



RESEARCH ARTICLE

REVISED Prevalence of responsible research practices among academics in The Netherlands [version 2; peer review: 2 approved]

Gowri Gopalakrishna ¹, Jelte M. Wicherts ², Gerko Vink³, Ineke Stoop⁴, Olmo R. van den Akker², Gerben ter Riet⁵, Lex M. Bouter ^{1,6}

¹Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, The Netherlands

²Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

³Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

⁴The Netherlands Institute for Social Research, Den Haag, The Netherlands

⁵Center of Expertise Urban Vitality, Faculty of Health, Amsterdam University of Applied Science, Amsterdam, The Netherlands

⁶Department of Philosophy, Faculty of Humanities, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

v2 First published: 28 Apr 2022, 11:471
<https://doi.org/10.12688/f1000research.110664.1>

Latest published: 08 Aug 2022, 11:471
<https://doi.org/10.12688/f1000research.110664.2>

Abstract

Background: Traditionally, research integrity studies have focused on research misbehaviors and their explanations. Over time, attention has shifted towards preventing questionable research practices and promoting responsible ones. However, data on the prevalence of responsible research practices, especially open methods, open codes and open data and their underlying associative factors, remains scarce.

Methods: We conducted a web-based anonymized questionnaire, targeting all academic researchers working at or affiliated to a university or university medical center in The Netherlands, to investigate the prevalence and potential explanatory factors of 11 responsible research practices.

Results: A total of 6,813 academics completed the survey, the results of which show that prevalence of responsible practices differs substantially across disciplines and ranks, with 99 percent avoiding plagiarism in their work but less than 50 percent pre-registering a research protocol. Arts and humanities scholars as well as PhD candidates and junior researchers engaged less often in responsible research practices. Publication pressure negatively affected responsible practices, while mentoring, scientific norms subscription and funding pressure stimulated them.

Conclusions: Understanding the prevalence of responsible research practices across disciplines and ranks, as well as their associated explanatory factors, can help to systematically address disciplinary- and academic rank-specific obstacles, and thereby facilitate responsible conduct of research.

Open Peer Review

Approval Status

	1	2
version 2		
(revision)		
08 Aug 2022		
version 1		
28 Apr 2022		

1. **Elisabeth M. Bik** , Harbers Bik LLC, CA, USA

2. **Dorothy Vera Margaret Bishop** , University of Oxford, Oxford, UK

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Responsible conduct of research, Responsible research practices, Research integrity, Open science



This article is included in the **Research on Research, Policy & Culture** gateway.

Corresponding author: Gowri Gopalakrishna (g.gopalakrishna@amsterdamumc.nl)

Author roles: **Gopalakrishna G:** Investigation, Methodology, Project Administration, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Wicherts JM:** Conceptualization, Investigation, Methodology, Visualization, Writing – Review & Editing; **Vink G:** Methodology, Visualization, Writing – Review & Editing; **Stoop I:** Methodology, Visualization, Writing – Review & Editing; **van den Akker OR:** Investigation, Visualization, Writing – Review & Editing; **ter Riet G:** Conceptualization, Funding Acquisition, Methodology, Supervision, Visualization, Writing – Review & Editing; **Bouter LM:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was funded by the Netherlands Organisation for Health Research and Development (ZonMw) 20-22600-98-401, awarded to Lex. M. Bouter, and the Consolidator Grant 726361 (IMPROVE) from the European Research Council (ERC, <https://erc.europa.eu>), awarded to Jelte M. Wicherts and Olmo R. van den Akker.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Gopalakrishna G *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Gopalakrishna G, Wicherts JM, Vink G *et al.* **Prevalence of responsible research practices among academics in The Netherlands [version 2; peer review: 2 approved]** F1000Research 2022, 11:471 <https://doi.org/10.12688/f1000research.110664.2>

First published: 28 Apr 2022, 11:471 <https://doi.org/10.12688/f1000research.110664.1>

REVISED Amendments from Version 1

We have made changes to the manuscript in this version which include some minor grammatical errors as well as modifying some parts of the introduction and results section which Reviewer 1 felt were in duplicate to our previous published manuscript PLOS ONE 2022, DOI: 10.1371/journal.pone.0263023. Other changes include clarification of some of the explanatory factor scales used in the study as well as a minor correction in the decimal point of the response proportion reported. With respect to changes introduced as a result of the review by Reviewer 2, the main aspects include highlighting the importance of representativeness of our study sample to the target population, a few additions to the Discussion section explaining our rationale for recoding NA answers to 1, correction of a registration link in the Methods section and a sentence in the Discussion section to better emphasize that these findings are not casual in nature due to the cross sectional nature of the study.

Any further responses from the reviewers can be found at the end of the article

Introduction

There has been a clear rise in publications and efforts aimed at promoting research integrity in recent years,^{1–8} including pleas for the adoption and promotion of open science and other RRP s aimed at increasing the trustworthiness of research through increased transparency. In particular, open methods (*e.g.* preregistration of study protocols), open codes (for data analysis), open data (following the FAIR principles⁹) and open access (rendering publications available at no cost for users) play an important role.⁴

A number of explanatory factors such as scientific norms subscription, fair distribution of resources, rewards and recognitions (*i.e.* organizational justice), perceived pressures researchers face (*e.g.* competition, work, publication and funding pressures), and support by mentors have been suggested to be important in fostering high-quality research.^{10–12} So far however, the body of research on research integrity has focused largely on how to minimize QRPs but not so much on the empirical evidence on the adoption of foster RRP s and the associative underlying factors that may be important in encouraging the uptake of these practices. These studies typically have a narrow disciplinary scope covering few possible explanatory factors.^{10–17}

The National Survey on Research Integrity (NSRI)¹⁸ was designed to take a balanced, research-wide approach to report on the prevalence of RRP s, QRPs and research misconduct, in addition to exploring the potential explanatory factors associated with these behaviors in a single survey. The NSRI targeted the entire population of academic researchers in The Netherlands, across all disciplinary fields and academic ranks.

The objectives of the NSRI were:

- 1) to estimate prevalence of RRP s, QRPs and research misconduct, and
- 2) to study the association between possible explanatory factors and RRP s, QRPs and research misconduct.

In this paper we focus on the prevalence of RRP s and the explanatory factors that may help or hinder responsible conduct of research. Elsewhere we report on QRPs, research misconduct and their associated explanatory factors.¹⁹

Methods**Ethics approval**

This study was performed in accordance with guidelines and regulations from Amsterdam University Medical Centers and the Declaration of Helsinki. In addition, the Ethics Review Board of the School of Social and Behavioral Sciences of Tilburg University approved this study (Approval Number: RP274). The Dutch Medical Research Involving Human Subjects Act (WMO) was deemed not applicable to this study by the Institutional Review Board of the Amsterdam University Medical Centers (Reference Number: 2020.286).

The full NSRI study protocol, ethics approvals, complete data analysis plan and final dataset can be found on Open Science Framework.²⁰ Below we summarize the salient study features.

Study design

The NSRI was a cross-sectional study using a web-based anonymized questionnaire. All academic researchers working at or affiliated to at least one of 15 universities or seven university medical centers (UMCs) in The Netherlands were invited by email to participate. To be eligible, researchers had, on average, to do at least eight hours of research-related activities weekly, belong to Life and Medical Sciences, Social and Behavioural Sciences, Natural and Engineering sciences, or the

Arts and Humanities, and had to be a PhD candidate or junior researcher, postdoctoral researcher or assistant professor, or associate or full professor.

The survey was conducted by a trusted third party, Kantar Public,²¹ which is an international market research company that adheres to the ICC/ESOMAR International Code of standards.²² Kantar Public's sole responsibility was to send the survey invitations and reminders by email to our target population and send the anonymized dataset at the end of the data collection period to the research team.

Universities and UMCs that supported NSRI supplied Kantar Public with the email addresses of their eligible researchers. Email addresses for the other institutes were obtained through publicly available sources, such as university websites and PubMed.

Researchers' informed consent was sought through a first email invitation which contained the survey link, an explanation of NSRI's purpose and its identity protection measures. Starting the survey after this section on informed consent implied written consent. Consenting invitees could therefore immediately participate in the survey thereafter. NSRI was open for data collection for seven weeks, during which three reminder emails were sent to non-responders, at a one- to two-week interval period. Only after the full data analysis plan had been finalized and preregistered on the Open Science Framework²⁰ did Kantar Public send us the anonymized dataset containing individual responses.

Survey instrument

NSRI comprised four components: 11 QRPs, 11 RRP, two research misconduct questions on falsification and fabrication (FF) and 12 explanatory factor scales (75 questions). The survey started with a number of background questions to assess eligibility of respondents. These included questions on one's weekly average duration of research-related work, one's dominant field of research, academic rank, gender and whether one was conducting empirical research or not.²⁰

All respondents, regardless of their disciplinary field or academic rank, were presented with the same set of RRP, QRP and research misconduct questions on FF. These questions referred to the last three years in order to minimize recall bias. The 11 RRP were adapted from the Dutch Code of Conduct for Research Integrity 2018¹¹ and a survey among participants of the World Conferences on Research Integrity.²³ The first author of this manuscript created the initial formulations of the RRP which covered study design, data collection, reporting, open science practices, conflicts of interest and collaboration. These 11 RRP formulations were reviewed and agreed upon in two rounds: first within the NSRI core research team, and subsequently by an external group of multidisciplinary experts who formed the NSRI Steering Committee.¹⁸ All 11 RRP had a seven-point Likert scale ranging from 1 = never to 7 = always, in addition to a "not applicable" (NA) answer option.

The explanatory factors scales were based on psychometrically tested scales in the research integrity literature and focused on action-ability. Twelve were selected: scientific norms, peer norms, perceived work pressure, publication pressure, pressure due to dependence on funding, mentoring (responsible and survival), competitiveness of the research field, organizational justice (distributional and procedural), and likelihood of QRP detection by collaborators and reviewers.^{10-12,18,23-27} Some of the scales were incorporated into the NSRI questionnaire *verbatim*, others were adapted for our population or newly created (see *Extended data*: Table 5). With respect to the explanatory factor scale on work pressure, this can be defined as "the degree to which an academic has to work fast and hard, has a great deal to do, but with too little time" while publication pressure can be defined as the degree to which an academic has to publish in high-prestige journals in order to have a sustainable career.²⁸ *Extended data* Table 6²⁹ provides the full list of questions we included in the questionnaire.

