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

Gowri Gopalakrishna[1]*, Gerben ter Riet[2], Gerko Vink[3], Ineke Stoop[4], Jelte M. Wicherts [5], Lex M. Bouter[1,6]

Affiliations:

[1] Department of Epidemiology and Data Science, Amsterdam University Medical Centers; Amsterdam, The Netherlands.

[2] Faculty of Health, Amsterdam University of Applied Science, Center of Expertise Urban Vitality; Amsterdam, The Netherlands.

[3] Department of Methodology & Statistics, Utrecht University; Utrecht, The Netherlands.

[4] The Netherlands Institute for Social Research; Den Haag, The Netherlands.

[5] Department of Methodology and Statistics, Tilburg University; Tilburg, The Netherlands.

[6] Department of Philosophy, Faculty of Humanities, Vrije Universiteit Amsterdam; Amsterdam, The Netherlands.

*Corresponding author. Email: g.gopalakrishna@amsterdamumc.nl

Abstract: The National Survey on Research Integrity reports on prevalence of research misconduct, questionable research practices (QRPs) and their associations with a range of explanatory factors among 6,813 academic researchers in The Netherlands. Prevalence of fabrication and falsification were 4.3% and 4.2%, respectively, and 51.3% of respondents engaged frequently in at least one QRPs. Scientific norm subscription and perceived likelihood of detection by reviewers were associated with less research misconduct. Publication pressure was positively associated with engaging more frequently in at least one QRP. 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 can promote research integrity.

One-Sentence Summary: Our survey shows that one in twelve researchers committed research misconduct and one in two engaged in sloppy science in the last three years.

Main Text: Society needs trustworthy research to meet important challenges (1). 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). To promote responsible research practices (RRP) over questionable research practices (QRP), solid evidence on the prevalence of research misconduct and QRPs as well as the factors promoting or curtailing such behaviours is 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 QRPs 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 (27-30). The National Survey on Research Integrity (NSRI) (14) targets the prevalence of QRPs, FF and responsible research practices (RRP) 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 (15).

NSRI's objectives are to estimate:

1) disciplinary field-specific prevalence of QRPs, FF and RRPs;

2) associations between explanatory factors and QRPs, FF and RRPs

In this paper, we focus on the NSRI results on QRPs, FF and postulated explanatory factors. Elsewhere (16), we report on our findings on RRPs and their postulated explanatory factors.

Results

Descriptive analyses

Of the 22 universities and UMCs in the Netherlands, eight supported the NSRI. A total of 63,778 emails were sent out (Figure 1) of which 9529 eligible respondents started the survey after passing the screening questions and 6813 completed it. The response could only be reliably calculated for the supporting institutions (fig S1a). This is 21.2%. Table S1a describes these respondents, stratified by background characteristics.

There are about equal proportions of male and female respondents. 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 (Table S1a). 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 organizational justice (4.1) (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 and organizational justice (3.5 and 3.9, respectively) (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: 14.5%) (Table 2, Figure 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.

"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%) (table S5). 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 (table S5). 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 3a and 3b 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 3a 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 3b 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 3b).

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 0.88 (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 3b).

Discussion

Summary of main findings

Our research integrity survey among academics across all disciplinary fields and ranks is one of the largest worldwide. 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 organizational justice lower. Arts and humanities scholars reported experiencing the highest work and publication pressures, the most competition and the least mentoring and organizational justice. 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 (23, 36).

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 (27-28). 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 (25, 27). 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 (37).

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 (38). Yet, surprisingly, we find neither strong nor statistically significant associations between mentoring and any frequent QRP or any FF. An earlier study (12) explored five different types of mentoring (including responsible and survival mentoring that we measured) and suggested that mentors can influence behaviour in ways that both increase and decrease the likelihood of problematic behaviours.

