QRPs). Here we are in the world of real research as we can see it happening all around us. Because they are numerous, QRPs are the true disruptors of our credibility, and they undermine our credibility slowly and unnoticed, until it is too late. John, Loewenstein (who is an honorary doctor at Tilburg University), and Prelec (2012)[6] collected data from approximately 2000 psychological researchers from different disciplines asking them whether they had ever been involved in any of the activities listed in Table 1. The uncorrected results show that 63% of the respondents reported at least once not having reported all measurements. This means that they may have provided a biased account of their results. Another 38% reported at least once having deleted cases when that affected the results, again possibly reporting biased results. Other studies report similar results (e.g., Fanelli, 2009)[7], suggesting that many researchers engage in research activities that take place in the grey area of QRPs. John et al. (2012) emphasize that the activities listed in Table 1 can be defensible in particular research projects, meaning that they are not QRPs in all contexts.

Obviously, in addition to not all activities listed in Table 1 (and other activities) counting as QRPs in all research projects, not everybody engages in QRPs and when researchers engage in QRPs, they often do not realize that they do!

[6] John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532.

[7] Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE, 4*(4), e5738. https://doi.org/10.1371/journal.pone.0005738.

Table 1: Examples of QRPs (John, Loewenstein, & Prelec, 2012; *N*=2,155) and Percentages of Researchers Admitting They Engaged at Least Once in a QRP.

	QRP	%
1.	Not reporting all measurements	63
2.	If test is non-significant, collecting more data, and repeating testing	56
3.	Not reporting all experimental conditions	28
4.	Reporting intermediate result as if it were the final result	16
5.	Favorable rounding of p-values (p = .058 reported as .05)	22
6.	Reporting only results favoring ones hypothesis	46
7.	Deleting cases if that produces better results	38
8.	Presenting unexpected result as if it were expected	27
9.	Denying relation with demographic variables when uncertain	3
10.	Falsifying data, etc—FFP; not QRPs	0.6

At Tilburg University, we have taken several measures to improve our research practices. Examples are our improved promotion regulations, data management regulations, a system of counselors and committees for handling alleged breaches of integrity, and a research integrity training for PhD students. Policy has a tendency to stick somewhere in the organization's hierarchy before it lands in the frontline where researchers are active. Did our policy land in the frontline? A recent inventory shows two results. First, there are fine examples of local initiatives that make me proud of our university. Second, similar to other professional organizations, making policy work requires a long-term effort. We are facing no less than the need for a culture change, taking effort and requiring perseverance and patience from all involved. Like all Dutch universities, Tilburg University has embraced the VSNU code of conduct for research integrity (see Footnote 2). All scientific personnel fall under this policy and must implement it into their research practice.

Over the past decades, Dutch science has seen a growing competition among researchers for jobs, job promotion, grant money, articles in top journals, reputation, prizes and status. Competition stimulates creativity but also drives people towards boundaries they must not cross. The VSNU's new vision on academic careers (see Footnote 3) intends to overcome the rat race that holds us hostage. In addition to emphasizing team science, quality will replace quantity, and the new Standard Evaluation Protocol 2021—2027 for research audits will undergo

a complete makeover in this respect. These and other measures may reduce the pressure that threatens spoiling the fun game science also is.

Measures to improve research quality

Being a statistician, my greatest concern is the way researchers sometimes mishandle their data and statistics for data analysis. Incorrect data handling and statistics use are QRPs, and researchers engage in QRPs mostly without realizing this. Elsewhere (Sijtsma, 2016)[8], I have argued that two measures in particular can improve research quality. The first measure is publishing one's data. This serves two purposes. Data publication enables other researchers doing secondary analyses, thus allowing a more efficient use of data that are often collected with tax money. But it also allows colleagues checking the results one reported, thus encouraging the researcher to publish results that are trustworthy. The second measure is to have a statistician help the researcher analyzing her data when the data are complex and statistical problems become really difficult. The training in statistics many researchers received is basic but does not provide them with experience needed to bypass the intricacies of statistics. Statistics is counterintuitive and difficult, and researchers untrained in advanced statistics readily fall into the traps statistics presents (Kahneman, 2011)[9]. Basic statistics training should not only make future researchers aware of the different statistical methods they can use, but also should convince them not to do everything themselves, especially when things get complex, but rather involve an expert.

Society has great trust in science. The Rathenau Instituut (Van den Broek-Honingh, & De Jonge, 2018)[10] reported last year that of all societal institutions people by far trust science the most, including information science provides about vaccination. That is good news but it would be naive to think trust is out there forever. Trust can take years to build but can be gone in moments. Considering what science has accomplished in the past century, transforming a couple of regulations supporting improved research practices into behavior should be child's play. The real obstacle in the way of this transformation is us; we have to change ourselves and this may not be easy. All the more reason to start today and get this done.

[8] Sijtsma, K. (2016). Playing with data—Or how to discourage questionable research practices and stimulate researchers to do things right. Psychometrika, 81, 1-15.
[9] Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books Ltd.
[10] Van den Broek-Honingh, N., & De Jonge, J. (2018). *Vertrouwen in de wetenschap – Monitor 2018*. Den Haag: Rathenau Instituut

Klaas Sijtsma studied personality psychology and statistics and measurement theory at the University of Groningen and specialized in methods and techniques of psychological research, in particular psychometrics (i.e., the statistical aspects of measuring psychological attributes). He has published many articles in leading international journals and three books on psychometric topics.

Sijtsma was affiliated with the University of Groningen, the Vrije Universiteit Amsterdam and Utrecht University and was appointed at Tilburg University in 1997 where he holds the chair of methods and techniques of psychological research. From 1999 to 2010 he was head of the Department of Methodology and Statistics and from 2011 to 2017 he was dean of the School of Social and Behavioral Sciences. In 2019 he was appointed as rector magnificus. He holds and has held many board positions. Sijtsma is member of the Supervisory Board of CITO Arnhem. He was President of the Psychometric Society from 2010 to 2011.

Michèle B. Nuijten

Assistant Professor at Tilburg School of Social and Behavioral Sciences, Tilburg University

Science under Pressure: Make Open Data the Default

In my *Introduction to Research Methods* course, I tell my first-year students that science is a good source of knowledge because it does not require one to rely on anyone's authority. You can check scientific conclusions for yourself. In a later lecture, I talk a bit more about the scientific enterprise and how science works in practice. If I then tell the students that many researchers do not share the data underlying their studies, the students are often confused. Wasn't the whole point of science that we do not have to simply take the researcher's word for it? If researchers refuse to share their data, how can we verify that their conclusions are correct?

I think my students' confusion is justified. Data are such an important part of scientific research, that it should be obvious that we need be able to access them. That is why I argue that open data should be the default.