Face validity of the NSRI questionnaire was tested in several ways. The QRP-related questions underwent extensive focus group testing in the instrument development stage of the project. Both the QRPs and RRP were further refined through several rounds of discussions with the core research team, with the project's Steering Committee and with an independent expert panel set up to review the entire questionnaire. Preliminary pilot testing was conducted for some of the explanatory factor scales, listed in *Extended Data* Table 5 along with the results of the factor analysis (factor loadings), whereas others were re-used from validated instruments, also detailed in Table 5 (*Extended data*).²⁹ Explanatory factor scales that are indicated as having been piloted will be reported on in future publications. In addition, internal consistency was tested and is reported as Cronbach's Alpha in *Extended Data* Table 1b. Inter-rater reliability was not applicable as the survey was self-administered; however test-retest reliability was not tested. Finally, the NSRI questionnaire's comprehensibility was pre-tested in cognitive interviews with 18 academics from different ranks and disciplines.³⁰ In summary, the comments centered around improvement in layout, such as the removal of an instructional video on the RR technique which was said to be redundant, improvement in the clarity of the instructions, and recommendations to emphasize certain

words in the questionnaire by using different fonts for improved clarity. The full report of the cognitive interview can be accessed at the Open Science Framework.²⁰

We used “missingness by design” to minimize survey completion time resulting in a total of 20 minutes on average for completion, each invitee received one of three random subsets of 50 explanatory factor items from the full set of 75 (see Table 5, *Extended data*²⁹). All explanatory factor items had seven-point Likert scales. In addition, the two perceived likelihood of QRP detection scales, the procedural organizational justice scale and the funding pressure scale had a NA answer option. There was no item non-response option as respondents had to either complete the full survey or withdraw.

Statistical analysis

We report on RRPp both in terms of prevalence and overall RRP mean. We operationalized prevalence as the proportion of participants that scored 5, 6 or 7 among the participants that deemed the RRP at issue applicable. Mean scores of individual RRPp only consider respondents that deemed the RRP to be applicable. In the multiple linear regression analysis, overall RRP mean was computed as the average score on the 11 RRPp, with the not-applicable scores recoded to 1 (*i.e.*, “never”). *Extended data*: Figures 2a to 2e show the distribution of responses, including the “not-applicable” category for the 11 RRPp.²⁹ The associations of the overall RRP mean with the five background characteristics (*Extended data*: Table 1a²⁹) and the explanatory factor scales were investigated with multiple linear regression.³¹

For the multivariate analyses of the explanatory factor scales, we used z-scores computed as the first principal component of the corresponding items.³² Missing explanatory factor item scores due to ‘not applicable’ answers were replaced by the mean z-score of the other items of the same scale. Multiple imputation with mice in R³² (version 4.0.3) was employed to deal with the missingness by design. Fifty complete data sets were generated by imputing the missing values using predictive mean matching.^{33,34} The linear regression models were fitted to each of the 50 data sets, and the results combined into a single inference. To incorporate uncertainty due to the nonresponse, the inferences were combined according to Rubin’s Rules.³⁵ All models contained all explanatory scales and the five background characteristics. The full statistical analysis plan, and analysis codes were preregistered on the Open Science Framework²⁰ including the following pre-specified subgroup analyses: field by rank, publication pressure by rank, funding pressure by rank, competition by disciplinary field, and detection (by reviewers or by collaborators) by disciplinary field.

Identity protection

Respondents’ identities were protected in accordance with the European General Data Protection Regulations (GDPR) and corresponding legislation in The Netherlands. In addition, we had Kantar Public conduct the survey to ensure that the email addresses of respondents were never handled by the research team. Kantar Public did not store respondents’ URLs and IP addresses. Only a fully anonymized dataset was sent to the research team upon closure of data collection and preregistration of the statistical analysis plan. Finally, we conducted analyses at aggregate levels only (*i.e.*, across disciplinary fields, gender, academic ranks, whether respondents conducted empirical research, and whether they came from NSRI supporting institutions).

Results

Descriptive analyses

A total of 63,778 emails were sent out (Figure 1) and 9,529 eligible respondents started the survey. Of these, 2,716 stopped the survey prematurely and 6,813 completed the survey. The response could only be reliably calculated for the eight supporting institutions (Figure 1a, *Extended data*²⁹) and was 21.1%. This percentage was derived by dividing total number of eligible individuals who opened the survey invitation (4414) by the total number of individuals who were invited from the eight supporting institutions (20879). *Extended Figure 1a*²⁹ provides a detailed explanation of this calculation.

Extended data: Table 1a gives a breakdown of all respondents stratified by background characteristics.²⁹ Male and female respondents were fairly equally split among the respondents. For the natural and engineering sciences, women accounted for 24.9% of respondents. In the highest academic rank of associate and full professors, women made up less than 30% of respondents (Table 1a, *Extended data*²⁹). Nearly 90% of all respondents are engaged in empirical research and about half (48%) come from the eight supporting institutions. Respondents from supporting and non-supporting institutions were fairly evenly distributed across disciplinary fields and academic ranks except for the natural and engineering sciences where less than one in four (23.5%) came from supporting institutions.

Extended data: Table 1a²⁹ describes the distribution of survey respondents by gender, disciplinary field, engagement in empirical research and if they were from a supporting institution or not. *Extended data* Table 1c²⁹ describes the distribution of the explanatory factor scales by disciplinary field and academic rank. The full description of these distributions can be read at Gopalakrishna *et al.*¹⁹

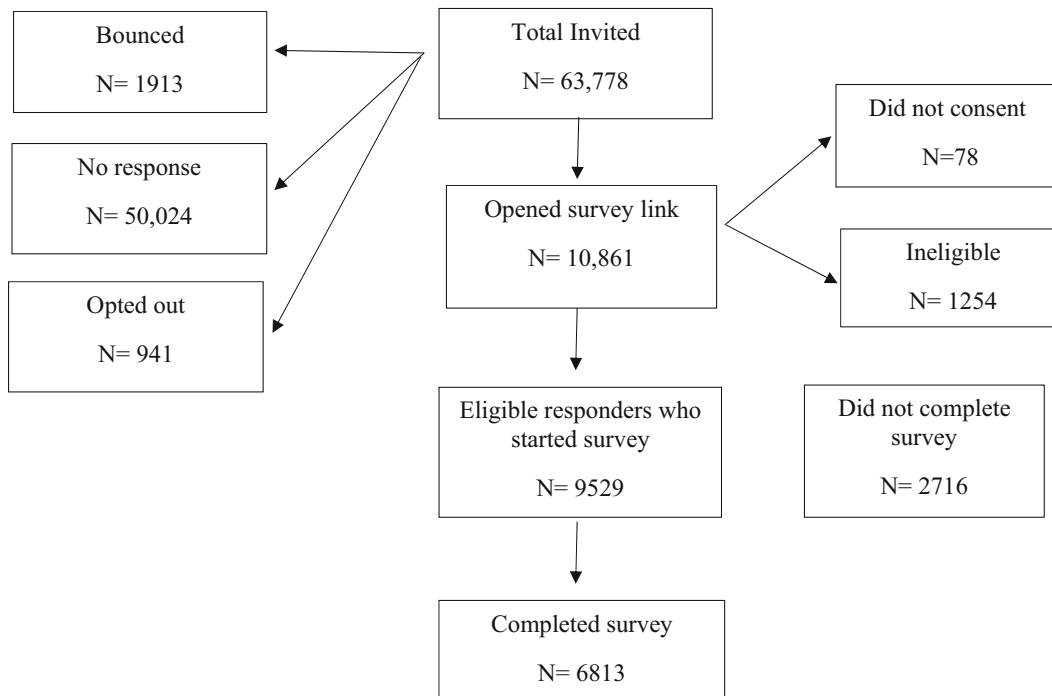


Figure 1. Flow chart of the survey.

Prevalence of RRP

The five most prevalent RRP (i.e. with a Likert scale score of 5, 6 or 7) had a prevalence range of 86.4% to 99% (Table 1; Figure 2, *Extended data*²⁹). Fair ordering of authorships (RRP 3) and preregistration of study protocols (RRP 6) showed the largest percentage differences between the Life and Medical Sciences and the Arts and Humanities (RRP 3: 75.7 vs 91.6% and RRP 6: 50.8% versus 30.2%). PhD candidates and junior researchers (74.2%) reported the lowest prevalence for RRP3 on fair allocation of authorships compared to associate and full professors (90.9%).

Extended data: Table 2 shows the discipline- and academic rank-specific prevalence of “not applicable” (NA) answers on the 11 RRP.²⁹ Arts and Humanities scholars reported the highest prevalence of NA for nine out of the 11 RRP. Similarly, across ranks, PhD candidates and junior researchers displayed the highest prevalence of NAs on nine out of the 11 RRP.

The four open science practices had an overall prevalence ranging from 42.8% to 75%: (i) following the FAIR principles (RRP 4: 75%); (ii) Publishing open access (RRP 8: 72.6%); (iii) Providing underlying data, computer codes, or syntaxes (RRP 10: 47.2%) and (iv) Preregistration of study protocols (RRP 6: 42.8%) (Table 1).

Surprisingly, the Arts and Humanities scholars had the highest prevalence for RRP 4 on following FAIR principles (84.6%). However, a closer look at RRP 4, reveals that this discipline also had the highest percentage of NA for RRP 4 (27.5%) (*Extended data*: Table 2²⁹). Life and Medical Sciences had the highest prevalence (50.8%) and the Arts and Humanities the lowest (30.2%) for preregistration of study protocols (RRP 6), where nearly 70% (67.8%) of the arts and humanities scholars rated RRP 6 as not applicable (Table 2, *Extended data*²⁹). Arts and Humanities scholars had the lowest prevalence (59.1%) and the Life and Medical Sciences the highest (75.1%) for publishing open access (RRP 8) (Table 1).

Regression analyses

Table 2a shows the results of the linear regression analysis for the five background characteristics while Table 2b shows the linear regression results for the explanatory factor scales.

Table 2a shows that the Arts and Humanities scholars had a significantly lower overall RRP mean score (-0.51; 95% CI -0.59, -0.42). Similarly, doing non-empirical research was associated with a significantly lower overall RRP mean score (-0.49; 95% CI -0.57, -0.42). Interestingly, females had a significantly lower RRP mean score than males

Table 1. Estimated prevalence (95% confidence intervals) of the 11 RRP's stratified by disciplinary field and academic rank.