Areas of focus within disciplines, academic ranks and gender

Lower perceived organizational justice among the arts and humanities has been previously reported (38). This disciplinary field also has the highest proportion of NAs for nine out 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 a higher 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 (38). The authors suggest that it is plausible that young researchers may be more prone to unintentionally committing sloppy science 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 (36). The need to focus on PhD candidates or junior researchers is again emphasized as these researchers reported 10 of the 11 QRPs as being not applicable. While some QRPs are indeed rank-specific such as QRPs 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 “sloppy science” may be more prevalent than previously reported. In other research integrity surveys, prevalence of self-reported QRPs were in the range of 13-33% (9, 13). Our finding of a high prevalence of any frequent QRP might be due to the cut-off we used in our analysis i.e. 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 QRPs and QRP definitions it render results between such surveys as not directly incomparable (9, 13). However, a recent systematic review of surveys on research integrity showed that papers published after 2011 reported higher prevalence of misbehaviour (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, 13) 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. This study found that 4.5% of their respondents admitted falsification. They did not assess fabrication (15, 41).

The higher prevalence estimate of any FF in the life and medical sciences has been previously reported by others (13). 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 for NSRI participation. Hence, we calculated the response based only on the eight supporting institutions. The 21.1% response is within the range of similar research integrity surveys (13, 38). 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 (12, 25, 27, 37, 38). Furthermore, prevalence estimates of fabrication and falsification may be more valid than those reported previously (9, 13) due to the use of the RR technique, which is a well-validated method known to elicit more honest answers on sensitive topics (15).

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 QRPs. 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 QRPs and defined them sometimes differently, hampering direct comparisons between our survey and others.

The NSRI is the largest research integrity survey in academia to-date that has looked at not only prevalence of QRPs 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 (15).