Benefits of sharing data 1: verifying results and conclusions

One major benefit of open data is that it becomes possible reanalyzing the data to verify the conclusions an author presented in her paper. This touches upon the core of scientific research (and the core of my students' confusion): In science we should not have to take someone's word for it, but we should be able verify the conclusions ("check the facts") for ourselves.

With available data, there are several things we can check. First of all, we can check the overall "numerical reproducibility" of a paper. Numerical reproducibility means that another researcher is allowed access to the original data and run the same analyses as described in the paper to see if she can find the same numbers as the ones reported in the paper.

It may seem obvious that if you take the same data and run the same analyses that the paper reported, the same numbers—that is, the numerical results reported in the paper—should be found. However, in practice this is not always the case. For instance, in some of my own work, I have found that roughly half of the published papers in psychology contain at least one statistical error. In about 12% of the papers this error could have affected the conclusion.[1] If data are available, such errors can be detected and corrected.

Besides simply checking whether you can reproduce the numbers a paper reports, open data also allow you to check how robust conclusions are to small changes in the statistical analyses. Researchers usually analyze their data in many different ways. Meta-research aimed at studying the way researchers do their research shows that researchers sometimes take advantage of this flexibility and try different analyses until one of them shows the desired result.
Trying all the different routes in this "garden of forking paths"[2] increases the chance of a false positive finding; that is, a result that suggests that an effect is present when in reality the effect is absent. This becomes especially problematic if researchers only report the statistical analysis "path" that led to the significant result, and omit the other paths from their paper.

This opportunistic use of the flexibility in data analysis usually is not the work of "evil scientists" who are deliberately trying to misrepresent results for their own gain. It is more likely that researchers fall prey to biases that make them believe that, in hindsight, one analysis leading to the significant result was actually the correct analysis all along!

It is very hard, if not impossible, to switch off cognitive biases that justify your choices. After all, scientists are only human. That is another reason why it is so important to have access to the raw data. We do not *have* to trust that researchers are unbiased, because we can check the robustness of their results. If it turns out that the entire conclusion hinges on the inclusion of one seemingly arbitrary covariate or on the removal of one outlier, other scientists can draw their own conclusions about the importance of the results.

Benefits of sharing data 2: answering new questions

If open data would be the new default, many more research questions could be answered. Data sets often contain a wealth of information that is not always used in the project that the data were collected for. Researchers sometimes collect data for more variables than they use for answering their research question. It would be a waste to keep these data locked up, when there is so much unused information in there that could give rise to new insights.

Let me give a concrete example. My study about statistical errors that I mentioned above was quite explorative in nature. In other words, my coauthors and I did not have concrete hypotheses we wanted to test, but we wanted to describe the prevalence and the nature of the statistical errors in the psychological research literature. Because we did not have an explicit testable question in mind when we started the data collection, we recorded many variables (e.g., the year the

[1] Nuijten, M. B., et al. (2016). "The prevalence of statistical reporting errors in psychology (1985-2013)." *Behavior Research Methods*, *48*(4), 1205-1226. https://doi.org/10.3758/s13428-015-0664-2

[2] http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

article was published, the number of statistical results reported, the number of these results that seemed incorrect, and the type of error made). We collected this information for over 15,000 articles. To keep our paper readable, we had to limit ourselves to describing only our most interesting findings, which meant that many interesting questions were left unanswered even when the data contained information for answering these questions. We were therefore very happy that, when we shared these data, other researchers used them to answer questions we had not even thought about.[3]

If raw data are available it also becomes a lot easier to include a study in a *meta-analysis*. A meta-analysis is a statistical summary of all the existing literature on a specific question. Given the exponential growth of the number of scientific papers, meta-analysis is an indispensable tool for making sense of the overwhelming literature. Unfortunately, the results sections in many papers do not contain sufficient information to allow them being included in a meta-analysis. If the raw data would have been available, this would have allowed the meta-analyst making fewer guesses and she would be in a better position to summarize the literature.

What if you have good reasons not to share your data?

Data sharing seems to have many advantages for science in general: filtering out errors, checking robustness of conclusions, answering new questions, and summarizing information. These are all assets that I expect will improve scientific quality and efficiency facilitated by secondary data analyses. However, there can also be situations in which sharing data might seem difficult.

One often-heard reason why researchers do not want to share their data, is to protect the privacy of their participants. This is a valid concern; it is of paramount importance that we are able to guarantee our participants confidentiality so that it is impossible to trace a particular data record back to a particular participant. Fortunately, there are ways to protect our participants' privacy and still share the data.

First of all, it is often straightforward to anonymize data: Simply remove all identifying information, such as names, or variables that could identify a person when seen in combination with other available information (e.g., if you see a nameless person in a data set who is female, 29 years old, lives in area 5042PN in Tilburg, and is an assistant professor, you can find out that this person is me). Removing such demographic variables from a data set is often not a problem, since the information is irrelevant for the research question anyway.

If anonymizing data is not possible, it might be an option to share your data if the receiving party agrees to certain restrictions. For instance, you might be able to draw up a contract that prohibits the person requesting the data from further sharing, or a contract that obliges the researcher to delete the data once the analyses are completed. In most cases, a compromise can be found between protecting privacy and allowing other researchers to check whether your numbers are correct.

Another reason why researchers can be reluctant to share their data has to do with the time it takes to collect those data. If you spent the last five years collecting your massive data set, it makes sense that you want to have the credits for this accomplishment. If others can immediately use the data you collected for their own publications, they might scoop you and write the papers that *you* wanted to write but without the effort you went through collecting the data.

I think this is also a valid concern about data sharing. Researchers are usually evaluated based on their publications, so it seems fair to give the person who collected the data a chance to publish results based on these data first. One solution to this problem is releasing the data "under embargo". The researcher archives the data and uploads them to a website that will remain "locked" for a certain period of time. After the embargo period has passed (and the researcher presumably has written everything she wanted to write about the data), the data are automatically made public.

And there is more good news for researchers who understandably want to have the credits for the data they collected: There is evidence that papers based on open data are cited more than papers not based on open data.[4] On top of that, it is even possible to publish the data themselves in scientific journals, so people can cite the actual data set.[5]

[3] For instance, Daniël Lakens used the data to "guess" likely causes of the errors (http://daniellakens.blogspot.com/2015/10/checking-your-stats-and-some-errors-we.html), and Chris Hartgerink and colleagues used the specific p-values to estimate if there was evidence for false negative findings (https://doi.org/10.1525/collabra.71).

[4] Piwowar and colleagues, 2007 (https://doi.org/10.1371/journal.pone.0000308)

[5] See for instance the Journal of Open Psychology Data

To summarize, both cases about data sharing show valid concerns, but also practical ways of dealing with these concerns. There are more situations imaginable where data sharing might not be straightforward, but I think that in most cases data sharing actually *is* quite straightforward, and that otherwise data can be shared with certain restrictions or adaptations.

Making open data the default

Science revolves around data, and to really be able to rely on scientific conclusions, we need access to those data. So how can we make that happen?