RRP	Description (In the last three years ...)	Disciplinary field					Academic rank					Overall
		Life and medical sciences	Social and behavioral sciences	Natural and engineering sciences	Arts and humanities	PhD candidates and junior researchers	Postdocs and assistant professors	Associate and full professors				
RRP1	I disclosed who funded my studies and all my relevant financial and non-financial interests in my publications	98.6 (98.0,99.0)	96.2 (95.1,97.0)	94.0 (92.6,95.2)	93.2 (90.3,95.3)	94.0 (92.6,95.1)	97.3 (96.6,97.9)	97.5 (96.7,98.2)	96.5 (96.0,97.0)			
RRP2	I took steps to correct errors in my published work whenever I and/or peers provided valid reasons for such a correction	88.9 (87.1,90.5)	83.4 (80.7,85.8)	85.5 (82.9,87.8)	86.5 (82.0,90.0)	87.9 (85.5,89.9)	84.5 (82.5,86.4)	87.7 (85.6,89.6)	86.4 (85.2,87.6)			
RRP3	The allocation and ordering of authorships in my publications, were fair and in line with the standards of my discipline	75.7 (74.0,77.3)	84.1 (82.4,85.8)	86.6 (84.7,88.3)	91.6 (88.7,93.8)	74.2 (72.1,76.3)	79.6 (78.0,81.1)	90.9 (89.5,92.1)	81.8 (80.8,82.7)			
RRP4	I contributed, where appropriate, to making my research data findable, accessible, interoperable and reusable in accordance with the FAIR principles	74.8 (73.1,76.5)	70.7 (68.4,72.8)	77.5 (75.1,79.7)	84.6 (80.9,87.7)	75.2 (73.0,77.4)	73.6 (71.8,75.3)	76.6 (74.6,78.4)	75.0 (73.9,76.1)			
RRP5	I kept a comprehensive record of my research decisions throughout my studies.	57.2 (55.3,59.2)	56.5 (54.2,58.8)	54.0 (51.2,56.7)	57.1 (52.5,61.6)	62.2 (59.9,64.4)	56.4 (54.4,58.3)	50.4 (48.1,52.7)	56.3 (55.1,57.6)			
RRP6	I pre-registered my study protocols in line with open science practices	50.8 (48.5,53.1)	38.9 (36.3,41.6)	31.9 (28.4,35.5)	30.2 (24.1,37.1)	44.3 (41.4,47.3)	40.0 (37.7,42.4)	45.2 (42.5,47.9)	42.8 (41.3,44.3)			
RRP7	I managed my research data carefully by storing both the raw and processed versions for a period appropriate to my discipline and methodology used	90.9 (89.7,91.9)	88.8 (87.2,90.2)	84.5 (82.4,86.5)	82.8 (78.7,86.3)	90.8 (89.3,92)	87.9 (86.5,89.1)	86.7 (85.1,88.3)	88.4 (87.6,89.2)			
RRP8	My research was published under open access conditions	75.1 (73.3,76.8)	72.7 (70.6,74.8)	73.7 (71.2,76.0)	59.1 (54.9,63.2)	73.8 (71.4,76.1)	72.0 (70.3,73.7)	72.6 (70.6,74.5)	72.6 (71.5,73.7)			
RRP9	When making use of other people's ideas, procedures, results and text in my publications, I cited the source accurately in accordance with the standards of my discipline	98.8 (98.3,99.2)	99.3 (98.8,99.6)	98.9 (98.1,99.3)	99.4 (98.2,99.8)	98.8 (98.2,99.2)	98.8 (98.3,99.1)	99.5 (99.1,99.8)	99.0 (98.7,99.2)			
RRP10	I fully disclosed and made accessible on open science platforms my underlying data, computer codes, or syntaxes used in my research	47.4 (45.2,49.5)	41.4 (38.8,44.1)	52.7 (49.8,55.6)	53.4 (46.3,60.3)	42.4 (39.6,45.2)	47.1 (44.9,49.2)	51.0 (48.6,53.5)	47.2 (45.8,48.6)			
RRP11	Before releasing results of my research, I meticulously checked my work to avoid errors and biases	94.3 (93.4,95.2)	94.8 (93.6,95.7)	93.6 (92.2,94.8)	94.2 (92,95.9)	94.3 (93.1,95.3)	94.4 (93.4,95.2)	94.2 (93.0,95.1)	94.3 (93.7,94.8)			

Prevalence is based on the RRP at issue having a Likert score of 5, 6 or 7 among respondents that deemed the RRP at issue applicable; All figures in this table are percentages and refer to the last 3 years.

Table 2a. Linear regression coefficients (95% confidence interval) of overall RRP mean score stratified by background characteristics.

		Overall RRP mean score
		Linear regression model Mean difference from reference category (95% CI)
Disciplinary field <i>Reference category: Life and medical sciences</i>	Social and behavioral sciences	-0.15 (-0.20, -0.10)
	Natural and engineering sciences	-0.03 (-0.09, 0.04)
	Arts and humanities	-0.51 (-0.59, -0.42)
Academic rank <i>Reference category: Postdocs and assistant professors</i>	PhD candidates and junior researchers	-0.31 (-0.37, -0.25)
	Associate and full professors	0.08 (0.03, 0.14)
Gender <i>Reference category: Male</i>	Female	-0.07 (-0.12, -0.02)
	Undisclosed	0.07 (-0.10, 0.24)
Engaged in empirical research <i>Reference category: Yes</i>	No	-0.49 (-0.57, -0.42)
Institutional Support <i>Reference category: No</i>	Yes	-0.06 (-0.1, -0.01)

Overall RRP mean score was computed as the average score on the 11 RRP's with the not applicable scores recoded to 1 (i.e. never). Model containing the five background variables and all 10 explanatory factor scales. Bold figures are statistically significant.

Table 2b. Linear regression coefficients (95% confidence intervals) of overall RRP mean score by explanatory factor scales.

	Overall RRP mean score
	Linear regression model Change in mean score per standard deviation increase (95 % CI)
Work pressure	0.03 (0.01, 0.06)
Publication pressure	-0.05 (-0.08, -0.02)
Funding pressure	0.14 (0.11, 0.17)
Mentoring (survival)	0.02 (-0.01, 0.05)
Mentoring (responsible)	0.15 (0.11, 0.18)
Competitiveness	0.02 (-0.01, 0.05)
Scientific norms	0.13 (0.10, 0.15)
Peer norms	0.00 (-0.03, 0.03)
Organizational justice*	0.03 (0.00, 0.06)
Likelihood of detection (collaborators)	0.05 (0.02, 0.08)
Likelihood of detection (reviewers)	0.00 (-0.03, 0.03)

Overall RRP mean score was computed as the average score on the 11 RRP's with the not applicable scores recoded to 1 (i.e. never). Model containing the five background variables (see Table 2a) and all 10 explanatory factor scales.*Two subscales (Distributional and Procedural Organizational Justice) were merged due to high correlation. Extended data: Table 4 shows the correlation of all the explanatory factor scales. Bold figures are statistically significant.

(-0.07; 95% CI -0.12, -0.02). Being a PhD candidate or junior researcher was associated with a significantly lower overall RRP mean (-0.31; 95% CI -0.37, -0.25).

One standard deviation increase on the publication pressure scale was associated with a significant decrease in overall RRP mean score (-0.05; 95% CI -0.08, -0.02) (Table 2b). An increase of one standard deviation in the following five explanatory factor scales was associated with higher overall RRP mean, namely: (i) responsible mentoring (0.15; 95% CI 0.11, 0.18); (ii) funding pressure (0.14; 95% CI 0.11, 0.17); (iii) scientific norms subscription (0.13; 95% CI 0.10,

0.15); (iv) likelihood of QRP detection by collaborators (0.05; 95% CI 0.02, 0.08); and (v) work pressure (0.03; 95% CI 0.01, 0.06).

Discussion

We found that overall RRP prevalence ranged from 42.8% to 99% with open science practices at the lower end (42.8% to 75.0%). The Arts and Humanities scholars had the lowest prevalence of preregistration of study protocols and open access publication. This disciplinary field also had the highest prevalence of NAs (nine out of the 11 RRP), as did the PhD candidates and junior researchers. Arts and Humanities scholars, as well as PhD candidates and junior researchers, were associated with a significantly lower overall RRP mean score, as was doing non-empirical research and being female in gender.

Publication pressure was associated with lower overall RRP mean score while responsible mentoring, funding pressure, scientific norms subscription, likelihood of QRP detection by collaborators and work pressure were associated with higher RRP mean scores.

The results of our regression analysis suggest that publication pressure might lower RRPs, although the effect was modest. This finding complements what we found for QRPs, where publication pressure was associated with a higher odds of engaging frequently in at least one QRP.¹⁹ These results suggest that lowering publication pressure may be important for fostering research integrity.

Our findings regarding scientific norms and peer norms subscription are noteworthy.^{10,12} These scales have previously been validated and used in a study among 3,600 researchers of different disciplines in the United States of America.^{12,24} In that study, respondents reported higher scientific norms subscription when asked about the norms a researcher should embrace, but they perceived the actual adherence to these norms by their peers to be lower. Our results corroborate these findings.¹²

Previous authors have made calls to institutional leaders and department heads to pay increased attention to scientific norms subscription within their research cultures.^{12,25} Our regression analysis findings reinforce these calls to revive subscription to the Mertonian scientific norms.²⁴

Mentoring was associated with a higher overall RRP mean score and was aligned with a similar study by Anderson *et al.*¹⁷ Interestingly, a lack of proper supervision and mentoring of junior co-workers was the third most prevalent QRP respondents reported in our survey.¹⁹ This finding was also reported in another recent survey among researchers in Amsterdam³⁶ which suggests that increased efforts to improve mentoring and supervision may be warranted within research institutions.

In our QRP analysis of the NSRI survey results, likelihood of detection by reviewers was significantly associated with less misconduct, suggesting that reviewers, more than collaborators, are important in QRP detection.³⁷ However, for RRPs, the reverse seems to be true: collaborators may be more important for fostering RRPs than reviewers.

To our surprise, we found that work pressure and funding pressure both had a small but significant association with higher RRP mean scores. One plausible explanation may be that adhering to RRPs requires a slower, more meticulous approach to performing research. However, given the cross sectional nature of our study, these findings do not indicate causality and must be interpreted with caution.

We found that scholars from the Arts and Humanities, as well as PhD candidates and junior researchers, reported RRPs more often as “not applicable”. We were unable to differentiate whether this is because these open science RRPs are truly not applicable or if these practices are simply not yet recognized as standard responsible practices in this discipline and rank. While it can be argued that not all open science practices, particularly those relating to the sharing of data and codes, are relevant for the non-empirical disciplines such as the Arts and Humanities,^{38,39} practices like preregistration of study protocols, publishing open access and making sources, theories and hypotheses explicit and accessible, seem relevant for most types of research, empirical or not.

Arts and Humanities scholars reported the highest work pressure and competitiveness, and the lowest organizational justice and mentoring support. While our sample size for this disciplinary field was relatively small (n = 636), the finding of lower organizational justice in this discipline is consistent with a recent study.³⁷ Our regression analysis shows that Arts and Humanities scholars had significantly lower overall RRP mean scores as well as the highest prevalence of “not applicables” for nine out of the 11 RRP. Research integrity efforts have largely focused on the biomedical, and social and

behavioural sciences.⁴⁰ However, these results point to a need to better understand responsible research practices that may be disciplinary field-specific, namely to the Arts and Humanities discipline.

We found that PhD candidates and junior researchers had the lowest prevalence across all RRP and were associated with the lowest overall RRP mean score. A recent Dutch survey of academics, as well as our own survey, point to inadequate mentoring and supervision of junior co-workers as a prevalent QRP.^{19,41} This seems to underline a clear message: adequate mentoring and supervision of PhD candidates and junior researchers appears to be consistently lacking and may be contributing to lower prevalence of RRP in this rank.

Women had a slightly lower, yet statistically significant, overall RRP mean score. While it has been previously reported that men engage in research misbehavior more than women,^{19,36,42} our finding of lower RRP engagement for women has not been reported earlier and is a finding we hope to explore in the qualitative discussions planned in the next phase of our project.