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

Figure 1: Flow chart of the survey

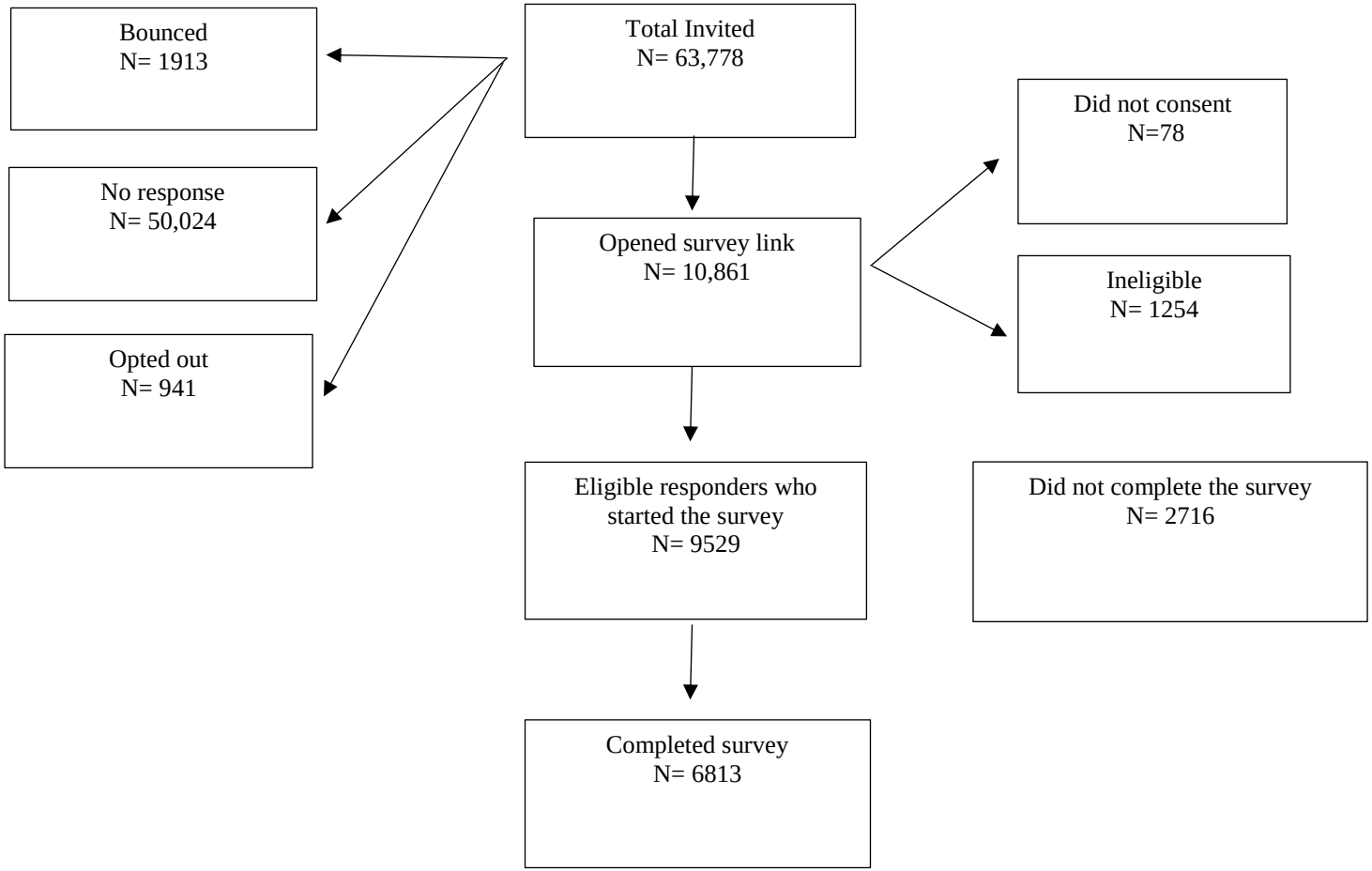


Table 1: Mean scores[#] (standard deviations) and z scores of explanatory factor scales stratified by disciplinary field and academic rank

	Disciplinary field				Academic rank			Overall
Explanatory factor scale	**Life and medical sciences**	**Social and behavioural sciences**	**Natural and engineering sciences**	**Arts and humanities**	**PhD candidates and junior researchers**	**Postdocs and assistant professors**	**Associate and full professors**	
Work pressure	4.5 (1.3)	4.5 (1.4)	4.4 (1.4)	4.8 (1.4)	3.9 (1.3)	4.7 (1.3)	4.8 (1.4)	4.5 (1.4)
Chronbach's alpha: 0.79	0	0.01	-0.10	0.20	-0.43	0.16	0.21	0
Publication pressure	3.8 (1.2)	4.0 (1.2)	3.9 (1.2)	4.1 (1.3)	3.8 (1.2)	4.2 (1.2)	3.7 (1.2)	3.9 (1.2)
Chronbach's alpha: 0.80	-0.06	0.05	0	0.13	-0.07	0.21	-0.21	0
Funding pressure	4.8 (1.4)	4.6 (1.4)	4.8 (1.4)	4.6 (1.4)	4.1 (1.5)	5.2 (1.2)	4.7 (1.3)	4.7 (1.4)
Chronbach's alpha: 0.76	0.05	-0.13	0.05	-0.10	-0.38	0.28	-0.06	-0.01
Mentoring *	4.0 (1.3)	3.8 (1.4)	3.8 (1.4)	3.5 (1.4)	4.2 (1.3)	3.9 (1.4)	3.5 (1.4)	3.9 (1.4)