In order to make open data the default, several things in the current scientific system need to change. Some of the change takes place bottom up. An increasing number of researchers is sharing their data, simply because they feel it is good practice that should be favored over not sharing data. I see this in young researchers, and PhD students in particular seem to be in the forefront of the open science movement, and practice what they preach by sharing what they can.

To make open data the default, we probably also need some top-down interventions. These interventions can be rewards or encouragements; for instance, some journals use a system where papers that come with open data get an "open data badge" as an acknowledgement, and some evidence suggests that this system works! It may also be necessary to impose rules for data sharing. Funders may add a requirement to their grants that researchers have to share their data. Furthermore, my Faculty at Tilburg University requires researchers to archive their data in such a way that at least one other person can reach and understand the data.

One-size-fits-all solutions are hard to come by. Not every strategy will work for every research area, every researcher, or every study. We need to test whether interventions increasing data sharing actually work, and we need to be sensitive to restrictions on sharing.

What we should *not* do, is see any hurdle or restriction as an invitation to stop thinking about other ways to share our data. Sharing data should become the new normal, and if you feel like you cannot be part of this reality, it is up to you to convince everyone why your situation is different.

Open data is not always possible, but it should be the default.

Michèle Nuijten is Assistant Professor at the Meta-Research Center of the Department of Methodology and Statistics of Tilburg University, where she studies reproducibility and replicability in psychology. She received her PhD in Methodology and Statistics at Tilburg University in 2018. Her PhD thesis, titled "Research on Research: A Meta-Scientific Study of Problems and Solutions in Psychological Science", was awarded the Tilburg University Dissertation Prize.

As part of her research, Nuijten co-developed the free tool statcheck; a "spellchecker" for statistics. The journals Psychological Science and the Journal of Experimental Social Psychology have made statcheck a standard element in their peer review process. Nuijten is involved in the Society for the Improvement of Psychological Science (SIPS), having been a member of the executive committee, and past-chair of the program committee. She is part of the Program Committee Replication Research of The Netherlands Organization for Scientific Research, advising them on distributing funding for replication research. Nuijten is also part of the Science Committee of the Tilburg School of Social and Behavioral Sciences, where she helps to develop guidelines on data handling and methods reporting.

Han L. J. van der Maas

Professor of Psychological Methods, Faculty of Social and Behavioural Sciences, University of Amsterdam

Science under pressure: questionable publishing practices

In the past ten years, many things in science have changed for the better: Research integrity is high on the agenda. One indication is our new vocabulary. We now talk about questionable research practices (QRP's), pre-registration, open science, etc. These concepts were unknown to most researchers ten years ago and are now part of the standard curriculum in first year undergraduate courses.

That does not mean that we are fine again and can proceed with 'normal' science. QRP's are still a significant threat because the incentives for implementing these practices have not been changed. Scientist need to (1) compete, (2) publish (a lot) in high impact journals, and (3) to obtain grant money (and lots of it). Of course, it is impossible to fundamentally change these incentives, and to some degree there isn't anything wrong with these goals.

Authors should decide on what they publish instead of journal editors

However, my proposition for this essay is that the organization of scientific publishing is problematic when attempting to reduce QRP's. I posit that for effectively reducing QRP's we have to re-organize the outdated publication process in science. The key idea is radical: Authors should decide on what they publish instead of journal editors. This will remove a main incentive for researchers to use QRP's.

I will clarify this proposition. By "outdated", I mean that our publishing model is based on printing limitations. For a long time, journals could only publish a very limited number of papers, as journals had to be printed and sent around the world by post. This limitation was eliminated about 25 years ago, though somehow we did not modernize the publication process accordingly.

The current publication process suffers from many problems:

a) It is still difficult to publish null results; that is, outcomes not showing expected effects or associations. This invites the use of QRP's. Publication bias is a huge problem.
b) In writing up our paper we avoid being self-critical, as it will often backfire in the review process.
c) We have to select a journal from an endless list of monodisciplinary titles. Frequently, papers are rejected because they do not fit the scope of the journal. As scientists nowadays rarely read journals (google scholar is my journal), this makes no sense.
d) I don't know a journal editor that does not complain about finding reviewers.
e) Reviewing is thankless and usually unaccredited work voluntarily performed by scientists for the sole purpose of improving the quality of a commercial journal.
f) Editorial comments and reviews do improve the quality of publications, but editors and reviewers suffer from bias.
g) The potential selective effect of editors deciding to accept/reject specific publications is ineffective, as scientists keep submitting their papers to different journals until publication. This means that papers are often reviewed and rejected multiple times before publication.
h) Endlessly trying to have the paper published makes the publication process extremely slow, expensive, frustrating, and inefficient, which is especially problematic for young researchers on temporary contracts.
i) The selective effect is very limited. The number of scientific publications doubles per decade.
j) There is huge problem with ownerships of journal papers, and in general the publication process in science is very costly. The scientific publishing system is kidnapped by commercial publishers.

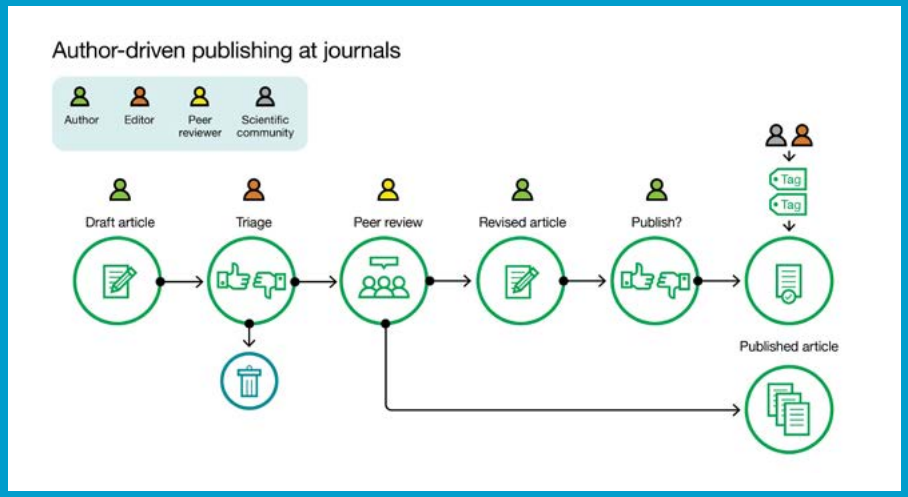
A radical change needed

What is the alternative? Improvements have been proposed and applied with some effectiveness, but a radical change is needed. I would prefer a model where: (1) we submit papers to a limited set of general journals, (2) technical editors quickly check if a submission meets basic requirements, (3) authors pay article processing charges (APC) for the reviewing part of this process, and (4) the editor collects reviews after which the authors write a revision. Following this review process, the authors themselves may ask for another round of reviews but also take a decision: (5) to publish or not, and (6) the reviews are published with the paper so that reviewing becomes a visible and valued contribution to science and expert criticism becomes accessible.