The response proportion to this survey could only be reliably calculated for the eight supporting institutions and was 21.1% (*Extended data* Figure 1a). This is within the range of other research integrity surveys.^{41,42} Since there were no reliable numbers at the national level that match our study's eligibility criteria, we were unable to assess our sample's representativeness including for the five background characteristics. Despite this limitation, we believe our results to be valid as our main findings corroborate the findings of other national and international research integrity surveys.^{13,17,37,43} Nonetheless, we wish to make clear that having solid data on the representativeness of our survey respondents in terms of our overall target population is vital. While this was unavailable at both the national level and within the supporting institutions, it is imperative that future surveys collect such data prior to the survey start.

A limitation in the analysis concerns the recoding NA answers into "never" for the multiple linear regressions. We expect our analyses reported in this manuscript to be an underestimation of the occurrence of true intentional RRP as a result of this re-coding because our recoding of NA into "never" cannot distinguish between not committing a behavior because it is truly not applicable versus intentionally refraining from doing so. Our analyses may therefore underestimate the occurrence of true, intentional RRP. We have studied other recodes of the NA answers and remain confident that our preregistered choice yields inferences that do not ignore the non-random distributions of the NA answers and do not violate theoretical and practical expectations about the relation between RRP and other studied practices.

The NSRI is one of the largest and most comprehensive research integrity surveys in academia to-date which looks at prevalence of RRP and the potential associated explanatory factors in a single study across disciplinary fields and academic ranks.

Data availability

Underlying data and extended data

Repository: National Survey on Research Integrity Open Science Framework: <https://doi.org/10.17605/OSF.IO/2K549>.²⁹

This project contains the following underlying data:

- Data file 1. (NSRI Data Analysis/ GitHub:NSRI2020/NSRI (main))
- Data file 2. (ERB Ethics Approvals)
- Data file 3. (NSRI Cognitive Interview Report)
- Data file 4. (NSRI Questionnaire, Raw Dataset & Data Analysis Plan)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Authors' contributions

Conceptualization: GtR, JMW, LMB

Methodology: GG, GtR, GV, IS, JMW, LMB

Investigation: GG, JMW, OvdA

Visualization: GG, GtR, GV, IS, JMW, LMB, OvdA

Funding acquisition: GtR, LMB

Project administration: GG, LMB

Supervision: GG, GtR, LMB

Writing – original draft: GG

All authors reviewed and edited the manuscript.

Acknowledgments

The authors wish to thank the NSRI Steering Committee members (Guy Widdershoven, Herman Paul, Joeri Tjink, Sonja Zuijgeest, Corrette Ploem) for their support. In addition, we wish to thank Sara Behrad, Frank Gerritse, Coosje Veldkamp, Brian Martinson and Melissa Anderson for their contributions.

References

- National Academy of Sciences, National Academy of Engineering, Institute of Medicine, Committee on Science and Engineering and Public Policy & Panel on Scientific Responsibility and the Conduct of Research. *Responsible science - Ensuring the integrity of the research process: volume 1*. Washington, DC: The National Academies Press; 1992.
[Reference Source](#)
- Steneck N, Mayer T, Anderson MS: *Singapore statement on research integrity*. Singapore: 2010.
[Reference Source](#)
- Shaw DM, Ten Thomas C: **simple rules for protecting research integrity**. *PLoS Comput. Biol.* 2015; **11**: e1004388.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moher D, et al.: **The Hong Kong Principles for assessing researchers: Fostering research integrity**. *PLoS Biol.* 2020; **18**: e3000737.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mejlgaard N, et al.: **Research integrity: nine ways to move from talk to walk**. *Nature.* 2020; **586**: 358–360.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hiney M: *Briefing paper on research integrity: what it means, why it is important and how we might protect it*. Science Europe; 2015.
[Reference Source](#)
- Committee on Responsible Science, Committee on Science Engineering Medicine and Public Policy, Policy and Global Affairs & National Academies of Sciences Engineering and Medicine. *Fostering integrity in research*. The National Academies Press; 2017.
- ALLEA - All European Academies: *The european code of conduct for research integrity (revised edition)*. Berlin: 2017.
- Wilkinson MD et al.: **The FAIR guiding principles for scientific data management and stewardship**. *Scientific Data.* 2016; **3**: 160018.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Merton RK: *The sociology of science: theoretical and empirical investigations*. University of Chicago Press; 1973.
- KNAW, NFU, NOW, TO2-federatie, Vereniging Hogescholen, VSNU: *Netherlands code of conduct for research integrity*. The Netherlands: 2018.
[Publisher Full Text](#)
- Anderson MS, Martinson BC, De Vries R: **Normative dissonance in science: results from a national survey of U.S. scientists**. *J. Empir. Res. Hum. Res. Ethics.* 2007; **2**: 3–14.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Xie Y, Wang K, Kong Y: **Prevalence of research misconduct and questionable research practices: a systematic review and meta-analysis**. *Sci. Eng. Ethics.* 2021; **27**: 41.
[PubMed Abstract](#) | [Publisher Full Text](#)
- De Vries R, Anderson MS, Martinson BC: **Normal misbehavior: scientists talk about the ethics of research**. *J. Empir. Res. Hum. Res. Ethics.* 2006; **1**: 43–50.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bonn NA, Pinxten W: **Advancing science or advancing careers? Researchers' opinions on success indicators**. *PLoS One.* 2021; **16**: e0243664.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Anderson MS, Ronning EA, De Vries R, et al.: **The perverse effects of competition on scientists' work and relationships**. *Sci. Eng. Ethics.* 2007; **13**: 461–637.
- Anderson MS, et al.: **What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists**. *Acad. Med.* 2007; **82**: 853–860.
[PubMed Abstract](#) | [Publisher Full Text](#)
- National Survey on Research Integrity: 2020.
[Reference Source](#)
- Gopalakrishna G, ter Riet G, Vink G, et al.: **Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands**. *PLoS One.* 2022; **17**(2): e0263023.
[PubMed Abstract](#) | [Publisher Full Text](#)
- National Survey on Research Integrity on Open Science Framework (OSF): 2020.
[Reference Source](#)
- Kantar Public: 2020.
[Reference Source](#)
- ESOMAR: **Kantar signs up to ICC/ESOMAR international code globally - inks new membership deal focused on employee development**. 2020.
[Reference Source](#)
- Bouter LM, Tjink JK, Axelsen N, et al.: **Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity**. *Res. Integr. Peer Rev.* 2016; **1**: 17.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Anderson MS, Ronning EA, De Vries R, et al.: **Extending the Mertonian norms: scientists' subscription to norms of research**. *J. High. Educ.* 2010; **81**: 366–393.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Martinson BC, Crain AL, De Vries R, et al.: **The importance of organizational justice in ensuring research integrity**. *J. Empir. Res. Hum. Res. Ethics.* 2010; **5**: 67–83.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Navarro MLA, Mas MB, Jiménez AML: **Working conditions, burnout and stress symptoms in university professors: Validating a structural model of the mediating effect of perceived personal competence**. *Span. J. Psychol.* 2010; **13**: 284–296.
[PubMed Abstract](#) | [Publisher Full Text](#)

27. Haven TL, de Goede MEE, Tijdink JK, *et al.*: **Personally perceived publication pressure: revising the Publication Pressure Questionnaire (PPQ) by using work stress models.** *Res. Integr. Peer Rev.* 2019; **4**: 7.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Karasek R, Theorell T: *Healthy Work: Stress, Productivity, and the Reconstruction of Working Life.* New York; Basic Books; 1990.
29. Gopalakrishna G, *et al.*: **NSRI Supplementary Materials.** *OSF.* 2022.
[Publisher Full Text](#)
30. Miller K, Willson S, Chepp V, *et al.*: *Cognitive interviewing methodology.* Hoboken, New Jersey: John Wiley & Sons, Inc; 2014.
31. Cruyff MJLF, Van den Hout ADL, Van der Heijden PGM: **The analysis of randomized-response sum score variables.** *J. R. Stat. Soc. Series B Stat. Methodol.* 2008; **70**: 21–30.
32. van Buuren S, Groothuis-Oudshoorn K: **mice: multivariate Imputation by chained equations in R.** *J. Stat. Softw.* 2011; **45**: 1–67.
[Publisher Full Text](#)
33. Rubin DB: **Statistical matching using file concatenation with adjusted weights and multiple imputations.** *J. Bus. Econ. Stat.* 1986; **4**: 87–94.
34. Little RJA: **Missing-data adjustments in large surveys (with discussion).** *J. Bus. Econ. Stat.* 1988; **6**: 287–296.
35. Rubin DB: *Multiple imputation for nonresponse in surveys.* John Wiley & Sons; 1987; vol. **76**.
36. Fang FC, Bennett JW, Casadevall A: **Males are overrepresented among life science researchers committing scientific misconduct.** *Am. Soc. Microbiol.* 2013; **4**: e0064012.
[Publisher Full Text](#)
37. Haven TL, Tijdink JK, Martinson BC, *et al.*: **Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam.** *PLoS One.* 2019; **14**: e0210599.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Severin A, Egger M, Eve MP, *et al.*: **Discipline-specific open access publishing practices and barriers to change: an evidence-based review.** *F1000 Res.* 2018; **7**(1925)
[Publisher Full Text](#)
39. Knöchelmann M: **Open Science in the Humanities, or: Open Humanities?** *MDPI.* 2019; **7**(65).
[Publisher Full Text](#)
40. Bonn NA, Pinxten W: **A decade of empirical research on research integrity: what have we (not) looked at?** *J. Empir. Res. Hum. Res. Ethics.* 2019; **14**: 338–352.
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Haven TL, *et al.*: **Researchers' perceptions of research misbehaviours: a mixed methods study among academic researchers in Amsterdam.** *Res. Integr. Peer Rev.* 2019; **4**: 25.
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Charness G, Gneezy U: **Strong evidence for gender differences in risk taking.** *J. Econ. Behav. Organ.* 2012; **83**: 50–58.
[Publisher Full Text](#)
43. Fanelli D: **How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data.** *PLoS One.* 2009; **4**: e5738.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 15 September 2022

<https://doi.org/10.5256/f1000research.135530.r146812>

© 2022 Bik E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Elisabeth M. Bik 

Harbers Bik LLC, CA, USA

I thank the authors for addressing all comments on Version 1, and for editing the manuscript. I have no further comments and congratulate the authors with their paper.

Competing Interests: I am an independent consultant who has received payments to investigate research integrity cases for publishers, journals, and research institutions. I also get speaker's fees for giving talks for publishers and research institutions and organizations.

Reviewer Expertise: Scientific integrity

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 22 August 2022

<https://doi.org/10.5256/f1000research.135530.r146813>

© 2022 Bishop D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Dorothy Vera Margaret Bishop 

Department of Experimental Psychology, University of Oxford, Oxford, UK

I thank the authors for their detailed response to my initial report.

While I personally would prefer some different approaches to data analysis, I am happy that the authors have justified their choices, and because the data is available, this allows other readers to explore the data as they like.