Chronbach's alpha: 0.93	0.12	-0.02	-0.07	-0.28	0.26	0.01	-0.27	0
Competitiveness	3.6 (1.0)	3.5 (1.0)	3.5 (1.0)	3.8 (1.0)	3.4 (0.9)	3.7 (1.0)	3.6 (1.0)	3.6 (1.0)
Chronbach's alpha: 0.70	0.03	-0.06	-0.06	0.19	-0.20	0.10	0.06	0
Scientific norms	6.1 (0.6)	6.1 (0.6)	6.2 (0.6)	6.2 (0.6)	5.9 (0.7)	6.2 (0.6)	6.2 (0.6)	6.1 (0.6)
Chronbach's alpha: 0.71	-0.04	-0.01	0.07	0.06	-0.29	0.07	0.19	0
Peer norms	4.2 (0.9)	4.2 (0.9)	4.4 (1.0)	4.1 (1.0)	4.3 (1.0)	4.1 (0.9)	4.3 (1.0)	4.2 (1.0)
Chronbach's alpha: 0.84	-0.03	-0.06	0.17	-0.09	0.05	-0.12	0.11	0
Organizational justice **	4.4 (1.1)	4.3 (1.2)	4.5 (1.2)	3.9 (1.3)	4.6 (1.0)	4.1 (1.1)	4.5 (1.2)	4.4 (1.2)
Chronbach's alpha: 0.91	0.02	-0.01	0.14	-0.31	0.14	-0.20	0.15	0
Likelihood of detection (collaborators)	3.6 (1.0)	3.5 (1.0)	3.6 (1.0)	3.5 (1.1)	3.5 (1.1)	3.5 (1.0)	3.7 (1.0)	3.6 (1.0)
Chronbach's alpha: 0.65	0.03	-0.05	0.04	-0.06	-0.04	-0.05	0.11	0
Likelihood of detection (reviewers)	4.2 (1.2)	4.3 (1.2)	4.4 (1.2)	4.2 (1.3)	4.1 (1.3)	4.3 (1.2)	4.4 (1.1)	4.3 (1.2)
Chronbach's alpha: 0.83	-0.07	0.04	0.05	-0.06	-0.14	0	0.08	0