My proposal is not new. Stern and O'Shea proposed this idea in 2018 and described their ideas in a recent PloS paper.[1] They propose that: "authors decide when and what to publish; peer review reports are published, either anonymously or with attribution; and curation occurs after publication, incorporating community feedback and expert judgment to select articles for target audiences and to evaluate whether scientific work has stood the test of time."
Figure 1 displays this process.

[1] Stern, B. M., & O'Shea, E. K. (2019). A proposal for the future of scientific publishing in the life sciences. *PLoS biology*, *17*(2), e3000116.

Figure 1: The author decides (figure by *By Bodo M. Stern and Erin K. O'Shea*, https://asapbio.org/digital-age)



The initial responses to this proposal of author driven publishing (ADP) are usually negative for three reasons. First, the review and editorial selection processes are expected to operate as a system of quality control. Reviews generally improve papers, and reviewers and editors prevent the publication of low-quality papers. Indeed, I believe that reviews are valuable, and they still play a crucial role in ADP. As reviews are published with the paper, compared with the traditional review format, their role becomes more important. With regard to quality control, we can deduce that the current system utterly failed to prevent the replication crisis and the abundant use of QRP's in science. I would even claim that it is partly responsible for this crisis.

The second objection against ADP is that we will see a further increase in the number of publications. However, I would not expect more publications, as scientists would be judged on impact, rather than on the number of publications. What would be the point of publishing 20 papers per year if everybody could do so? Rather, with ADP publishing that many papers would present a risk to authors. Why would researchers read any papers by person X if these papers are merely copies of each other, if they only contain quick underpowered studies, etc.? In the assessment of researchers, ADP will shift from a focus on the number of publications to a focus on citations and citations per publication. In the long term, I predict a paradoxical effect; because it is easier to publish, the number of publications per scientist will decrease. A third objection is rewarding the transition to ADP, which I will discuss at the end of this essay.

Does ADP solve the problems of our current publishing system? Let me check the list:

a) It is up to the authors to publish null results. The incentive to use QRP's is reduced.
b) Self-criticism will be less risky as there will no longer be a risk that the paper is rejected. In my view, it is really important to invite authors to be self-critical as they are often the main experts.
c) With the abolishment of the traditional editor, we will also see the end of monodisciplinary journals. These journals will be replaced by general non-commercial journals and completely new initiatives that I discuss below.
d) When reviews are published and become a visible contribution to science, I would review more papers.
e) Many journals already experiment with publishing reviews.[2] One interesting initiative is Publons, which collects information about peer review.
f) In ADP, the contribution of the reviews remains, but the enormous bias in the decision to publish disappears.
g) The miserable and time-wasting practice of serial submissions for papers ends.
h) Young researchers on temporary contracts, for whom the current system can be a career breaker, have greater chances to compete with staff on fixed contracts, for whom the current system is merely a nuisance.
i) As previously stated, I expect a deceleration in the growth of the number of publications.
j) ADP ends the era of commercial publishers in science. The reduced role of editors, and thus traditional journals, will cause these journals being replaced by innovative publishing concepts.

Open access breaks the monopoly of commercial publishers

Back to the objections. Is it realistic that authors will first attempt to publish in high-impact traditional journals and only when they fail dump their papers in an

[2] A recent review did not show any strong effects on willingness to review.
Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., & Squazzoni, F. (2019). The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature communications*, *10*(1), 322.

ADP journal? Is a smooth transition to ADP possible? The answer is yes and this transition is all but hypothetical. We are already witnessing the start of this new era.

The first important trend is open science and open access, increasingly enforced by grant agencies. Open access breaks the monopoly of commercial publishers. Both scientist and publishers feel the need for new concepts. Unfortunately, for commercial publishers their existence is not a part of the new concept.

Second, some journals already successfully apply the ADP. An example is *Frontiers*. The impact statistics of *Frontiers* are remarkably high, they are fast, and they publish almost everything.3 The main disadvantage is the APC's, which could be much lower if universities took the publishing role in their own hands. A serious attempt to do so are the university journals.4

Third, a radical form of ADP is already in place. In the past decade, we witnessed a rapid increase in popularity of archive publications. The most famous is arXiv.org, a repository of electronic preprints, with over 15,000 submission per month. Cornwell University maintains arXiv.org as a service to the field for a fraction of the costs of traditional publishing. arXiv.org has inspired many similar initiatives.

Fourth, different systems for post-publication article assessment have emerged. One option is to start overlay journals (see https://www.episciences.org for an example). These open access academic journals do not produce their own contents, but select from texts that are already freely available from preprint servers. An even simpler form of post-publication article evaluation take place in the form of badges or tags. These could be badges for sharing data, pre-registered studies, etc.

Although most of these developments are still in their initial phase, they already affect our publishing habits substantially. Among young researchers in Psychology, it is common sense to publish preprints on PsyArXiv, where some papers are already well cited before they are published in traditional journals. This is a rapid change, as this popular website has only existed since 2016.
I see this revolution in publishing as an inevitable consequence of the open science movement, and I think that it will significantly help in promoting good research practices in science.

3 http://www.colinphillips.net/how-to-create-a-top-journal-by-accepting-almost-everything/
4 https://universityjournals.eu/

Han L. J. van der Maas (1966) is Professor of Psychological Methods, Faculty Professor of Complex systems in the Social and Behavioral Sciences, and director of the Research institute of Psychology at the University of Amsterdam (UvA).

Van der Maas (1966) studied Psychology at the University of Amsterdam and received his master (Psychological Methods) in 1989 and his PhD (Developmental Psychology) in 1993 for research on methods for the analysis of phase transitions in cognitive development. After a five year KNAW fellowship, he joined the faculty of the developmental group of the University of Amsterdam, first as associate professor and in 2003 as full professor. In 2005 he became professor there and chair of Psychological Methods group at the University.

In 2009 he founded Oefenweb.nl, a spin-off company of the UvA-holding, selling a unique game-based web-based adaptive child monitoring system. He is scientific director of Oefenweb.nl. His general research theme is the formalization and testing of psychological theories in areas such as cognition, expertise, development, attitudes and intelligence.

Anne de Vries

PhD candidate at Tilburg Law School,
Tilburg University

Science under pressure: the PhD premium – quantity over quality[1]

[1] This essay reflects the personal opinion of the author and is meant to stimulate further debate on the topic.

