Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands

Gowri Gopalakrishna, Gerben ter Riet, Gerko Vink, Ineke Stoop, Jelte M. Wicherts, Lex M. Bouter

Published: February 16, 2022 • https://doi.org/10.1371/journal.pone.0263023

Abstract

Prevalence of research misconduct, questionable research practices (QRPs) and their associations with a range of explanatory factors has not been studied sufficiently among academic researchers. The National Survey on Research Integrity targeted all disciplinary fields and academic ranks in the Netherlands. It included questions about engagement in fabrication, falsification and 11 QRPs over the previous three years, and 12 explanatory factor scales. We ensured strict identity protection and used the randomized response method for questions on research misconduct. 6,813 respondents completed the survey. Prevalence of fabrication was 4.3% (95% CI: 2.9, 5.7) and of falsification 4.2% (95% CI: 2.8, 5.6). Prevalence of QRPs ranged from 0.6% (95% CI: 0.5, 0.9) to 17.5% (95% CI: 16.4, 18.7) with 51.3% (95% CI: 50.1, 52.5) of respondents engaging frequently in at least one QRP. Being a PhD candidate or junior researcher increased the odds of frequently engaging in at least one QRP, as did being male. Scientific norm subscription (odds ratio (OR) 0.79; 95% CI: 0.63, 1.00) and perceived likelihood of detection by reviewers (OR 0.62, 95% CI: 0.44, 0.88) were associated with engaging in less research misconduct. Publication pressure was associated with more often engaging in one or more QRPs frequently (OR 1.22, 95% CI: 1.14, 1.30). We found higher prevalence of misconduct than earlier surveys. Our results suggest that greater emphasis on scientific norm subscription, strengthening reviewers in their role as gatekeepers of research quality and curbing the “publish or perish” incentive system promotes research integrity.


Editor: Sergi Fàbregues, Universitat Oberta de Catalunya, SPAIN

Received: September 3, 2021; Accepted: January 10, 2022; Published: February 16, 2022

Copyright: © 2022 Gopalakrishna 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 author and source are credited.

Data Availability: All data files are available from the Open Science Platform (https://osf.io/dp6zf). DOI: 10.17605/OSF.IO/DP6ZF.

Funding: - Awarded to Lex M. Bouter Grant No.: 20-22600-98-401 Netherlands Organisation for Health Research and Development (ZonMw) https://www.zonmw.nl/en/news-and-funding/news/detail/item/largest-study-ever-on-research-integrity-launches-aimed-at-all-researchers-in-the-netherlands/ The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. - Awarded to Jelte M. Wicherts Grant No.: Consolidator Grant 726361 (IMPROVE) European Research Council (ERC) https://erc.europa.eu he funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The basis of sound public policy relies on trustworthy and high quality research [1]. This trust is earned by being transparent and by performing research that is relevant, replicable, ethically sound and of rigorous methodological quality. Yet trust in research and replicability of previous findings [2] are compromised by researchers engaging in research misconduct, such as fabrication and falsification (FF) and subtle trespasses of ethical and methodological principles [3]. Continued efforts to promote responsible research practices (RRPs) which include open science practices like open data sharing, pre-registration of study protocols, open access publication over questionable research practices (QRPs) are therefore needed. In order to support the need for such continued efforts, solid evidence on the prevalence of research misconduct and QRPs as well as the factors promoting or curtailling such behaviours are needed.

QRPs include subtle trespasses such as not submitting valid negative results for publication, not reporting flaws in study design or execution, selective citation to enhance one’s own findings and so forth. The global discussion of the ‘replication crisis’ [2] has highlighted common worries about these QRPs becoming alarmingly prevalent and suggests underlying systematic factors, such as increased publication and funding pressures and lowered behavioural norms. After several major cases of misconduct [4], the global research community is converging to a common view on ways to foster research integrity [5].
While many integrity promoting initiatives exist [3, 6–8], strong evidence on which factors prevent these trespasses is lacking. The studies addressing this [9–13] are discipline-specific and focus on few factors to explain the occurrence of QRPFs and FF. A broad range of explanatory factors such as scientific norm subscription, organizational justice in terms of distribution of resources and promotions, competition, work, publication and funding pressures, and mentoring need to be considered in order to comprehensively understand the occurrence of QRP incidence [14–17]. The National Survey on Research Integrity (NSRI) [18] targets the prevalence of QRPFs, FF and RRPFs as well as their postulated explanatory factors. It targets all academic researchers in The Netherlands across all disciplinary fields and uses a randomized response (RR) technique to assess engagement in FF as it is a well-validated method known to elicit more honest answers on highly sensitive topics [19].