*[#] Scales ranging from 1 (never, totally disagree, very unlikely) to 7 (always, totally agree, very likely); *Two scales (responsible mentoring and survival mentoring) were merged due to high correlation; **Two subscales (distributional and procedural organizational justice) were merged due to high correlation; Supplementary Table 4 shows the correlation of all the explanatory factor scales.*

Table 2: Prevalence[+] (95% confidence intervals) of the QRPs, any frequent QRP[#] and fabrication or falsification stratified by disciplinary field and academic rank[^]

		Academic field				Academic rank			
QRP	**Description (In the last three years..)**	**Life and medical sciences**	**Social and behavioural sciences**	**Natural and engineering sciences**	**Arts and humanities**	**PhD candidates and junior researchers**	**Postdocs and assistant professors**	**Associate and full professors**	**Overall**
QRP1	Insufficient attention to the equipment, skills or expertise	15.2 (13.9,16.7)	14.7 (13,16.5)	13.4 (11.6,15.4)	16.2 (13,20)	15.9 (14.2,17.7)	14.6 (13.2,16.1)	13.7 (12.2,15.4)	14.7 (13.8,15.7)
QRP2	Insufficiently supervised or mentored junior co-workers	16.1 (14.7,17.6)	13.8 (12.1,15.6)	14.9 (13,17)	13.4 (10.5,16.9)	12.9 (11.1,14.9)	14.4 (13,15.8)	17 (15.4,18.7)	15 (14.1,15.9)
QRP3	Inadequate research designs or unsuitable measurement instruments	4.4 (3.7,5.3)	4.6 (3.7,5.7)	4.3 (3.3,5.6)	2.9 (1.6,5)	6 (4.9,7.2)	4 (3.3,4.9)	3.2 (2.5,4.1)	4.3 (3.9,4.9)
QRP4	Unfairly reviewed manuscripts, grant applications or colleagues	0.7 (0.4,1.2)	0.9 (0.5,1.5)	1.1 (0.6,1.9)	0.4 (0.1,1.6)	1.2 (0.6,2.1)	0.6 (0.3,1)	0.9 (0.5,1.4)	0.8 (0.6,1.1)
QRP5	Conclusions not sufficiently substantiated	3.7 (3,4.5)	4 (3.2,5.1)	4.3 (3.3,5.5)	4.9 (3.3,7.1)	6.1 (5,7.3)	3.5 (2.9,4.3)	2.8 (2.2,3.7)	4 (3.6,4.5)
QRP6	Improper referencing of source	0.6 (0.4,1)	0.4 (0.2,0.8)	0.9 (0.5,1.6)	0.8 (0.3,2)	1.1 (0.7,1.7)	0.6 (0.3,1)	0.3 (0.1,0.7)	0.6 (0.5,0.9)
QRP7	Inadequate notes of research process	13.8 (12.5,15.2)	14.4 (12.8,16.2)	16.1 (14.1,18.3)	14.6 (11.5,18.3)	15 (13.4,16.7)	15 (13.7,16.5)	13.4 (11.8,15.1)	14.5 (13.7,15.5)
QRP8	Failed to report important study details in publications	2.9 (2.3,3.7)	3 (2.3,3.9)	2.4 (1.7,3.4)	2.9 (1.7,5)	3.1 (2.3,4)	2.6 (2.1,3.4)	2.9 (2.2,3.8)	2.8 (2.4,3.3)
QRP9	Not submitting or resubmit valid negative studies for publication	14.5 (13,16.2)	17.2 (15.1,19.5)	25.3 (22.3,28.5)	19.9 (14.4,26.7)	17.1 (14.8,19.6)	19.5 (17.6,21.4)	15.5 (13.7,17.5)	17.5 (16.4,18.7)
QRP10	Insufficient inclusion of study flaws and limitations in publications	17.8 (16.4,19.4)	17.2 (15.5,19.1)	15.8 (13.9,17.9)	15.2 (12.1,19)	21.2 (19.3,23.3)	16.9 (15.5,18.4)	13.7 (12.2,15.3)	17 (16.1,18)
QRP11	Selectively cited references to enhance findings or convictions	15.8 (14.5,17.3)	11.8 (10.4,13.4)	13.8 (12.1,15.8)	13.4 (10.9,16.5)	20 (18.2,22)	13.5 (12.2,14.9)	9.5 (8.3,10.9)	14 (13.2,14.9)
Any frequent QRP	Score 5, 6 or 7 on at least 1 of the 11 QRPs	55.3 (53.4, 57.1)	50.2 (48.0, 52.5)	49.4 (46.8, 52.0)	42.1 (38.3, 46.1)	52.5 (50.3, 54.7)	52.3 (50.4, 54.2)	48.9 (46.7, 51.0)	51.3 (50.1, 52.5)
Fabrication	Making up of data or results	5.5 (3.2, 7.7)	4.8 (2.2, 7.5)	2.5 (0, 5.5)	0.7 (0, 5.1)	4.0 (1.4, 6.6)	4.9 (2.6, 7.1)	3.6 (1.1, 6.1)	4.3 (2.9, 5.7)
Falsification	Manipulating research materials, data or results	4.9 (2.7, 7.2)	2.0 (0, 4.6)	5.3 (2.2, 8.4)	6.1 (1.4, 10.9)	5.5 (2.8, 8.1)	2.6 (0.4, 4.8)	5.3 (2.7, 7.9)	4.2 (2.8, 5.6)
Any FF	Fabrication and/or Falsification	10.4 (7.1, 13.7)	5.7 (1.8, 9.5)	7.6 (3.1, 12.1)	8.4 (1.6, 15.3)	8.9 (5.0, 12.6)	7.3 (4.1, 10.6)	8.9 (5.1, 12.7)	8.3 (6.2, 10.3)