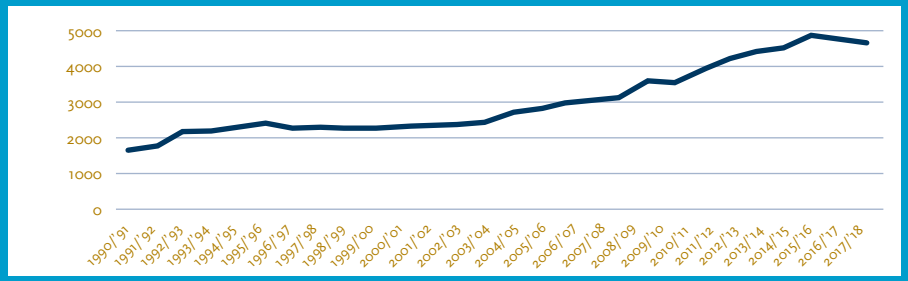
PhD dissertations are an important part of the research output of Dutch universities. However, in this essay I will argue that the quality of our PhD system is under pressure due to the PhD premium (in Dutch: de *promotiepremie*). This premium is a financial incentive the Ministry of Education, Culture and Science provides to the Dutch universities for each completed PhD dissertation. The premium steers towards maximizing the *number* of PhD dissertations instead of focusing on maximizing the *quality* of the PhD research. The PhD premium can make supervising PhD candidates a lucrative business. This may even lead to questionable research practices being tolerated or even institutionalized if these practices increase the number of PhD dissertations. Therefore, I will propose an alternative to the current PhD premium that could eliminate this potentially perverse incentive.

A brief history: some facts and figures

In 1993, the Ministry introduced the PhD premium. Until 2009, the premium was considerably higher for PhD dissertations in the exact, technical and medical sciences than in the other research areas. In 2009, the PhD premium was set at € 95.000 for all disciplines.[2]

Since the early 1990s, the annual number of completed PhD dissertations at the 14 Dutch universities has almost doubled from 2,478 dissertations in 1999/2000 to 4,781 in 2017/2018. The graph shows a particular sharp increase since 2002/2003, and a small decrease in 2017/2018:

PhD dissertations per year 1990/91-2017/18
Source: https://opendata.cbs.nl/statline/, data set 'Wetenschappelijk onderwijs; gepromoveerden aan universiteiten 1990/'91-2015/'16' and 'Wetenschappelijk onderwijs; promoties, studierichting'



[2] H. Vossensteyn, H. De Boer & B. J loed, *Chronologisch overzicht van ontwikkelingen in de bekostigingssystematiek voor het Nederlandse hoger onderwijs*, CHEPS: Deventer 2017.

Already in 1994, Frans W. Saris, a Dutch professor in physics and later the dean of his School, predicted that the PhD premium would lead to a rise in PhD dissertations.[3]

The strong increase of the number of PhD dissertations had a disruptive effect on the public funding of academic research. It meant that a larger part of the structural funding for academic research went to PhD premiums. According to the Minister of Education, Culture and Science this caused an unwanted steering effect.[4] Therefore, in 2016 the total budget available for PhD premiums was maximized to 20% of the total (structural) government funding of academic research. As a result, the PhD premium has dropped to € 80.329 per dissertation in 2019.

The PhD premium: a profitable model?

The Association of Dutch Universities (VSNU) claims that the PhD premium does not promote quantity at the expense of quality, because the costs for supervising PhD candidates are higher than the PhD premium.[5] This would mean that the strong increase of the number of PhD dissertations in the last two decades only increased costs for universities. However, this argument does not distinguish between the various types of PhD candidates.

Indeed, the costs for employing a PhD candidate, which are approximately € 200.000, are much higher than the PhD premium, but VSNU figures show that the number of employed PhD candidates has stayed almost the same since 2008, while the number of PhD dissertations increased by almost 50% from 3,214 in 2007/2008 to 4,781 in 2017/2018.[6] Most likely, the strong increase of the number of PhD dissertations comes from non-employed PhD candidates, who do not receive a salary and thus are much 'cheaper' for universities than employed PhD candidates.

Common types of non-employed PhD candidates are international scholarship PhD candidates and external PhD candidates. The first category receives a scholarship from a foreign organization. This means that a university's costs are limited to providing supervision, office and research facilities, graduate courses and sometimes a research budget for traveling and attending conferences. In addition, some universities, such as the universities of Delft and Groningen, provide

[3] F.W. Sairs, 'Universitair leven Frans W. Saris De promotiepremie', De Gids 1994-157.
[4] *Wetenschapsvisie 2025*, p. 22.
[5] *Factsheet gepromoveerden. Gepromoveerden van belang voor Nederland,* VSNU: Den Haag 2014
[6] See the second graphic on https://www.vsnu.nl/f_c_promovendi.html

scholarship PhD candidates with a monthly financial top-up to their (often low) scholarship. To my knowledge, there is no reliable information available about the average costs for universities facilitating international scholarship PhD candidates. Presumably, since universities do not need to pay for salary, pension or other social benefits, the costs for these PhD candidates are much lower than for regular PhD employees.

Even less costly are external PhD candidates. External PhD candidates conduct their research in their own time and/or in their employer's time. For example, an academically educated financial specialist employed at a bank may have the ambition to obtain a doctoral degree in financial law next to his job at the bank. The costs for the university are limited to supervision and in the future, based on a plan launched by the VSNU in 2019, some basic facilities (e.g., e-mail, access to the library). Access to additional facilities, such as access to graduate courses, is at the discretion of the universities.[7] Again, solid information about the costs of supervising external PhD candidates is lacking, but it is likely that external PhD candidates can indeed be profitable. In particular, in 2017 and 2018, radio show Argos branded our university a "promotion factory" because of the large numbers of external PhD candidates some professors supervised. According to Argos some of the supervisors received a financial reward for each completed PhD dissertation, which suggests a direct financial interest in the completion of dissertations. I consider these incidents a symptom of a structural problem in the current funding system. During my term as the President of the national PhD council (PNN), at different universities I noticed practices seemingly aimed at increasing the number of PhD dissertations. In particular, when external PhD candidates pay an annual fee for their supervision (sometimes as high as € 12.000) that covers part (if not all) of the costs, external PhD candidates can become even more lucrative.