NSRI’s objectives are to estimate:

1. disciplinary field-specific prevalence of QRPFs, FF and RRPFs;
2. associations between explanatory factors and QRPFs, FF and RRPFs

In this paper, we focus on the NSRI results on QRPFs, FF and postulated explanatory factors. Elsewhere [20], we report on our findings on RRPFs and their postulated explanatory factors.

Materials and methods

Ethics approval

The Ethics Review Board of the School of Social and Behavioral Sciences of Tilburg University approved the NSRI (Approval Number: RP274). The Dutch Medical Research Involving Human Subjects Act was deemed not applicable by the Institutional Review Board of the Amsterdam University Medical Centers (Reference Number: 2020.286). The full NSRI questionnaire, its raw anonymized dataset, the complete data analysis plan, its source codes and version controls of the analysis (displayed in Github) can be found on the Open Science Framework [21].

Study design

The NSRI is 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 7 University Medical Centers in The Netherlands were invited by email to participate. To be eligible, researchers had, on average, to do at least 8 hours of research-related activities weekly and belong to life and medical sciences; social and behavioral sciences; natural and engineering sciences; or the arts and humanities; and be a PhD candidate or junior researcher (who is defined in The Netherlands as an individual with a Masters or PhD degree doing a minimum of 8 hours per week of research related tasks under close supervision) postdoctoral researcher or assistant professor; or associate or full professor.

The survey was conducted by a trusted third party, Kantar Public [22] which is an international market research company that adheres to the ICC/ESOMAR International Code of Standards [23]. Kantar Public’s sole responsibility was to send the survey invitations and reminders by email to our target group and, at the end of the data collection period, send the research team the anonymized dataset.

Universities and University Medical Centers 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. Consenting invitees could immediately participate. 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 Framework [21], Kantar Public sent us the anonymized dataset containing individual responses.

Survey Instrument

NSRI comprises of four components: 11 QRPFs, 11 RRPFs, two FFs and 12 explanatory factor scales (75 questions, detailed in S6 Table). 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 if one was doing empirical research or not [21].

All respondents obtained the same set of questions on QRPFs, RRPFs and FF, referring to one’s behavior in the previous three years. A three year timeframe was chosen to limit recall bias and is also a timeframe used in other similar studies [9, 10]. The 11 QRPFs were adapted from a recent study where 60% of the surveyed participants came from the biomedical disciplinary field [24]. As the NSRI targeted disciplinary fields including those outside of the biomedical field, we conducted a series disciplinary field specific focus groups to ensure the 11 QRPFs from Bouter et al. were applicable to our multidisciplinary target group. All QRPFs had 7-point Likert scales ranging from 1 to 7 where 1 = never and 7 = always (no intermediate linguistic labels were used) plus a “not applicable” (NA) answer option. The two FF questions used the RR technique with only a yes or no answer option [25]. The RR technique is best known to elicit more honest answers, the more sensitive in nature the questions are [19, 25]. Additionally, because the technique takes longer to apply, the survey would end up taking too long when all questions would use the technique. Hence, we chose to limit its use to only the most sensitive questions on research misconduct.

The explanatory factors scales were based on psychometrically tested scales most commonly used 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 [16, 24, 26–30]. Some of the scales were incorporated into the NSRI questionnaire verbatim, others were adapted for our population or newly created (see S5 Table). The scales on scientific norms, peer norms, competitiveness, organizational justice (procedural and distributional), and perceived likelihood of QRP detection by collaborators and reviewers were piloted. The other explanatory factor scales were either used previously in highly similar samples (e.g., publication pressure scale) [27] or in samples in earlier studies which were sufficiently similar to our current sample [31, 32] except for the funding pressure scale which was newly created but could not be piloted due to resource constraints. However, in the NSRI, this scale performed well in terms of psychometric properties (with a Cronbach’s alpha of 0.76) and in terms of convergent validity (i.e., positive correlations with publication pressure and competitiveness (S4 Table)).