[+]Prevalence is based on the QRP at issue having a Likert score of 5, 6 or 7 among respondents that deemed the QRP at issue applicable; [#]Any frequent QRP is based on the presence of at least one of the 11 QRPs; [^]All figures in this table are percentages and refer to the last 3 years.

Figure 2: Percentage of observed answer categories of QRPs across 6813 respondents

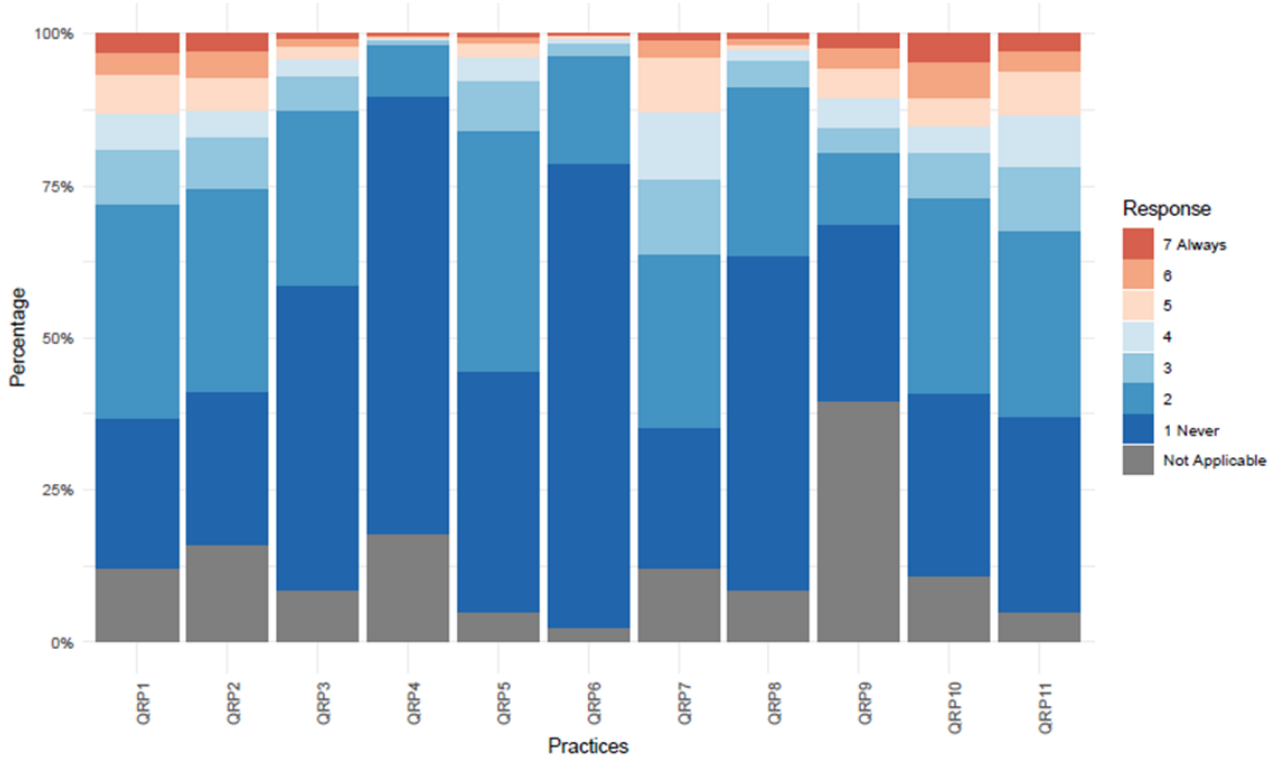


Table 3a: Regression coefficients and odds ratios (95% confidence interval) of overall QRP mean[^], any frequent QRP[¶] and any FF[#] stratified by five background characteristics

		Overall QRP Mean	Any Frequent QRP	Any FF
		Linear regression model[††] coefficient (95% CI)	**Logistic regression model[††] OR (95% CI)**	**Ordinal regression model[††] OR (95% CI)**
Disciplinary field	Social and behavourial sciences	-0.09 (-0.13, -0.05)	0.81 (0.72, 0.92)	0.81 (0.44, 1.48)
Reference category: Life and medical sciences	Natural and engineering sciences	-0.07 (-0.11, -0.03)	0.80 (0.69, 0.92)	0.92 (0.47, 1.79)
	Arts and humanities	-0.25 (-0.31, -0.19)	0.61 (0.50, 0.74)	1.16 (0.53, 2.54)
Academic rank	PhD candidates and junior researchers	*0.02 (-0.02, 0.06)*	**1.16 (1.01, 1.32)**	0.94 (0.49, 1.80)
Reference category: Postdocs and assistant professors	Associate and full professors	-0.01 (-0.05, 0.02)	0.95 (0.84, 1.08)	1.52 (0.82, 2.79)
Gender	Female	**-0.09 (-0.12, -0.06)**	**0.77 (0.69, 0.85)**	1.26 (0.73, 2.16)
Reference category: Male	Undisclosed	**-0.18 (-0.29, -0.07)**	**0.65 (0.45, 0.96)**	1.00 (0.30, 3.29)
Engaged in empirical research	No	**-0.15 (-0.20, -0.10)**	**0.76 (0.64, 0.91)**	0.63 (0.27, 1.45)
Reference category: Yes				
Institutional support	Yes	-0.03 (-0.06, 0.00)	0.93 (0.83, 1.03)	1.09 (0.64, 1.85)
Reference category: No				

[^] Overall mean QRP was computed as the average score on the 11 QRPs with the not applicable scores recoded to 1 (i.e. never); [¶]Any frequent QRP is defined as at least one of the 11 QRPs having a score of 5, 6 or 7 on the Likert scale; [#]Any FF refers to fabrication or falsification; [††]All models contain the five background characteristics (see Table 3a) and all 10 explanatory factor scales; Bold figures are statistically significant.