Is profit making still an option, now that the PhD premium is maximized at 20% of the structural research budget of universities? This is a valid question, since the PhD premium per dissertation will be lower when the number of PhD dissertations in the Netherlands increases. A lower PhD premium will also lead to a weaker perverse incentive. However, the incentive on quantity rather than quality persists, because the maximum of 20% of the total annual budget applies to all Dutch universities *combined*. This means that universities may still compete for the biggest piece of the pie, even though that pie has a maximum size. Prof. van Gestel, member of the Tilburg Scientific Integrity Committee, noticed that this leads to privatizing of profits and communizing of costs.[8] This means that the potential gain (i.e., more PhD premiums) of increasing the number of PhD dissertations a particular university realizes is accompanied by the potential loss (i.e., a lower PhD premium) that is spread over all universities. I think that this system does not promote responsible behavior unless universities start to cooperate for the greater good rather than compete. So far, I have seen very little evidence of such cooperation when it comes to putting a stop on the universities' growth ambitions.[9]

System focused on quantity

These concerns are not new. During my term as the president of PNN, I voiced similar concerns.[10] I was not the first,[11] and not the last.[12] Some parties are more positive about the current system than I am though.[13] In 2012, the Inspection for Education noted that there are generally enough warranties to counter the effect of the PhD premium. On the other hand, the Inspection warned that universities should pay more attention to eliminating potential perverse incentives in their internal financial allocation models.[14] To my surprise, a year earlier (before I was president), PNN argued that the PhD premium should be transferred directly to research groups and supervisors, as – according to PNN – already happens in some cases. The idea was that, to increase the chances of successfully finishing a PhD trajectory, all deans should follow this example by giving the promotor such a direct financial incentive.[15] In 2018, a Dutch university suggested transferring PhD premiums to research groups and providing the PhD candidate with a financial reward if they finish their PhD dissertation within four years.[16] I think that other incentives that promote quantity, such as a minimum number of PhD dissertations per professor per annum, are equally undesirable. Such incentives are especially risky in alfa and gamma sciences, where research grants are generally harder to

[7] *Een gezonde praktijk in het Nederlandse promotiestelsel,* VSNU: Den Haag 2019.

[8] R. van Gestel, 'Buiten promovendi gerekend', *Ars Aequi* 2019, p. 242-247.

[9] A similar observation could be made in respect of the anually growing number of students at Dutch universities.

[10] A. de Vries & K. Hoyer, 'Ontneem hoogleraar promotierecht bij ondermaatse begeleiding De promotiefabriek: wachten op een volgend schandaal', *ScienceGuide*, 14 September 2018.

[11] See a.o.: 'De perverse promotiepremie mag wel worden afgeschaft', *NRC Handelsblad* 21 April 2011; B. Sprecher, 'Het promovendi-overschot', *Mare* 15 October 2015.

[12] K. Marée, J. Been & B. Hekkema, 'Voor elke promotie krijgt de universiteit geld. En de promovendus betaalt de prijs', *De Correspondent*, 9 May 2019; Prof. Paola Gori Giorgi cited in: J. Chaudron, 'De promovendus die tegen het zere been van de hoogleraar schopt, vliegt eruit', *Trouw* 11 May 2019

[13] For example, Lawson argues that the PhD premium is not a perverse incentive, because it is outweighed by the costs of PhD research (R.A. Lawson, 'Promoties onder het vergrootglas', *Ars Aequi* 2017(11), p. 886-892). However, he does not distinguish between external and internal PhD candidates.

[14] *Verkenning naar de kwaliteitsborging van promotietrajecten en promoties*, Inspectie OCW, October 2012.

[15] M.M. Meijer, Behoud Talent! Een rapportage over de verschillende aspecten die een rol spelen bij de begeleiding van promovendi, PNN 2011, p. 16, p. 42.

[16] C. Boomsma, FSE wants bonus for fast PhD students, UKrant 9 April 2018.

obtain and where the number of annual PhD dissertations on average is considerably lower than in the beta and medical domain.[17] In particular in case of limited financial resources, quantity incentives may unintendedly promote recruiting "low cost" external PhD candidates whereas the focus should be with the PhD project's content and relevance to the School's research program, the supervisor's expertise and the candidate's qualities. In sum, all steering mechanisms that are meant to increase the PhD output are inherently risky in a system that is already heavily focused on quantity.

More is less?

One may argue that an incentive that steers towards quantity does not necessarily steer away from quality. However, there is a clear tension between the two goals. If an external PhD has paid thousands of euro's for receiving PhD supervision, doesn't this put pressure on the supervisors to accept the manuscript at some point, even if it is below standard? If a professor supervises dozens of PhD candidates simultaneously, can he or she still properly supervise the quality of the research being done? If a research group or an individual has a direct financial interest in the completion of a PhD project, does this not interfere with quality requirements that ask for patience rather than pressure? If a School or a research group becomes dependent on external PhD candidates for generating income, will it remain critical with respect to the quality of the research?

I do not believe that there is one answer to each of these questions. Many academics are able to maintain their scientific integrity, also when under pressure to produce more PhD dissertations, but some may not and this may be true in particular in a system that rewards quantity of output. This is the essence of this issue: the PhD premium may amplify a rather persistent quantity-driven culture at universities. As such, the PhD premium may contribute to an unhealthy environment where science becomes profit-driven. This steers away from what academic research should be about an independent quest for knowledge.

Solutions

Some argue that the current system provides sufficient checks and balances to safeguard the quality of PhD dissertations. For example, for each dissertation the PhD committee carries out a final quality check. However, because the PhD supervisors have a strong say in the selection of the members of the PhD committee, the PhD committee's judgment should not be the only quality check.

[17] Likewise: R.A. Lawson, 'Promoties onder het vergrootglas', *Ars Aequi* 2017(11), p. 887.

The VSNU seems to support this view to some extent. In 2019, the VSNU presented several recommendations for safeguarding the quality of PhD trajectories, also for external PhD candidates. For example, PhD candidates should be registered with a Graduate School at least two years before the defence of the thesis, PhD committees should always consist of at least one, and preferably two, members from another university, and each PhD candidate should have at least two supervisors.[18] Another option, not proposed by the VSNU, would be to regulate the distribution of the PhD premiums within universities in a way that perverse incentives are abolished. For example, a university could pay a lump sum for PhD projects to research groups and departments rather than passing on the PhD premium if a dissertation is completed.

The VSNU recommendations are necessary, but are less suited to fix a financial system that favors quantity over quality. It is like prescribing medicine to a patient for an unhealthy diet, rather than changing the diet. In a system that heavily promotes quantity, the proposed checks and balances require extensive supervision and control to make sure that they do not become a paper tiger. Therefore, I believe that the root causes of the problem should be eliminated. One of the root causes lies with the PhD premium that rewards quantity rather than quality and allows universities to make a profit from uncritically increasing the numbers of certain "low-cost" PhD candidates.

If we accept that making a profit should not be a reason for conducting independent academic research, the solution is simple: change the system in a way that no profit can be made on PhD dissertations. A logical solution has already been proposed in 2018 by the national PhD Council (PNN)[19] and in 2019 (tentatively) by the Committee van Rijn when it advised restructuring the funding of universities.[20] The proposal is to align the PhD premium with the actual (average) costs that a university makes in relation to the different types of PhD candidates. This means assigning a higher premium for PhD employees and a lower premium for external PhD candidates. A third category could be introduced for international scholarship PhD candidates. If a PhD candidate has already covered the costs for his own supervision by paying a fee, a PhD premium should not be awarded at all. In October of 2019, the Ministry of Education, Culture and Science suggested that it would investigate the real costs of the different types of PhD candidates

[18] Een gezonde praktijk in het Nederlandse promotiestelsel, VSNU: Den Haag 2019.
[19] Actieplan: Naar een Gezond Promotiesysteem, PNN, 16 October 2018 (I was the president of PNN around that time).
[20] *Adviesrapport bekostiging hoger onderwijs 'Wissels om', Adviescommissie Bekostiging Hoger Onderwijs en Onderzoek*, 19 July 2019

and how the funding of universities can be better aligned to those costs.[21] A differentiated PhD premium is not entirely new, and two Dutch universities already use a differentiated PhD premium for low versus high-cost PhD projects, the latter leading to a 60-70% higher 'internal PhD premium'.[22]

I expect that the proposed change in the PhD premium stimulates a more balanced investment in young researchers and softens the high investments that universities make to employ a PhD candidate. Simultaneously, the change prevents profit-driven practices in relation to external and scholarship PhD candidates. This allows universities to focus on what really matters, selecting promising young researchers to conduct sound and independent PhD research.