We used “missingness by design” to minimize survey completion time. Thus, each invitee received one of three random subsets of 50 explanatory factor items from the full set of 75 (see S5 Table). All explanatory factor items had 7-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 “not applicable” (NA) answer option. There was no item non-response as respondents had to either complete the survey or withdraw. We pre-tested the NSRI questionnaire’s comprehensibility in cognitive interviews [15] with 8 academics from different ranks and disciplines. In summary, the comments centered around improvement in layout such as the removal of an instruction video on the RR technique which was said to be redundant, improvement in the clarity of the instructions and to emphasize certain words in the questionnaire by use of different fonts for improved clarity. The full report of the cognitive interview can be accessed at the Open Science Framework [21].

Statistical analysis

In this paper, we focus on three outcomes: (i) overall mean QRP, (ii) prevalence of any frequent QRP and (iii) any FF. The associations of these three outcomes with the five background characteristics (S1 Table) and the explanatory factor scales (Table 1) were investigated with multiple (i) linear regression, (ii) binary logistic regression and (iii) ordinal logistic regression, respectively [17]. Mean scores of individual QRP’s only consider respondents that deemed the QRP at issue applicable meaning for each of the QRP columns, mean scores were calculated only over values 1–7 and “not applicable” answers were not part of this calculation. In the multiple linear regression analysis (Tables 3 and 4), overall mean QRP was computed as the average score on the 11 QRP’s, after recoding not applicable scores to 1 (i.e. never). Prevalence was operationalized as the percentage of respondents who scored at least one QRP as 5, 6 or 7 among the respondents for that QRP. This definition allows for comparability to other studies [8, 10]. S2A–S2E Fig show the distribution of responses for the 11 QRP’s. The label ‘any FF’ was assigned if a respondent had admitted to at least one instance of falsification or fabrication.

| Table 1. Mean scores (standard deviations) and z-scores of explanatory factor scales stratified by disciplinary field and academic rank. |

https://doi.org/10.1371/journal.pone.0263023.t001

For the multivariable analyses of the explanatory factor scales we used z-scores computed as the first principal component of the corresponding items [14]. 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 R (version 4.0.3) was employed to deal with the missingness by design [33, 34]. Fifty complete data sets were generated by imputing the missing values using predictive mean matching [35, 36]. The regression models were fit to each of the 50 datasets, and the results combined into a single inference. To incorporate uncertainty due to the nonresponse, the standard errors were computed according to Rubin’s Rules [37]. All multivariable models contain the five background variables and the explanatory factor scales. The subscales distributional and procedural organizational justice were highly correlated (correlation factor of >0.8 (S4 Table)). They were thus merged to gain precision leading to the formation of one Organizational Justice scale. Results in S4 Table demonstrate that the correlations for the separate subscales were highly similar to those obtained from combining these scales. The full statistical analysis plan, and statistical analysis codes were preregistered on the Open Science Framework [21].

Identity protection

Respondents’ identity protection was ensured in accordance to the European General Data Protection Regulation (GDPR) and corresponding legislation in The Netherlands as follows: first, Kantar Public conducted the survey to ensure that the email addresses of respondents were never handled by the research team. Second, Kantar Public did not store respondents’ URLs and IP addresses. The anonymized dataset was sent to the research team upon closure of data collection and preregistration of the
statistical analysis plan. Third, we used the RR method for the two most sensitive questions [25]. RR creates a probabilistic and not a direct association between a respondent’s answer and the pertinent behaviour, adding an additional layer of confidentiality. Finally, we conducted analyses at aggregate levels only, that is across disciplinary fields, gender, academic rank, whether respondents conducted empirical research and were employed by an NSRI-supporting research institution (see S1 Table).