Congratulations on an important paper in this field.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Psychology, Neuropsychology, Language, Reproducibility

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 27 May 2022

<https://doi.org/10.5256/f1000research.122291.r137500>

© 2022 Bishop D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Dorothy Vera Margaret Bishop 

Department of Experimental Psychology, University of Oxford, Oxford, UK

Summary

This is the 2nd paper reporting results from the National Survey on Research Integrity in the Netherlands (NSRI), this time with a focus on frequency of, and drivers of responsible research practices (RRPs). The previous study reported on questionable research practices and a future report will consider the associations between the two. This seems a reasonable way to divide up the report of findings. The authors report that frequency of RRP varies with discipline and they conduct a pre-registered regression analysis to look at predictors of RRP. I evaluated this work without looking at the existing reviewer report, in order to give an independent opinion.

Evaluation

This is a substantial piece of work; the survey materials that the authors have developed are impressive and they home in on key constructs for tackling issues around responsible research. It was particularly impressive that the researchers achieved the support of many research institutions in the Netherlands. This is a rich dataset, which is openly available to other researchers, greatly enhancing its value.

I have many suggestions for improvement, however, as I think the value and comprehensibility of the work could be enhanced. Since the F1000 model requires reviewers to give approval in order for the work to be indexed, I will note which points I see as most important to achieve that.

1. Response rate.

A major limitation of this study is the low response rate (20%). The authors mention this in the Discussion and note that their response rate is comparable to other studies of this kind. That is

certainly true (in fact, it is a relatively high rate!), and I sympathise, having had similar experiences in studies of this kind I've been involved in. Nevertheless, it really limits the conclusions one can draw, especially if there is no information about how this self-selection bias affected who responded. The authors note that they cannot assess the sample's representativeness even for the five background characteristics, but "Nevertheless, we believe our results to be valid as our main finding align well with the findings of other national and international research integrity surveys". But those other surveys suffer from the same problem: self-selection bias. Given that one of the goals of the study is to assess the prevalence, there is serious potential for biased estimates. If we have a lot of studies all with the same bias, we are in serious danger of creating illusory validity. I have two suggestions for starting to address this:

- At least for the institutions who supported the survey, gather information on the numbers of academics at the institution who fall into each discipline, and the number who fall into each academic rank. Even if these numbers are approximate, and do not describe the specific sample targeted here, they would be helpful for giving some idea about response rates in each cell of a discipline x academic rank table. Supplementary Table 1a gives some information on those completing the survey but does not actually report discipline x academic rank, which I think is an important feature (as indicated by Table 2a). I do not see it as essential to do this, as I appreciate it may be difficult to gather this information, but it would be very useful. If it is not possible to do it, maybe flag up the importance to gather this information upfront in future surveys.
- Discuss possible strategies for achieving better response rates in future surveys. An obvious one is incentives. I could not find any information about this in the paper, but in supplementary materials, I note the survey takes 20-25 minutes to complete. A survey that does not incentivise people to respond is going to be problematic because anyone who is busy and/or regards the subject of the survey uninteresting or irrelevant won't reply, and these people may have given different responses to those who do respond. Probably the most cost-effective way of incentivising people is to offer a lottery with one or more high-stakes prizes - e.g. enter people into a prize draw with the chance to win one of five prizes of €1000. It's possible ethics committees would object, but I think the case for doing this is very strong - and it could be argued it is unethical to do a study that is likely to give biased findings. The most ethical solution would be to offer each respondent an adequately motivating reward (comparable to minimum wage rate) for the time spent completing the survey. With a potential pool of 60K respondents, this would get very expensive, but the research would be more valid with a smaller pool of representative academics, than with a large pool of unrepresentative people. I think some discussion of this issue, perhaps combined with some discussion of point (2) below, would be easy to incorporate in a revision and worth doing.

2. Arts and humanities

On the one hand, it is good to include arts and humanities. But on the other hand, they frequently responded NA, and one can see why. Around 1/3 of respondents were not doing empirical research. The wording of questions to refer to 'open science', 'scripts', and 'data' is not ideal for this field. In my experience, academics in this area can get pretty irritated and feel they are having scientific practices imposed on them. The survey also has questions on adherence to 'scientific norms' - again, that wording is really not appropriate language for people in arts and humanities. 'Scholarly norms' would be better. Open access publishing has been a thorny issue in the

humanities, especially in areas where the main output is a monograph, and there may be no funds to pay for open access. (Indeed, lack of funds for open access may be a limiting factor in other disciplines and failure to ask about that is one limitation of an otherwise very well-motivated and comprehensive survey). My inclination would be to remove the Arts and Humanities subgroup from the analysis, as they are so very different in many respects, and I suspect the survey lacks face validity for many in those disciplines. (Of course, given that the authors have provided their data and scripts, it would be straightforward for other interested scholars to do this, so I don't insist on this as a condition for giving peer reviewer approval).

3. Pre-registration

I found the pre-registration status of the paper confusing. A link, <https://osf.io/2k549>, is provided under Data Availability, but that refers to a Belgian pilot study. I think that is probably just an error, but it was extremely confusing and I wasted time wading through that material looking for details of the current questionnaire. Then under 'statistical analysis' I found 'The full statistical analysis plan, and analysis codes were preregistered on the Open Science Framework', and a link to an OSF page that contains data, materials, and analysis scripts, <https://osf.io/ehx7q/>. This material is well-organised and reasonably easy to navigate, but it does not appear to have been formally pre-registered, in the sense of having a fixed date-stamped version and I could not find a document with the data analysis plan.

I did find data-analysis.rmd, which says "This document contains the analyses as described in NSRI data analysis plan - VERSION 7 - 20120126.docx", but I could not find that document on the OSF. Apologies if I missed it: hopefully this can be made more prominent, as this is a key aspect for evaluating the analysis. This is an essential point.

4. Skewed data.

For many of the items, data are skewed - in effect, these are items which amount to asking whether the respondent approves of motherhood and apple pie - everyone strongly agrees. I noticed at least one item (I did not check all), in the opposite direction, and this was one which perhaps should have been reverse scored - item F27 - most people responded 1. Some brief discussion of how this might affect results would be warranted - e.g. how does the restriction of range on some scales affect the regression coefficients?

5. Treatment of NA responses.

We are told that for the regression analysis NA was coded as 1. The justification for this is questionable. My general sense is that it would be preferable to have a smaller sample for whom the survey items were valid (i.e. where NA was not used), rather than to shoehorn all respondents into an analysis which might give a misleading picture.

It would be reassuring to readers if the analysis could be repeated by excluding all participants who responded NA, to check how this influenced findings. I see this as essential for clarifying the results.

6. Difficulty in getting the sense of the main findings.

The underlying motivation for this work includes identification of potential explanatory factors for RRP, presumably so interventions can be designed to modify these. Yet I could get no sense of how useful various explanatory factors would be, because data are reported largely as regression coefficients and confidence intervals, with the predictors shown on a z-score scale, which I think is derived from a principal component analysis. This would make it difficult for anyone else to use the same survey and try to replicate the results in a new sample - that would be easier if an average score from sets of items were used as an independent variable.

Minimally, a measure of effect size, such as percentage of variance explained, would be useful. I had to go to Supplementary Materials to find more detail of basic results of interest, and when I did that there were some anomalies - see point j below. It is a very large and complex dataset and I appreciate that the authors did not want to overwhelm readers with information. Nevertheless, I feel they have gone too far in the direction of economical presentation so that the reader has less of an immediate sense of what the results mean. This is not helped by having Methods placed after Results (see point a below).

I was interested, for instance, in understanding more about the unexpected association between work pressure/funding pressure and RRP. I didn't really understand the authors' explanation 'adherence to RRP requires a slower, more meticulous approach' - I can see that might increase work pressure because there is more to do, but it wasn't so clear for funding pressure. Why would increased funding pressure increase RRP? Perhaps funders are these days demanding that evidence of RRP is shown in proposals? What's interesting though is that some might leap on this finding to justify putting more pressure on researchers, with some kind of 'more pain, more gain' argument. Of course, they could be right! This gets right to the heart of research culture: in the past, many disciplines had a 'survival of the fittest' approach with ECRs - there was an implicit ethos that research was tough, that putting pressure on ECRs would select for the best researchers, with less committed researchers dropping out. If the most committed are also those who adopt new, open practices, then you might get this kind of association. I'm not advocating this as an explanation, which is completely against current ideas of nurturing ECRs to get the best from them! But it is important to get a more detailed picture of what is going on here.

Accordingly, I wrote a little script to explore this finding and this suggests that it's a complex picture with the effect influenced by the combination of Field and Rank, as well as Empirical Research. It also looked as if the association was at least in part driven by the NA responders doing non-empirical research. If I am able to attach figures I'll do so, but otherwise, here is the code, where the data is just the first imputed dataset, d50[[1]].

```
myplot <- ggplot(mydat, aes(x=WorkPress, y=RRP_ave, color=Research)) +  
  geom_point(shape=1, size=.5)+  
  geom_smooth(method=lm)  
myplot<- myplot + facet_grid(Field ~ Rank)
```

The plots can be viewed on Github:

<https://github.com/oscci/miscellaneous/blob/master/FundPress.pdf>

<https://github.com/oscci/miscellaneous/blob/master/WorkPress.pdf>

Assuming my plots are accurate, I'd be very cautious about making any general claims about the

impact of either Work Pressure or Publication Pressure on the adoption of RRP.

I don't regard it as essential for the authors to add such plots, but I would like to see some discussion of the possible variation across disciplines/ranks, and the substantive importance of the effect sizes in real life.

More minor points

a. I dislike the practice of putting Methods at the end of the paper. I see it as symptomatic of a tendency that the authors want to avoid - treating methods as less important than results. I can't make sense of the results until I have seen the methods. In fact, numerous questions occurred as I read the Results, which then were answered at the end of the paper. So please put this important material in its rightful place, after the introduction.

b. The file codebook.rmd does not run. The variable v is length 143, whereas descriptions and values are length 139, so they won't combine in a data frame. Ideally, we just need a document file with this information (I tried the html file, but it did not give sensibly formatted output; a simple .csv would be best). (I'm sure this is easily fixed)

c. It would be good to have a table showing the breakdown of numbers in each cell formed by cross-classifying by discipline and academic rank. Some information is available on 2-way cross-classification in the supplementary tables, but it is not enough.

d. p 3, end of Introduction; 'associative' change to 'associated'

e. Mention how long the survey took to complete: I eventually found this in the online material, but it would be better in text.

f. p 3. It seems odd to say that there were 6813 completed surveys out of 63,778 emails and then describe a 21% response rate. I assume this reflects the fact that for nonsupporting institutions, you do not know if emails were correct? Just a word of explanation is needed here.

g. Is Extended Data the same as Supplementary Materials? I assume it is and refers to the material I found at this link: <https://osf.io/w9vhc/>.

h. Why is Supplementary Table 1b not in the main text? This seems like key information from the survey, which readers will want to have readily available without needing to open supplementary documents from a link.

i. Similarly, I felt that to get a sense of the data, I needed to see the items corresponding to the scales shown in Supplementary Table 1b. These should be included in the main text.

j. Supplementary Figure 2 would also be worth including in the main text, but it needs a key indicating what each of the RRP codes is. The order of the codes seems different from the order in which each RRP is mentioned in the Tables. This figure illustrates the skew that I mentioned that affects especially Scientific Norms (I assume RRP1?) and RRP9 and RRP11. I tried to work out which scales were RRP9 and RRP11 by looking for scales with means above 5 (since most responses in

Fig 2 are 6 and 7 for these scales), but there weren't any others than Scientific Norms, so this again is confusing and needs clarifying. I eventually worked it out by comparing the main paper Table 1 and the Supplementary material, but I am still confused as to why the mean scores for F9 and F11 are not higher in Supplementary Table 1b.

k. I recommend being more cautious in the use of causal language, e.g. talking of 'explanatory factors'. This is observational rather than experimental data, based on self-report, and it is possible that there are subject-specific factors that lead to specific kinds of responses on the 'dependent' variables and also affect reporting of 'independent' predictors. In effect, any reporting biases by participants are confounded with both independent and dependent measures. The difficulty in assigning causality is apparent in the authors' own explanation of why work pressure predicts RRP - this could actually be because adoption of RRP makes more work.