Anne de Vries is a PhD researcher in private law at Tilburg Law School and a member of the Tilburg Environmental Law Team. In her research she focuses on contract and liability law and environmental law, and regularly speaks, advises and publishes in these fields. As a member of the Tilburg PhD platform (TiPP) and the University Council, she represented the interests of the PhD community at Tilburg University. In 2018-2019 she was the president of the National PhD Network (PNN). In that capacity she raised awareness in the national media and politics about problems faced by PhD candidates. Among other things, she strongly criticized the PhD premium because for its strong emphasis on quantity of research. Anne writes columns on academic life and culture for Univers, the independent magazine of Tilburg University and was elected as one of ten top opinion makers of the academic year 2018-19 by news platform ScienceGuide.

[21] *Opdrachtbeschrijving onderzoek toereikendheid en doelmatigheid macrobudget en kosten(-toerekening) middelbaar beroepsonderwijs, hoger onderwijs en (wetenschappelijk en praktijkgericht) onderzoek*, Ministry of OCW, 18 October 2019.

[22] B. Jongbloed et. al., *Bekostiging van het Nederlandse hoger onderwijs: kostendeterminanten en varianten*, CHEPS: Deventer 2018 p. 50. For PDEngs, which only take two years, university also receive a lower premium.

Lex Bouter

Professor of Methodology and Integrity,
Amsterdam University Medical Centers and
Faculty of Humanities, Vrije Universiteit Amsterdam

What research institutes can do to foster research integrity[1]

[1] A slightly different version of this contribution will appear in *Science and Engineering Ethics*, https://doi.org/10.1007/s11948-020-00178-5

It is a real honour and a great pleasure to be invited to lecture on the ninety-second birthday of this interesting university. Explaining to the academic community of Tilburg University what they should do to foster research integrity is to some extent like 'preaching to the converted'. The data fabrication incident in this university that shocked many within and outside Academia turned out to be a blessing in disguise as well. Due to the adequate response by rector Philip Eijlander and the excellent analyses2 of how this could have happened, the Stapel case served as a call to arms for this and many other universities. As your current rector Klaas Sijtsma said at the 5th World Conference on Research Integrity in Amsterdam: 'Never waste a good crisis'[3].

Traditionally research integrity has focussed on the prevention, identification and handling the three deadly sins of scientific and scholarly research: fabrication, falsification and plagiarism[4]. In recent years attention has shifted to the lesser breaches of research integrity that are commonly referred to as questionable research practices or QRPs[5,6]. The idea is that these are much more prevalent and thus collectively do more harm to the validity of and the trust in the results of research. Examples are selective reporting, P-hacking, and hypothesising after the results are known or HARK-ing. In an excellent paper from the Meta-Research Center of this university, 34 QRPs are identified as researcher degrees of freedom that should be avoided in hypothesis-testing research[7]. QRPs have in common that they can help to make study results more exciting, more positive and more statistically significant, which in its turn increases the likelihood to be accepted by a high impact journal, to get many citations, and to obtain the next grant or academic tenure.

[2] Levelt, Noort and Drenth Committees. Flawed science: the fraudulent research practices of social psychologist Diederik Stapel. Tilburg, 2012 (https://www.rug.nl/about-us/news-and-events/news/news2012/stapel-eindrapport-eng.pdf)

[3] Sijtsma K. Never waste a good crisis: towards responsible data management. Amsterdam, 2017 (https://www.youtube.com/watch?v=bddRx-LN8lo and https://www.wcrif.org/images/2017/documents/2.%20Tuesday%20May%2030,%202017/1.%20Aula/K.%20Sijtsma%20%20Never%20waste%20a%20good%20crisis%20-%20Towards%20responsible%20data%20management.pdf)

[4] National Academies of Sciences, Engineering and Medicine. Fostering integrity in research. Washington, 2017 (https://www.nap.edu/download/21896#)

[5] Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Research Integrity and Peer Review 2016; 1: 17 (http://rdcu.be/mPZT)

[6] Haven T, Tijdink J, Pasman HJ, Widdershoven G, ter Riet G, Bouter L. Do research misbehaviours differ between disciplinary fields? A mixed methods study among academic researchers in Amsterdam. OSF preprints 4 April 2019 (https://osf.io/7d4qz)

[7] Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. Frontiers of Psychology 2016; 7: 1832 (https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01832/full)

Research Integrity Promotion Plan

Almost all researchers want to deliver good quality science, to avoid QRPs, and to follow their moral compass to steer a course of research integrity. Like any compass, the functioning of a moral compass depends on its quality and on external factors. The quality is determined by the virtuousness of the individual at issue. Major external factors that can corrupt the moral compass concern the local research climate and the perverse incentives of the science system as a whole. Researchers need help from their research institute in optimising the functioning of their moral compass. That help involves adequate education and skills training, good facilities and expert help, and clear codes and procedures.

The Netherlands Code of Conduct for Research Integrity specifies 61 standards for good research that mirror in fact as many QRPs to be avoided[8]. A unique feature of the code is that it also contains a chapter on the duties of care research institutes have to empower their research staff to steer away from QRPs. In other words, research institutes need to have a Research Integrity Promotion Plan. The Horizon 2020 funded consortium Standard Operating Procedures for Research Integrity (SOPs4RI)[9] offers research institutes help to formulate this plan. Having such a plan operational likely will become a contractual obligation for institutes accepting grants from the next EU framework program Horizon Europe.