Results

Descriptive analyses

Of the 22 universities and University Medical Centers in the Netherlands, eight supported the NSRI. A total of 63,778 emails were sent out (Fig 1) of which 9529 eligible respondents started the survey after passing the screening questions and 6813 completed it. The percentage response could only be reliably calculated for the supporting institutions (S1A Fig). This is 21.2%. S1 Table describes these respondents, stratified by background characteristics.

![Flow chart of the survey](https://doi.org/10.1371/journal.pone.0263023.g001)

There are about equal proportions of male and female respondents. Further breakdown by disciplinary field, academic rank, research type and institutional support is detailed in S1 Table. Of respondents in the natural and engineering sciences, 24.9% are women. In the rank of associate and full professors, women make up less than 30% of respondents (S1 Table). Nearly 90% of all respondents are engaged in empirical research. Respondents from supporting and non-supporting institutions are fairly evenly distributed across disciplinary fields and academic ranks, except for the natural and engineering sciences where less than one in four (23.5%) come from supporting institutions. Postdocs and assistant professors report the highest scale scores for publication pressure (4.2), funding pressure (5.2) and competitiveness (3.7), and the lowest scale score for peer norms (4.1) and organizational justice (4.1) when compared to the other academic ranks (Table 1). Respondents from the arts and humanities have the highest scale scores for work pressure (4.8), publication pressure (4.1) and competitiveness (3.8). They also have the lowest scores for mentoring, peer norms organizational justice (3.5, 4.1 and 3.9, respectively) when compared to the other disciplinary fields (Table 1). The scientific norms scale scores, although much higher than the peer norms scale scores, show a similar trend of higher scientific norm scores and lower peer norm scores, across disciplinary fields and academic ranks.

Prevalence of QRPs and research misconduct

Table 2 shows the prevalence of the QRPs and FFs. The five most prevalent QRPs (i.e. Likert scale score 5, 6 or 7) are: (i) “Not submitting or resubmitting valid negative studies for publication” (QRP 9: 17.5%), (ii) “Insufficient inclusion of study flaws and limitations in publications” (QRP 10: 17%), (iii) “Insufficient supervision or mentoring of junior co-workers” (QRP 2: 15%), (iv) “Insufficient attention to the equipment, skills or expertise” (QRP 1: 14.7%), and (v) “Inadequate note taking of the research process” (QRP 7: 14.6%) (Table 2, Fig 2). Less than 1% of respondents said they unfairly reviewed manuscripts, grant applications or colleagues (QRP 4: 0.8%) or engaged in “improper referencing of sources” frequently (QRP 6: 0.6%) in the last three years.

![Percentage of observed answer categories of QRPs across 6813 respondents](https://doi.org/10.1371/journal.pone.0263023.g002)
Table 2. Prevalence (95% confidence intervals) of the QRPs, any frequent QRP and fabrication or falsification stratified by disciplinary field and academic rank.

https://doi.org/10.1371/journal.pone.0263023.t002

“Not (re)submitting valid negative studies for publication” (QRP 9) has the highest prevalence of “not applicable” (NA) across all disciplines with the arts and humanities on top (72.3%) (S2 Table). About one in two PhD candidates and junior researchers (48.7%) reported QRP 4 (i.e. “unfairly reviewed manuscripts, grant applications or colleagues”) as not applicable to them. Overall, the arts and humanities scholars have the highest prevalence of NAs for nine out of the 11 QRPs. PhD candidates and junior researchers have the highest NA prevalence for 10 out of 11 QRPs (S2 Table). This group also has the highest prevalence for 8 out of 11 QRPs across ranks (Table 2).

Respondents from the life and medical sciences have the highest prevalence of any frequent QRP compared to the other disciplinary fields (55.3%, Table 2). The life and medical sciences respondents also have the highest prevalence estimate for any FF (10.4%). Less than 1% of arts and humanities scholars reported fabrication. However, for falsification, these scholars have the highest prevalence estimate (6.1% 95% CI: 1.4, 10.9; Table 2).

Regression analyses

Tables 3 and 4 show the results of the regression analyses for the five background characteristics and the explanatory factor scales, respectively. All models include the five background characteristics and all explanatory factor scales.