Table 3b: Regression coefficients and odds ratios (95% confidence interval) of overall QRP mean[^], any frequent QRP[¶] and any FF[#] stratified by explanatory factor scales

	Overall QRP mean	Any frequent QRP	Any FF
	Linear regression model[††] coefficient (95% CI)	**Logistic regression model[††] OR (95% CI)**	**Ordinal regression model[††] OR (95% CI)**
Work pressure	**0.02 (0.00, 0.04)**	1.04 (0.98, 1.11)	0.93 (0.67, 1.31)
Publication pressure	**0.10 (0.08, 0.12)**	**1.22 (1.14, 1.30)**	1.09 (0.75, 1.59)
Funding pressure	0.01 (-0.01, 0.03)	1.01(0.94, 1.08)	1.06 (0.74, 1.54)
Mentoring *	**0.02 (0.01, 0.04)**	1.01 (0.95, 1.07)	1.02 (0.73, 1.44)
Competitiveness	**0.02 (0.00, 0.04)**	1.04 (0.98, 1.12)	1.13 (0.79, 1.62)
Scientific norm	**-0.12 (-0.13, -0.10)**	**0.88 (0.83, 0.94)**	**0.79 (0.63, 1.00)**
Peer norms	**-0.04 (-0.05, -0.02)**	**0.91 (0.86, 0.97)**	1.20 (0.87, 1.65)
Organizational justice **	**-0.04 (-0.06, -0.02)**	**0.91 (0.85, 0.98)**	0.96 (0.67, 1.38)
Likelihood of detection (collaborators)	0.01 (-0.01, 0.03)	0.99 (0.93, 1.06)	0.96 (0.63, 1.47)
Likelihood of detection (reviewers)	0.00 (-0.02, 0.02)	0.99 (0.93, 1.06)	**0.62 (0.44, 0.88)**

*[^]Overall mean QRP was computed as the average score on the 11 QRPs with the not applicable scores recoded to 1 (i.e. never); [¶]Any frequent QRP is defined as at least one of the 11 QRPs having a score of 5, 6 or 7 on the Likert scale; [#]Any FF refers to fabrication or falsification; [††]All models contain the five background characteristics (see Table 3a) and all 10 explanatory factor scales; *Two scales (responsible mentoring and survival mentoring) were merged due to high correlation ; * Two subscales (distributional and procedural organizational justice) were merged due to high correlation; Bold figures are statistically significant.*

References

1. IAP - the Global Network of Science Academies, "Doing global science: a guide to responsible conduct in the global research enterprise," (Princeton University Press - Interacademy partnership, 2016).
2. National Academies of Sciences Engineering and Medicine, "Reproducibility and replicability in science," (The National Academies Press, 2019).
3. 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), vol. 1.
4. Levelt Committee, Noort Committee, Drenth Committee, "Flawed science: the fraudulent research practices of social psychologist Diederik Stapel," (Tilburg University, The Netherlands, 2012).
5. N. Mejlgaard *et al.*, Research integrity: nine ways to move from talk to walk. *Nature* **586**, 358-360 (2020).
6. M. Hiney, "Briefing paper on research integrity: what it means, why it is important and how we might protect it," (Science Europe, 2015).
7. D. M. Shaw, C. Thomas, Ten simple rules for protecting research integrity. *PLoS Computational Biology* **11**, e1004388 (2015).
8. 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).
9. Y. Xie, K. Wang, Y. Kong, Prevalence of research misconduct and questionable research practices: a systematic review and meta-analysis. *Science and Engineering Ethics* **27**, 41 (2021).
10. R. De Vries, M. S. Anderson, B. C. Martinson, Normal misbehavior: scientists talk about the ethics of research. *Journal of Empirical Research on Human Research Ethics* **1**, 43-50 (2006).
11. N. A. Bonn, W. Pinxten, Advancing science or advancing careers? Researchers' opinions on success indicators. *PLoS One* **16**, e0243664 (2021).
12. M. S. Anderson *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. *Academic Medicine* **82**, 853-860 (2007).
13. D. Fanelli, How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* **4**, e5738 (2009).
14. National Survey on Research Integrity. (Accessed Date: 12 July 2021 www.nsri2020.nl).
15. G. J. Lensvelt-Mulders, J. J. Hox, P. G. M. Van der Heijden, C. J. Maas, Meta-analysis of randomized response research: thirty five years of validation. *Sociological Methods and Research* **33**, 319-348 (2005).
16. G. Gopalakrishna *et al.*, "Prevalence of responsible research practices and their potential explanatory factors: a survey among academic researchers in The Netherlands," (Accessed Date: 12 July 2021 https://doi.org/10.31222/osf.io/xsn94, 2021).
17. National Survey on Research Integrity on Open Science Framework (OSF). (Accessed Date: 12 July 2021 https://osf.io/dp6zf/).
18. Kantar Public. (Accessed Date: 12 July 2021 http://www.nipo.nl/).
19. ESOMAR, "Kantar signs up to ICC/ESOMAR international code globally - inks new membership deal focused on employee development," (Accessed Date: 12 July 2021 https://www.esomar.org/what-we-do/news/23/news/261/Kantar-signs-up-to-ICC-ESOMAR-