The results are still of interest but need to be reported with appropriate cautions.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Psychology, Neuropsychology, Language, Reproducibility

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Jun 2022

Gowri Gopalakrishna, Amsterdam University Medical Centers, Amsterdam, The Netherlands

We thank the reviewer for these suggestions. We have already tried to gather data on the five

background variables for the eight supporting institutions, but this proved difficult as some institutions had deleted the dataset they specially generated for our survey by the time we made this request (ie shortly after the survey collection closed in Nov 2020) or did not have this data that specifically matched the definitions we used in our survey. For example academic rank category we had a number of different ranks which were not always synonymous to how these institutions categorized their staff.

We have therefore chosen to address this concern of the reviewer by flagging the importance of gathering this type of information upfront for future surveys. Discussion section now includes the following sentences: "Nonetheless, we believe having solid data on the representativeness of our survey respondents in terms of our overall target population is vital. While this was unavailable at both the national level and within the supporting institutions, it is imperative that future surveys collect such data prior to the survey start."

Discuss possible strategies for achieving better response rates in future surveys. An obvious one is incentives. I could not find any information about this in the paper, but in supplementary materials, I note the survey takes 20-25 minutes to complete. A survey that does not incentivise people to respond is going to be problematic because anyone who is busy and/or regards the subject of the survey uninteresting or irrelevant won't reply, and these people may have given different responses to those who do respond. Probably the most cost-effective way of incentivising people is to offer a lottery with one or more high-stakes prizes - e.g. enter people into a prize draw with the chance to win one of five prizes of €1000. It's possible ethics committees would object, but I think the case for doing this is very strong - and it could be argued it is unethical to do a study that is likely to give biased findings. The most ethical solution would be to offer each respondent an adequately motivating reward (comparable to minimum wage rate) for the time spent completing the survey. With a potential pool of 60K respondents, this would get very expensive, but the research would be more valid with a smaller pool of representative academics, than with a large pool of unrepresentative people. I think some discussion of this issue, perhaps combined with some discussion of point (2) below, would be easy to incorporate in a revision and worth doing.

Author Response: Thank you for raising this relevant point. We did explore with our international advisors and steering committee of the NSRI the issue of incentives. However, we decided against this for several reasons. Firstly, the literature (a selection of which is provided below this response) and our experts we consulted were divided on the usefulness of incentives to higher the response rate. Secondly, incentivizing would likely not adequately address the issue of selection bias and may even possibly exacerbate it. There is evidence suggesting if incentives are to be used, they should be offered to all participants, unconditionally in order to be effective which in our study was not feasible due to the sample size. Finally, because of our strict privacy measures, handing out incentives to responders only was not a feasible option as we didn't know who the responders were. We, therefore, chose to focus our efforts to increase participation by 1. drafting an attractive, yet succinct personalized invitation letter to each of the 63,778 participants in our target group 2. clearly outlining the importance of this survey, and how its outcome would help shape the research integrity climate for Netherlands and beyond, 3. ensuring strict privacy protection measures in the design of the survey (detailed in our methods section), 4. developing a

reasonably short, easy to answer questionnaire suitable for a laptop and other handheld devices, 5. testing the online survey layout and usability through cognitive testing, 6. conducting a broad media campaign consisting of advertising the survey on social media and news outlets (university newsletters, national newspapers, national and international academic news magazines, personalized email invitations on a last name basis), 7. sending 3 reminders over a seven week period.

Literature:

Görizt, Anja. (2006). *Incentives in Web Studies: Methodological Issues and a Review*. *International Journal of Internet Science*. 1;

Görizt, A. S., & Neumann, B. P. (2016). *The longitudinal effects of incentives on response quantity in online panels*. *Translational Issues in Psychological Science*, 2(2), 163–173.

<https://doi.org/10.1037/tps0000071>

Suzer-Gurtekin, Z. Tuba, Mahmoud ElKasabi, Mingnan Liu, James M. Lepkowski, Richard Curtin, and Rebecca McBee. 2016. "Effect of a Pre-Paid Incentive on Response Rates to an Address-Based Sampling (ABS) Web-Mail Survey." *Survey Practice* 9 (4). <https://doi.org/10.29115/SP-2016-0025>.

Evans, J.R. and Mathur, A. (2018), "The value of online surveys: a look back and a look ahead", *Internet Research*, Vol. 28 No. 4, pp. 854-887. <https://doi.org/10.1108/IntR-03-2018-0089>

2. Arts and humanities

On the one hand, it is good to include arts and humanities. But on the other hand, they frequently responded NA, and one can see why. Around 1/3 of respondents were not doing empirical research. The wording of questions to refer to 'open science', 'scripts', and 'data' is not ideal for this field. In my experience, academics in this area can get pretty irritated and feel they are having scientific practices imposed on them. The survey also has questions on adherence to 'scientific norms' - again, that wording is really not appropriate language for people in arts and humanities. 'Scholarly norms' would be better. Open access publishing has been a thorny issue in the humanities, especially in areas where the main output is a monograph, and there may be no funds to pay for open access. (Indeed, lack of funds for open access may be a limiting factor in other disciplines and failure to ask about that is one limitation of an otherwise very well-motivated and comprehensive survey). My inclination would be to remove the Arts and Humanities subgroup from the analysis, as they are so very different in many respects, and I suspect the survey lacks face validity for many in those disciplines. (Of course, given that the authors have provided their data and scripts, it would be straightforward for other interested scholars to do this, so I don't insist on this as a condition for giving peer reviewer approval).

Author Response: We thank the reviewer for this valuable comment. The reviewer is right in pointing out that certain terms used in the survey may be less applicable to the Arts and Humanities including certain Open Science practices such as open access for specific fields and open sharing of data, syntaxes and codes. However, we wish to point out that many of the RRP's (about 7 of the 11) are not exclusionary to the Arts and Humanities for example RRP 2 "I took steps to correct errors in my published work whenever I and/or peers provided valid reasons for such a correction", RRP 5 "I kept a comprehensive record of my research decisions throughout my studies", RRP 9 "When making use of other people's ideas, procedures, results and text in my

publications, I cited the source accurately in accordance with the standards of my discipline". We did strive to have face validity of our survey instrument by including a sample of researchers from all four disciplinary fields including the Arts and Humanities in our focus group sessions, in the cognitive testing of the survey instrument (see Methods section, subheading survey instrument) and in our Steering Committee. Whilst we agree that some of the RRP's may truly be not applicable to this discipline, this would not be the case for the majority of the RRP's studied. Because this field is especially understudied in the field of research integrity, we feel it is all the more important to not exclude this group from our results but rather prompt a debate on why this discipline may be so different and that we need greater understanding of this discipline in the research integrity, responsible research and open science debates. Excluding this group from our analysis and study will not help the dialogue on the need for better understanding of the challenges and knowledge production methods in this discipline.

3. Pre-registration

I found the pre-registration status of the paper confusing. A link, <https://osf.io/2k549>, is provided under Data Availability, but that refers to a Belgian pilot study. I think that is probably just an error, but it was extremely confusing and I wasted time wading through that material looking for details of the current questionnaire. Then under 'statistical analysis' I found 'The full statistical analysis plan, and analysis codes were preregistered on the Open Science Framework', and a link to an OSF page that contains data, materials, and analysis scripts, <https://osf.io/ehx7q/>. This material is well-organised and reasonably easy to navigate, but it does not appear to have been formally pre-registered, in the sense of having a fixed date-stamped version and I could not find a document with the data analysis plan.

I did find data-analysis.rmd, which says "This document contains the analyses as described in NSRI data analysis plan - VERSION 7 - 20120126.docx", but I could not find that document on the OSF. Apologies if I missed it: hopefully this can be made more prominent, as this is a key aspect for evaluating the analysis. This is an essential point.

Author Response: Thank you for your meticulous checking of the links provided under Data availability. We have corrected the link under Data Availability with the correct link: <https://osf.io/dp6zf/>. This should have been the correct link listed which when clicked takes you to several subfolders. The data analysis plan can be found under the subfolder: "NSRI Data Analysis > OSF Storage (United States)>NSRI Questionnaire, Raw Data & NSRI Data Analysis Plan. The NSRI Data Analysis Plan is a pdf document which is time stamped as 2021-08-03 03:11PM.

4. Skewed data.

For many of the items, data are skewed - in effect, these are items which amount to asking whether the respondent approves of motherhood and apple pie - everyone strongly agrees. I noticed at least one item (I did not check all), in the opposite direction, and this was one which perhaps should have been reverse scored - item F27 - most people responded 1. Some brief discussion of how this might affect results would be warranted - e.g. how does the restriction of range on some scales affect the regression coefficients?

Author Response: All scales have been recoded - when applicable - such that they measure in the same direction. In the case for F27 which measures the question: "Publication pressure sometimes leads me to cut corners." Here, most respondents did respond with a "1" indicating "never" reflecting that most respondents did not tend to favor cutting corners. On skewness and its probably effect on coefficients, this was carefully checked by our statisticians to ensure skewness did not affect our regression coefficients.

5. Treatment of NA responses.

We are told that for the regression analysis NA was coded as 1. The justification for this is questionable. My general sense is that it would be preferable to have a smaller sample for whom the survey items were valid (i.e. where NA was not used), rather than to shoehorn all respondents into an analysis which might give a misleading picture.

It would be reassuring to readers if the analysis could be repeated by excluding all participants who responded NA, to check how this influenced findings. I see this as essential for clarifying the results.