Table 3. Regression coefficients and odds ratios (95% confidence interval) of overall QRP mean, any frequent QRP and any FF stratified by five background characteristics.

https://doi.org/10.1371/journal.pone.0263023.t003

Table 4. Regression coefficients and odds ratios (95% confidence interval) of overall QRP mean, any frequent QRP and any FF stratified by explanatory factor scales.

https://doi.org/10.1371/journal.pone.0263023.t004
Table 3 shows that being a PhD candidate or a junior researcher is associated with a statistically significantly higher odds of any frequent QRP. Being non-male (i.e., female or gender undisclosed) and doing non-empirical research is associated with a lower overall QRP mean and lower odds of any frequent QRP. The associations of the background characteristics with any FF have wide 95% confidence intervals and none are statistically significant.

Table 4 shows that a standard deviation increase on the publication pressure scale is associated with an increase of 0.10 in the overall QRP mean score. Similarly, each standard deviation increase on the scientific norms, peer norms and organizational justice scales is associated with a lower overall QRP mean scores of 0.12, 0.04, and 0.04, respectively (Table 4).

Logistic regression shows that for each standard deviation increase on the publication pressure scale, the odds of any frequent QRP increases by a factor of 1.22, while scientific norms, subscription, peer norms and organizational justice scales worked the other way around for these three explanatory factors, i.e., the odds of any frequent QRP decreases by a factor of 0.86 (scientific norms), 0.91 (peer norms) and 0.91 (organizational justice), respectively.

Ordinal regression shows that for each standard deviation increase on scientific norms, subscription or perceived likelihood of detection by reviewers scale, the odds of any FF decreases by a factor 0.79 and 0.62, respectively (Table 4).

Discussion

Summary of main findings

Our research integrity survey among academics across all disciplinary fields and ranks is one of the largest worldwide [2, 10]. Here, we share our findings on QRPs, fabrication and falsification as well as the explanatory factor scales that may be associated with the occurrence of these research misbehaviours. We find that over the last three years one in two researchers engaged frequently in at least one QRP, while one in twelve reported having falsified or fabricated their research at least once.

Postdocs and assistant professors rate publication pressure, funding pressure and competitiveness higher than other academic ranks, but peer norms and organizational justice lower. Arts and humanities scholars reported experiencing the highest work and publication pressures, the most competition and the lowest in mentoring, peer norms and organizational justice compared to other disciplinary fields. PhD candidates and junior researchers engage more often in any frequent QRP than other academic ranks as do males and those doing empirical as opposed to those doing non-empirical research.

Scientific norm subscription was the explanatory factor scale associated with the lowest prevalence of any frequent QRP and any FF. We also found that higher perceived likelihood of QRP detection by reviewers was associated with less FF. More publication pressure was associated with higher odds of any frequent QRP. Surprisingly, work pressure and competitiveness were only marginally associated with higher QRP mean while mentoring was only weakly negatively associated with overall mean QRP and not at all with the odds of any frequent QRP or any FF.

Explanatory factors that may drive or reduce research misbehaviour and misconduct

Publication pressure appears to lead to the largest increase in the odds of any frequent QRP. This finding supports recent initiatives to change the "publish or perish" reward system in academia [26, 27, 30].

Our findings on the discrepancy between subscription to scientific norms espoused by respondents and their perceived adherence to such norms by their peers corroborate earlier findings in a study among 3600 researchers in the USA [15, 16]. Previous researchers have made calls to institutional leaders and department heads to pay increased attention to these scientific norms in order to improve adherence and promote responsible conduct of research [16, 28]. Scientific norms subscription was one of two explanatory factor scales with the largest significant association in lowering any frequent QRP and FF in our regression analyses.

Perceived likelihood of detection by reviewers is significantly associated with lower odds of any FF suggesting that reviewers may have an important role in preventing research misconduct. The increased transparency offered by open science practices such as data sharing, is likely to boost chances of detection of research misconduct whether through formal journal reviewers or otherwise such as through post publication peer review or other types of scholarly reviews such as comments on preprints [31].