The idea is that the Research Integrity Promotion Plan explains what the research institute sets out to do - in the context of its mission, disciplinary focus and type of research it performs - to promote research integrity. The plan needs to cover a set of mandatory topics and will typically describe a mix of education programs, codes, manuals, policy measures, regulations, facilities, audit schemes, and support systems. SOPs4RI will produce a toolbox filled with Standard Operating Procedures (SOPs) and guidelines that can help research institutes to formulate their Research Integrity Promotion Plan[10,11,12,13]. The difference between a SOP and a guideline is gradual, with SOPs being more strict step-by-step recipes and guidelines offering

[8] Netherlands code of conduct on research integrity. The Hague, 2018 (https://www.vsnu.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf)

[9] Standard Operating Procedures for Research Integrity (SOPs4RI) (https://www.sops4ri.eu/)

[10] ORI. Guidelines for institutes and whistleblowers (https://ori.hhs.gov/ori-guidelines-institutions-and-whistleblowers-responding-possible-retaliation-against)

[11] ENRIO. Recommendations for the investigation for research misconduct (http://eneri.eu/wp-content/uploads/2019/03/INV-Handbook_ENRIO_web_final.pdf)

[12] PRINTIGER. Working with research integrity: guidance for research performing organizations (https://link.springer.com/content/pdf/10.1007%2Fs11948-018-0034-4.pdf)

[13] ENERI. Manual of research integrity and ethics (http://eneri.eu/wp-content/uploads/2018/10/ENERI-e-Manual.pdf)

some freedom of choice. It is important to make this not another box ticking exercise, but to ensure that researchers appreciate and use the guidance offered by their institute.

Arguably, one of the most important things research institutes should do is to avoid perverse incentives in assessing researchers for career advancement. The current dominant focus on bibliometric indicators derived from publication and citation counts sends a strong message that only these things really matter[14]. During recent years, the myopic use of quantitative indicators in research evaluations has been criticised. This led to initiatives like the Leiden Manifesto[15] and the San Francisco Declaration on Research Assessment (DORA)[16]. In line with this, the Hong Kong Principles for assessing researchers[17] were formulated and endorsed at the 6th World Conference on Research Integrity last June. These principles will help research institutes that adopt them to minimise perverse incentives that invite to engage in questionable research practices or worse.

The Hong Kong Principles are chosen with a view to explicitly recognise and reward researchers for behaviour that leads to trustworthy research by avoiding QRPs. The principles have been developed with the idea in mind that their implementation could assist in how researchers are assessed for career advancement with a focus on behaviours that strengthen research integrity. We formulated five principles: assess responsible research practices, value complete reporting, reward the practice of open science, acknowledge a broad range of research activities, and recognise essential other tasks like peer review and mentoring. For each principle, we provide a rationale for its inclusion and give examples of research institutes where these principles are already being adopted.

[14] Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN. Assessing scientists for hiring, promotion, and tenure. PLoS Biol 2018; 16: e2004089 (https://journals.plos.org/plosbiology/article/file?rev=2&id=10.1371/journal.pbio.2004089&type=printable)

[15] Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. The Leiden Manifesto for research metrics. Nature 2015; 520: 429-31 (http://www.leidenmanifesto.org/)

[16] San Francisco Declaration on Research Assessment (https://sfdora.org/)

[17] Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, Coriat AM, Foeger N, Dirnagl U. The Hong Kong principles for assessing researchers: fostering research integrity. OSF Preprints September 17, 2019 (https://osf.io/m9abx)

Research institutes should make their research integrity policies as evidence-based as possible. In hindsight, it is difficult to understand why it took us so long in establishing a solid tradition in research on research integrity. That only started to happen recently and was fuelled by granting programs like the Horizon 2020 Science with and for Society (SwafS) calls for research ethics and research integrity[18]. In the Netherlands, the programs on Fostering Responsible Research Practices[19] and Replication Studies[20] contributed to the emerging field of research on research.

Invitation to participate in National Survey on Research Integrity

That being said there is still a lot we do not know about research integrity in Dutch research institutes. To fill this gap in May 2020 all researchers in Dutch universities and university medical centres will be invited to participate in the National Survey on Research Integrity led by a team that includes Professor Jelte Wicherts of this university. The survey is expected to provide valid and reliable knowledge on how often specific QRPs occur and what their underlying explanatory variables are. This will provide insights that help research institutes to improve their policies and to fulfil their duties of care in fostering research integrity better. Given the sensitivity of some of the questions, the survey will pay particular attention to ensuring the protection of the identity of the participants and their research institutions. The Randomised Response technique that will be used is expected to elicit a strong sense of trust in respondents because their answers can never be linked to them[21]. In addition, to keep the time to complete the survey short we make use of missingness by design.

Finally, I would like to say that there are many stakeholders with a responsibility to foster research integrity. First and foremost, the researchers themselves are responsible to behave well and to refrain from QRPs. Researchers should also be a good role model and help others to keep on track. Second – and that was the topic of my lecture – research institutes must empower researchers to act according to the standards of good research. In addition, funding agencies and scientific journals have important roles to play. There is no magic pill or a quick fix.

[18] European Commission. Horizon 2020: Science with and fore society (Swafs) (https://ec.europa.eu/research/swafs/index.cfm?pg=funding)

[19] ZonMw. Fostering Responsible Research Practices (https://www.zonmw.nl/en/research-and-results/fundamental-research/programmas/programme-detail/fostering-responsible-research-practices/)

[20] NWO. Replication studies (https://www.nwo.nl/en/funding/our-funding-instruments/sgw/replication-studies/replication-studies.html)

[21] Lentsvelt-Mulders GJLM, Hox JJ, van der Heijden PGM, Maas CJM. Meta-analysis of randomized response research: thirty-five years of validation. Sociological Methods and Research 2005; 33 : 319-48

The dilemmas and distractions researchers face are real and universal. We owe it to society to collaborate and do our utmost best to prevent QRPs and to foster research integrity.

Professor Lex M Bouter, PhD, has a tenured chair in Methodology and Integrity at the Department of Epidemiology and Biostatistics of the Amsterdam University Medical Centers and the Department of Philosophy of the Faculty of Humanities of the Vrije Universiteit. Before taking up the current position he was professor of Epidemiology since 1992. In 1988 he published a textbook on epidemiology, the seventh revised Dutch edition of which appeared in 2016 and the first English edition in 2018.

He served his university as its rector between 2006 and 2013. Subsequently, Professor Bouter prepared for a return to science during a sabbatical leave. In 2014 his tenured professorship was broadened to Methodology and Integrity. He is currently involved in teaching and research regarding responsible conduct of research. Professor Bouter has supervised 76 PhD students, of whom to date 16 were appointed as professor.

In 2017, Professor Bouter organized the 5th World Conference on Research Integrity in Amsterdam and became chair of the World Conferences on Research Integrity Foundation.

Credits

Realization: Division Marketing & Communication, Stijn van Kruijsdijk, Tilburg University
Design: De Studio | powered by Canon
Photographs: Maurice van den Bosch, (cover, 5, 15, 16-17, 41), Tagwaran/SS (p. 8-9), Sportoakimirka/SS (p. 24-25), Monster Ztudio/SS (p. 32-33), Kyrill Savenko/SS (p. 42-43).

Tilburg University
Warandelaan 2
5037 AB Tilburg
T +31 13 466 9111
www.tilburguniversity.edu

ISBN: 978-94-6167-415-9

www.tilburguniversity.edu