Author Response: Thank you for this comment. We wish to clarify that the "not applicable" values are bonafide missing values. While we understand removing them may seem semantically intuitive, there are valid statistical and procedural reasons why we chose to replace these values with the lowest observed category (1 = Never). Please allow us to explain these here: First, the recoding of "not applicable" to 1 is part of our pre-registered data analysis plan. Second, we did run extensive sensitivity analyses to study the validity of our pre-registered choice. Based on these analyses we concluded that our pre-registered choice is the most valid solution to this issue in this data set i.e. we deliberately chose recoding NAs into 1 as we know the direction of any potential bias: it underestimates the true effect, thereby limiting the statistical power of our analysis. Despite this we still found some effects. These sensitivity analyses can be found in the OSF data analysis folder > subfolder entitled "Figures and Tables > Table 3 Regressions": <https://osf.io/ehx7q/>.

Third, replacing "not applicable" with the value 0 is not a bonafide value that could have been observed. Using non-bonafide constants to fill in missing values is unreliable and statistically invalid. As such filling in zero would underestimate any parameters to a much greater extent than using a bonafide observed value would. We believe inducing such deliberate bias would therefore be undesirable. Fourth, coding "not applicables" as zero yields a positive correlation between QRP and RRP in a confirmatory factor analysis. This is counterintuitive and not in line with theoretical, nor practical expectations. Coding the "not applicables" as 1 (or any other bonafide observed value, for that matter) yields an expected negative correlation between the factors QRP and RRP.

Lastly, the validity of our pre-registered data analysis plan with respect to the "not applicable" has been confirmed by two independent replications on two different data structures reference: De Koning and Van der Sluis (2021). Modeling not applicable answers when data are incomplete. Master Thesis. Utrecht University.

To ensure this is also explained in the Limitations section of the manuscript, we have included this change in the Discussion section:

"We have studied other recodes of the NA answers and remain confident that our preregistered choice yields inferences that do not ignore the non-random distributions of the NA answers and do not violate theoretical and practical expectations about the relation between RRP and other studied practices."

6. Difficulty in getting the sense of the main findings.

The underlying motivation for this work includes identification of potential explanatory factors for RRP, presumably so interventions can be designed to modify these. Yet I could get no sense of how useful various explanatory factors would be, because data are reported largely as regression coefficients and confidence intervals, with the predictors shown on a z-score scale, which I think is derived from a principal component analysis. This would make it difficult for anyone else to use the same survey and try to replicate the results in a new sample - that would be easier if an average score from sets of items were used as an independent variable.

Minimally, a measure of effect size, such as percentage of variance explained, would be useful. I had to go to Supplementary Materials to find more detail of basic results of interest, and when I did that there were some anomalies - see point j below. It is a very large and complex dataset and I appreciate that the authors did not want to overwhelm readers with information. Nevertheless, I feel they have gone too far in the direction of economical presentation so that the reader has less of an immediate sense of what the results mean. This is not helped by having Methods placed after Results (see point a below).

I was interested, for instance, in understanding more about the unexpected association between work pressure/funding pressure and RRP. I didn't really understand the authors' explanation 'adherence to RRP requires a slower, more meticulous approach' - I can see that might increase work pressure because there is more to do, but it wasn't so clear for funding pressure. Why would increased funding pressure increase RRP? Perhaps funders are these days demanding that evidence of RRP is shown in proposals? What's interesting though is that some might leap on this finding to justify putting more pressure on researchers, with some kind of 'more pain, more gain' argument. Of course, they could be right! This gets right to the heart of research culture: in the past, many disciplines had a 'survival of the fittest' approach with ECRs - there was an implicit ethos that research was tough, that putting pressure on ECRs would select for the best researchers, with less committed researchers dropping out. If the most committed are also those who adopt new, open practices, then you might get this kind of association. I'm not advocating this as an explanation, which is completely against current ideas of nurturing ECRs to get the best from them! But it is important to get a more detailed picture of what is going on here.

Author Response: Thank you for raising this valid important concern. We agree that these findings must be interpreted with caution given the cross sectional nature of our study. We have emphasized this by including a sentence in the Discussion section on this topic which reads as

follows "However, given the cross sectional nature of our study, these findings do not indicate causality and must be interpreted with caution"

Assuming my plots are accurate, I'd be very cautious about making any general claims about the impact of either Work Pressure or Publication Pressure on the adoption of RRP.

I don't regard it as essential for the authors to add such plots, but I would like to see some discussion of the possible variation across disciplines/ranks, and the substantive importance of the effect sizes in real life.

Author Response: Thank you for this thorough analysis on this issue, However, please allow us to explain why the proposed procedure may be cause for concern. Our analysis are based on the pooled inferences obtained after multiple imputation. We use Rubin's rules (See reference 42, pp. 76) to properly combine parameter estimates and correctly calculate the corresponding variances. Looking at a single imputed data set and interpreting the effects of that set can be dangerous and may lull the analyst into a false sense of discovery . For example, re-running the code snippet on the next imputed data set yields different plots, with different trends and interactions. This is indicative of the sampling and missing data uncertainty in the analysis problem. Rubin's rules take the maximum likelihood estimate over the imputed sets and increase the variance about the estimates (i.e. the square of the standard error) accordingly. The procedure we followed ensures that the variability and likelihood of the obtained inference after pooling is properly incorporating the missing data and parameter uncertainty associated with the problem. Having said this, we do agree that making general claims that may imply causality is to be avoided. In that light we have included the following sentence at the end of our Discussion section as follows: "Even so, it is important to emphasize that this is a cross-sectional study and therefore does not imply causality."

More minor points

- **I dislike the practice of putting Methods at the end of the paper. I see it as symptomatic of a tendency that the authors want to avoid - treating methods as less important than results. I can't make sense of the results until I have seen the methods. In fact, numerous questions occurred as I read the Results, which then were answered at the end of the paper. So please put this important material in its rightful place, after the introduction.**

Author Response: We have moved Methods such that it appears after Introduction.

- **The file codebook.rmd does not run. The variable v is length 143, whereas descriptions and values are length 139, so they won't combine in a data frame. Ideally, we just need a document file with this information (I tried the html file, but it did not give sensibly formatted output; a simple .csv would be best). (I'm sure this is easily fixed)**

Author Response: This discrepancy in length is now fixed and can be viewed here: <https://osf.io/2yhpq/>

- **It would be good to have a table showing the breakdown of numbers in each cell formed by cross-classifying by discipline and academic rank. Some information is available on 2-way cross-classification in the supplementary tables, but it is not enough.**

Author Response: We have included an extra Supplementary Table 1b showing the proportion of respondents by discipline and rank in the supplementary files which can be directly viewed here <https://osf.io/wju6e/>

- **p 3, end of Introduction; 'associative' change to 'associated'**

Author Response: Thank you. We have made this change.

- **Mention how long the survey took to complete: I eventually found this in the online material, but it would be better in text.**

Author Response: We have included the following sentence in Methods: "We used "missingness by design" to minimize survey completion time resulting in a total of 20 minutes on average for completion"

- **p 3. It seems odd to say that there were 6813 completed surveys out of 63,778 emails and then describe a 21% response rate. I assume this reflects the fact that for non-supporting institutions, you do not know if emails were correct? Just a word of explanation is needed here.**

Author Response: We have included a reference to Supplementary Figure 1a which provides the reader with a full and detailed explanation on how we derived the 21% response proportion.

The changes made now read as follows: "This percentage was derived by dividing total number of eligible individuals who opened the survey invitation (4414) by the total number of individuals who were invited from the eight supporting institutions (20879). Extended Figure 1a²⁰ provides a detailed explanation of this calculation"

- **Is Extended Data the same as Supplementary Materials? I assume it is and refers to the material I found at this link: <https://osf.io/w9vhc/>.**

Author Response: Yes, this is correct. Extended Data is how F1000 refers to supplementary materials.

- **Why is Supplementary Table 1b not in the main text? This seems like key information from the survey, which readers will want to have readily available without needing to open supplementary documents from a link.**

Author Response: In our previous submission of this work to a different journal, there were restrictions on the number of tables and figures allowed in the main text as well as word

limits, hence we had to make choices on the most important Tables and Figures and results to discuss. As part of this we moved Table 1b to extended data which we prefer to leave as is given F1000 whilst not having a Figures and Table limit does have a word limit which we are already at.

- **Similarly, I felt that to get a sense of the data, I needed to see the items corresponding to the scales shown in Supplementary Table 1b. These should be included in the main text.**

Given the length of our manuscript, including Supplementary 1b which is already rather large to also now show all items per scale we felt would make the Table even larger and cumbersome. We, therefore, included it as a supplementary table. As for the items in each scale, these are presented Supplementary Table 5 to make it easier for the reader to digest all the information we have included in the entire manuscript.

- **Supplementary Figure 2 would also be worth including in the main text, but it needs a key indicating what each of the RRP codes is. The order of the codes seems different from the order in which each RRP is mentioned in the Tables. This figure illustrates the skew that I mentioned that affects especially Scientific Norms (I assume RRP1?) and RRP9 and RRP11. I tried to work out which scales were RRP9 and RRP11 by looking for scales with means above 5 (since most responses in Fig 2 are 6 and 7 for these scales), but there weren't any others than Scientific Norms, so this again is confusing and needs clarifying. I eventually worked it out by comparing the main paper Table 1 and the Supplementary material, but I am still confused as to why the mean scores for F9 and F11 are not higher in Supplementary Table 1b.**

Author Response: We wish to clarify that the eleven RRPs (dependent variables) shown in Supplementary Figure 2 are NOT the explanatory factor scales (independent variables). It appears that perhaps there was some confusion in the review with RRPs no. 1-11 (as shown in supplementary Figure 2) and with the 10 explanatory factors scales shown in supplementary Table 1b. Scientific Norm Subscription for example which the reviewer makes reference to is not an RRP but one of the 10 explanatory factor variables. Figure 2 show the distribution of respondent answers on the Likert answer scale and not of the explanatory factor scales.

- I recommend being more cautious in the use of causal language, e.g. talking of 'explanatory factors'. This is observational rather than experimental data, based on self-report, and it is possible that there are subject-specific factors that lead to specific kinds of responses on the 'dependent' variables and also affect reporting of 'independent' predictors. In effect, any reporting biases by participants are confounded with both independent and dependent measures. The difficulty in assigning causality is apparent in the authors' own explanation of why work pressure predicts RRPs - this could actually be because adoption of RRPs makes more work.

Author Response: Thank you for this comment. We have included a sentence at the end of the Discussion to this effect which reads as follows: "Even so, it is important to emphasize

that this is a cross-sectional study and therefore does not imply causality."

Competing Interests: None to declare.

Reviewer Report 16 May 2022

<https://doi.org/10.5256/f1000research.122291.r136272>

© 2022 Bik E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Elisabeth M. Bik 

Harbers Bik LLC, CA, USA

This manuscript is the second publication derived from a large questionnaire sent out to universities and academic medical centers in the Netherlands. While the first paper (PLOS ONE 2022, DOI: 10.1371/journal.pone.0263023; henceforth: "PONE22"¹) focused on questions and answers related to questionable research practices (QRP) and research misconduct, this manuscript reports on the survey part related to responsible research practices (RRP).

This study found a negative correlation between pressure to publish and RRP, whereas its counterpart PONE22 found a positive correlation between publication pressure and QRP. Both findings are important because they strongly suggest that the increasing pressure put on scholars to publish might lead to less RRP and more QRP and science misconduct. In addition, the paper finds a correlation between the lack of adequate mentoring of junior researchers and a lower prevalence of RRP.