Lack of proper supervision and mentoring of junior co-workers was one of the three most prevalent QRPs. A recent study of 1080 researchers in Amsterdam reported similar findings [32]. Unsurprisingly, we find a moderate yet statistically significant association between survival mentoring and higher overall QRP mean suggesting that survival mentoring may be associated with higher QRPs while an association in the opposite direction, again moderate but significant, is observed for responsible mentoring and lower overall QRP mean. Both results as expected, and reported in an earlier study [13] which explored five different types of mentoring (including responsible and survival mentoring that we measured). Our study and that of Anderson et al. [13] suggests that mentors can influence behaviour in ways that both increase (in the case of survival mentoring) or decrease (in the case of responsible mentoring) the likelihood of problematic research behaviours such as QRPs.

Areas of focus within disciplines, academic ranks and gender

Lower perceived organizational justice among the arts and humanities has been previously reported [32]. This disciplinary field also has the highest proportion of NAs for nine of the 11 QRPs, suggesting that what is deemed as a QRP in the selection of 11 we have chosen for the NSRI may differ within the arts and humanities.

Among academic ranks, we find that being a PhD candidate or junior researcher is associated with the highest odds of engaging in any frequent QRP. This rank also has the highest prevalence for eight out of the 11 QRPs we measured. A recent Dutch study of academics postulated that this may be in part explained by the consistent lack of good supervision and mentoring of junior
researchers [32]. The authors suggest that it is plausible that young researchers may be more prone to unintentionally committing QRP given their lack of research experience in combination with poor supervision.

Additionally, a research environment where mistakes cannot be openly discussed may further deter newcomers from admitting errors made. A safe and supportive learning environment with adequate supervision is increasingly recognized as key in this regard [38]. The need to focus on PhD candidates or junior researchers is again emphasized as these researchers reported 10 of the 11 QRP’s as being not applicable. While some QRP’s are indeed rank-specific such as QRP’s 2 and 4 on supervision and review of grant proposals respectively, the remaining nine are not rank-specific. Our finding that identifying as male is associated with higher odds of any frequent QRP and higher overall mean QRP agrees with findings by others [39, 40].

QRP and FF prevalence

The prevalence of any frequent QRP was 51.3% which suggests that QRP may be more prevalent than previously reported. In other research integrity surveys, prevalence of self-reported QRP’s were in the range of 13–33% [9, 10]. Our finding of a high prevalence of any frequent QRP might be due to the cut-off we used in our analysis, that is at least one QRP with a score of 5, 6 or 7 (with 1 being never and 7 being always). As other studies have used different cut-offs, answer scales and different number of QRP’s and QRP definitions it render results between such surveys as not directly incomparable [9, 10]. However, a recent systematic review of surveys on research integrity showed that papers published after 2011 reported higher prevalence of misbehavior [9] which may be due to the increased awareness of research integrity in recent years although this cannot be ascertained conclusively.

When it comes to misconduct, previous surveys report the prevalence to be in the range of about 2–3% [9, 10] rising to as much as 15.5% when the questions concern misconduct observed in others [9]. In our study, the prevalence estimate of self-reported fabrication is 4.3% and self-reported falsification, 4.2%, while the prevalence estimate of any FF is 8.3%. When looking at disciplinary field-specific estimates of misconduct, life and medical sciences have the highest estimate of any FF (10.4%). These numbers are concerning and only comparable to one other smaller study (n = 140) that also used the RR technique [41]. This study found that 4.5% of their respondents admitted falsification. They did not assess fabrication [41].

The higher prevalence estimate of any FF in the life and medical sciences has been previously reported by others [10]. Unfortunately, it cannot be concluded if this is due to more misconduct actually taking place or because researchers in this particular disciplinary field are simply more aware of the issue and thus more willing to report it.

Strengths and limitations

The email addresses of researchers affiliated to non-NSRI-supporting institutions were web-scraped from open sources. Therefore, we are unable to credibly verify if the scraped email addresses matched our eligibility criteria prior to participation in the survey. Hence, we could only reliably calculate the response to the NSRI based on the eight supporting institutions. The 21.1% response is within the range of similar research integrity surveys [10, 32]. Given this response, one may wonder how representative the NSRI sample is of the target population i.e. all academic researchers in the Netherlands. Unfortunately, there are no reliable numbers at the national level that match our study’s eligibility criteria. Therefore, we cannot assess our sample’s representativeness even for the five background characteristics. Nevertheless, we believe our results to be valid as our main findings align well with the findings of other research integrity surveys [13, 16, 28, 31, 32]. Furthermore, prevalence estimates of fabrication and falsification may be more valid than those reported previously [9, 10] due to the use of the RR technique [19].