The two papers nicely complement each other, but since they both refer to the same questionnaire, there is a lot of textual and data overlap. One might argue that keeping both parts of the study together in one publication would have been a better strategy because by splitting up these papers, parts of this manuscript read as redundant and not novel. While in some cases - such as in the methods - this might be acceptable, some other parts - such as the introduction and results - should be checked for identical or very similar phrases. Some specific examples of such similarities are noted below.

Strengths of this paper are the scale and anonymity of the survey, the focus on both good as well as bad practices (in combination with the PONE22 paper), and the analysis of multiple science disciplines and influencing factors. This paper nicely builds upon and extends on previous surveys on research integrity and misconduct, and offers a view of the factors that drive researchers to choose between good or bad practices. This study and its PONE22 companion could and should be used for future roadmaps on how to best structure academic research.

Specific comments

1. Introduction: the first two sentences are (nearly) identical to those of PONE22. Could they be worded a bit differently?
2. Results - "A total of 63,778 emails were sent out (Figure 1) and 9,529 eligible respondents started the survey." Figure 1 is the same as in PONE22 - the authors could just leave this sentence and add reference 19 for more details.
3. Results. "The response could only be reliably calculated for the eight supporting institutions (Figure 1a, Extended data20) and was 21.1%." - This sentence was not clear. Does this refer to the institutions (8 out of 22 invited institutions) that provided the email addresses, as mentioned in PONE22? Does the 21.1% refer to the number of invited vs completed surveys or institutions? A similar statement was found in PONE22, but there it was 21.2%.
4. The whole first part of the results "Descriptive analyses" seems nearly identical to that in PONE22. To prevent a partial duplication of already published results, it might be better to just give a short summary here, and refer to reference 19.
5. Results: Table 1: In the description of RRP1, check "finan1bcial"
6. Results. In the paragraph starting with "The four open science practices" perhaps change 75% to 75.0% to match the number of digits to the other reported percentages.
7. Discussion. The sentence starting with "In our QRP analysis of the NSRI survey results.." refers to an older survey, reference 24. Should this not be reference 19, PONE22? Perhaps the authors could check.
8. Discussion. Could the authors perhaps define the difference between publication pressure and work pressure? I was not sure what was meant by these or how they were related to questions in the survey. This is particularly relevant since these two factors had opposite effects on RRP. Perhaps this should be better defined in the results or here in the discussion?
9. Discussion: The last paragraphs, starting with "The email addresses of researchers..." are nearly identical to those in PONE22. Could the authors please rewrite these?
10. Methods: Again, large chunks of the text here are similar to those in PONE22. I am not sure how much textual similarity in the Methods would be acceptable for F1000, but it might still be needed to partially rewrite these and refer to reference 19.
11. Methods: Could the authors check the part starting with " The explanatory factors scales were based..."? The second sentence seems to talk specifically about QRPs, and might not be applicable for this paper.
12. This is very minor, but the text switched between "behavior" and "behaviour" and its derivatives.

References

1. Gopalakrishna G, Ter Riet G, Vink G, Stoop I, et al.: Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLoS One*. 2022; **17** (2): e0263023 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: I am an independent consultant who has received payments from scientific publishers and institutions to investigate particular cases of research misconduct. I also have received payments from publishers and institutions to give talks and workshops about research integrity and misconduct. In addition, I receive donations to support my work through Patreon.com.

Reviewer Expertise: Scientific integrity

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Jun 2022

Gowri Gopalakrishna, Amsterdam University Medical Centers, Amsterdam, The Netherlands

We thank the reviewer for these comments. With respect to the similarities between both manuscripts, we have modified parts of the Introduction, Results and Discussion sections to emphasize the differences between this manuscript and PONE22 more. The specific changes are outlined in detail below against each of the reviewer comments.

As the reviewer correctly points out both manuscripts, this and PONE22, are complementary to

each other. Whilst the PONE22 article reports on the fabrication and falsification (FF) and questionable research practices (QRPs), this manuscript presents the responsible research practice questions (RRPs) from the National survey on Research Integrity (NSRI) and the association between each explanatory factor and the overall RRP mean score. Because both PONE22 and this manuscript report on data of the NSRI the Methods sections are indeed quite similar.

We did consider the option of combining both papers into one but decided against this for the following reasons:

1. Much of the research integrity literature has focused on negative research behaviours (i.e. QRPs and misconduct) and hardly on the empirical evidence around RRP and the factors that may be associated with them. Combining both papers may dilute the emphasis on the RRP which we felt warranted in line with the discussions that increasingly center around rewarding responsible research practices and open science.
2. Combining both papers into one would have complicated a focused discussion considerably given the large number of descriptive as well as regression-based findings. We also believed that two separate papers with clear cross-references would make the NSRI findings more assessable than everything piled in one lengthy piece.

For descriptive results on the explanatory factors which are similar in both papers, we have now made a reference to PONE22 in this manuscript to avoid repetition.

Specific comments

Introduction: the first two sentences are (nearly) identical to those of PONE22. Could they be worded a bit differently?

Author Response: Thank you for this comment. We have changed the introduction such that it now begins with the following paragraph

"There has been a clear rise in publications and efforts aimed at promoting research integrity in recent years, including pleas for the adoption and promotion of open science and other RRP aimed at increasing the trustworthiness of research through increased transparency. In particular, open methods (e.g. preregistration of study protocols), open codes (for data analysis), open data (following the FAIR principles) and open access (rendering publications available at no cost for users) play an important role"

Results - "A total of 63,778 emails were sent out (Figure 1) and 9,529 eligible respondents started the survey." Figure 1 is the same as in PONE22 - the authors could just leave this sentence and add reference 19 for more details.

Author Response: Thank you for this comment. While we agree this part of the results is the same as in PONE22, we feel it is important to explicitly mention these results so a reader can immediately have the response proportion of our survey without needing to look up this important information in another publication.

Results. "The response could only be reliably calculated for the eight supporting institutions (Figure 1a, Extended data20) and was 21.1%." - This sentence was not

clear. Does this refer to the institutions (8 out of 22 invited institutions) that provided the email addresses, as mentioned in PONE22? Does the 21.1% refer to the number of invited vs completed surveys or institutions? A similar statement was found in PONE22, but there it was 21.2%.

Author Response: Thank you for spotting this error in the decimal point. The figure should be 21.1% and refers to the same percentage explained in Figure 1a of the extended data reference 20. This is derived by dividing total number of eligible individuals who opened the survey invitation (4414) by the total number of individuals who were invited from the 8 supporting institutions. We have further clarified by this by changing the following sentence to read as: "The response could only be reliably calculated for the eight supporting institutions (Figure 1a, Extended data20) and was 21.1%. This percentage was derived by dividing total number of eligible individuals who opened the survey invitation (4414) by the total number of individuals who were invited from the eight supporting institutions (20,879)."

The whole first part of the results "Descriptive analyses" seems nearly identical to that in PONE22. To prevent a partial duplication of already published results, it might be better to just give a short summary here, and refer to reference 19.

Author Response: We have replaced this first paragraph as suggested with the following sentence: Extended Data Table 1a describes the distribution of survey respondents by gender, disciplinary field, engagement in empirical research and if they were from a supporting institution or not. Extended Data Table 1b describes the distribution of the explanatory factor scales by disciplinary field and academic rank. The full description of these distributions can be read at Gopalakrishna et al19 .

Results: Table 1: In the description of RRP1, check "finan1bcial"

Author Response: This should have been "financial" and is now corrected.

Results. In the paragraph starting with "The four open science practices" perhaps change 75% to 75.0% to match the number of digits to the other reported percentages.

Author Response: We have changed the percentage to read as 75.0%

Discussion. The sentence starting with "In our QRP analysis of the NSRI survey results.." refers to an older survey, reference 24. Should this not be reference 19, PONE22? Perhaps the authors could check.

Author Response: Thank you for spotting this mistake. The reference should be 19 as mentioned and is now corrected.

Discussion. Could the authors perhaps define the difference between publication pressure and work pressure? I was not sure what was meant by these or how they were related to questions in the survey. This is particularly relevant since these two factors had opposite effects on RRP. Perhaps this should be better defined in the

results or here in the discussion?

Author Response: We now make this distinction more clearly in the Methods section where we have included the following sentences:

"With respect to the explanatory factor scale on work pressure, this can be defined as "the degree to which an academic has to work fast and hard, has a great deal to do, but with too little time" while publication pressure can be defined as the degree to which an academic feels s/he has to publish in high-prestige journals in order to have a sustainable career (Karasek & Theorell, 1990)," Extended data Table 620 provides the full list of questions we included in the questionnaire."

Discussion: The last paragraphs, starting with "The email addresses of researchers..." are nearly identical to those in PONE22. Could the authors please rewrite these?

Author Response: We have modified these paragraphs as requested. They now read as follows:

"The response proportion for this survey could only reliably be calculated for the eight supporting institutions and this was 21.1%¹⁹. This is within the range of other research integrity surveys 24,30. Since there were no reliable numbers at the national level that match our study's eligibility criteria, we were unable to assess our sample's representativeness including the five background characteristics. Despite this limitation, we believe our results to be valid as our main findings corroborate the findings of other national and international research integrity surveys 12,17,22,24,31.

A limitation in our analysis concerns the recoding of the NA answers into "never" for the multiple linear regressions. We expect our analyses reported in this manuscript to be an underestimation of the occurrence of true intentional RRP as a result of this re-coding. Because our recoding of NA into "never" cannot distinguish between not committing a behavior because it is truly not applicable versus intentionally refraining from doing so, our analyses may therefore underestimate the occurrence of true, intentional RRP.

The NSRI is the one of the largest and most comprehensive research integrity surveys in academia to-date which to study prevalence of RRP and the potential explanatory factors that may be associated with these behaviours in a single study across disciplinary fields and academic ranks."

Methods: Again, large chunks of the text here are similar to those in PONE22. I am not sure how much textual similarity in the Methods would be acceptable for F1000, but it might still be needed to partially rewrite these and refer to reference 19.

Author Response: Thank you for this comment. We feel it is appropriate for there to be textual similarity in the Methods as this is the same survey and hence using the same methods as in PONE22 with the important exception that this manuscript reports on a different dataset from the survey namely that on the prevalence and associative factors relating to the responsible research practices.

Methods: Could the authors check the part starting with " The explanatory factors scales were based..."? The second sentence seems to talk specifically about QRPs, and

might not be applicable for this paper.

Author Response: This is making reference to the two explanatory factor scales namely, likelihood of QRP detection by collaborators and likelihood of QRP detection by reviewers. As such they are not making reference to the PONE22 QRP paper but to two of the twelve explanatory factor scales which are named as such.

This is very minor, but the text switched between "behavior" and "behaviour" and its derivatives.

Author Response: Thank you for pointing this out. We have checked the manuscript to ensure behavior is spelled consistently.

Competing Interests: None to report

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research