A limitation of our analysis concerns recoding NA answers into “never” for the multiple linear regressions since there is a difference between not committing a behaviour because it is truly not applicable and intentionally refraining from doing so. Our analyses may therefore underestimate the occurrence of true intentional QRP’s. 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 QRP and other studied practices. Another limitation is our definition of “any frequent QRP”, which we assigned to scores of 5, 6 or 7 on the Likert scale. Widening the definition of “frequent” would have resulted in higher prevalence estimates. Furthermore, other surveys assessed a different number of QRP’s and defined them sometimes differently, hampering direct comparisons between our survey and others.

Another potential limitation we wish to mention is misclassifications in academic rank due to promotion of individuals to a higher rank less than 3 years prior to completing our survey. Their responses are therefore likely to partly represent their behaviors whilst in a lower academic rank. However we did not collect information on years of experience of respondents in a rank due to strict privacy design of the survey. As such we are unable to comment on the impact of this misclassification on our results but we believe it to be relatively minor. Future surveys on this topic may, however, wish to take this into account in their design and analysis. The NSRI is the largest research integrity survey in academia to-date that has looked at not only prevalence of QRP’s and FF but also at the largest range of possible explanatory factors in one single study across all disciplinary fields and academic ranks using the RR technique [19].

As a follow up to the NSRI, we plan to conduct in-depth interactive workshops to further understand the major drivers or suppressors of QRP’s and FF in order to elucidate the nuances that a survey cannot capture.

Supporting information

81 Fig.

a. Flowchart of supporting institutions (n = 8). b. Flowchart of non-supporting institutions (n = 14).
https://doi.org/10.1371/journal.pone.0263023.s001

(DOCX)

82 Fig.
a. Percentage of observed answer categories of QRPCs stratified by disciplinary field. b. Percentage of observed answer categories of QRPCs stratified by academic rank. c. Percentage of observed answer categories of QRPCs stratified by gender. d. Percentage of observed answer categories of QRPCs stratified by research type (empirical Y/N). e. Percentage of observed answer categories of QRPCs stratified by institutional support (Y/N).
https://doi.org/10.1371/journal.pone.0263023.s002

**Fig.** Scatter plot of mean explanatory factor scale scores by disciplinary field and academic rank.
https://doi.org/10.1371/journal.pone.0263023.s003

**Table.** Characteristics of all respondents by disciplinary field, academic rank, gender, research type and institutional support.
https://doi.org/10.1371/journal.pone.0263023.s004

**Table.** Prevalence (%) of the “not applicable” answers stratified by disciplinary field and academic rank.
https://doi.org/10.1371/journal.pone.0263023.s005

**Table.** Mean score (95% confidence interval) of QRPCs stratified by disciplinary field and academic rank. Mean score (95% confidence interval) and prevalence (95% confidence interval) of QRPCs stratified by gender, research type and institutional support.
https://doi.org/10.1371/journal.pone.0263023.s006

**Table.** Correlation matrix of the z scores of the principal component analysis of the explanatory factor scales.
https://doi.org/10.1371/journal.pone.0263023.s007

**Table.** Full list of the explanatory factor scales and their corresponding items showing which were adapted, newly created or piloted.
https://doi.org/10.1371/journal.pone.0263023.s008

**Table.** National Survey on Research Integrity questionnaire.
https://doi.org/10.1371/journal.pone.0263023.s009

**Acknowledgments**

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

**References**


7. Hiney M. Research integrity: what it means, why it is important and how we might protect it. Science Europe; 2015.


   View Article • PubMed/NCBI • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

   View Article • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

   View Article • Google Scholar


   View Article • Google Scholar

   View Article • Google Scholar


   View Article • Google Scholar

   View Article • PubMed/NCBI • Google Scholar

   View Article • Google Scholar

   View Article • Google Scholar