International-Code-globally-inks-new-membership-deal-focused-on-employee-development 2020).
20. L. M. Bouter, J. K. Tijdink, N. Axelsen, B. C. Martinson, G. ter Riet, Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review* **1**, 17 (2016).
21. A. Y. C. KUK, Asking sensitive questions indirectly. *Biometrika* **77**, 436-438 (1990).
22. T. L. Haven, M. E. E. de Goede, J. K. Tijdink, F. J. Oort, Personally perceived publication pressure: revising the Publication Pressure Questionnaire (PPQ) by using work stress models. *Research Integrity and Peer Review* **4**, 7 (2019).
23. R. K. Merton, *The sociology of science: theoretical and empirical investigations*. (University of Chicago Press, 1973).
24. M. S. Anderson, E. A. Ronning, R. De Vries, B. C. Martinson, Extending the Mertonian norms: scientists' subscription to norms of research. *The Journal of Higher Education* **81**, 366-393 (2010).
25. M. S. Anderson, B. C. Martinson, R. De Vries, Normative dissonance in science: results from a national survey of U.S. scientists. *Journal of Empirical Research on Human Research Ethics* **2**, 3-14 (2007).
26. M. L. A. Navarro, M. B. Mas, A. M. L. Jiménez, Working conditions, burnout and stress symptoms in university professors: Validating a structural model of the mediating effect of perceived personal competence. *The Spanish Journal of Psychology* **13**, 284-296 (2010).
27. B. C. Martinson, A. L. Crain, R. De Vries, M. S. Anderson, The importance of organizational justice in ensuring research integrity. *Journal of Empirical Research on Human Research Ethics* **5**, 67-83 (2010).
28. K. Miller, S. Willson, V. Chepp, J. L. Padilla, "Cognitive interviewing methodology," (John Wiley & Sons, Inc, Hoboken, New Jersey, 2014).
29. M. J. L. F. Cruyff, A. D. L. Van den Hout, P. G. M. Van der Heijden, The analysis of randomized-response sum score variables. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **70**, 21-30 (2008).
30. W. Revelle, in *psych: procedures for psychological, psychometric, and personality research*. (Northwestern University, Evanston, Illinois, 2020).
31. S. van Buuren, K. Groothuis-Oudshoorn, mice: multivariate Imputation by chained equations in R. *Journal of Statistical Software* **45**, 1-67 (2011).
32. R Core Team, "R: A language and environment for statistical computing. R foundation for statistical computing," (Vienna, Austria, 2021).
33. R. J. A. Little, Missing-data adjustments in large surveys (with discussion). *Journal of Business Economics and Statistics* **6**, 287–296 (1988).
34. D. B. Rubin, Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics* **4**, 87-94 (1986).
35. D. B. Rubin, *Multiple imputation for nonresponse in surveys*. (John Wiley & Sons, New York, 1987).
36. D. Moher *et al.*, The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology* **18**, e3000737 (2020).
37. J. M. Wicherts, Psychology must learn a lesson from fraud case. *Nature* **480**, (2011).
38. T. L. Haven, J. K. Tijdink, B. C. Martinson, L. M. Bouter, Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. *PLoS One* **14**, e0210599 (2019).
39. G. Charness, U. Gneezy, Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization* **83**, 50-58 (2012).

40. F. C. Fang, J. W. Bennett, A. Casadevall, Males are overrepresented among life science researchers committing scientific misconduct. *American Society for Microbiology* **4**, e0064012 (2013).
41. J. A. List, C. D. Bailey, P. J. Euzent, T. L. Martin, Academic economist behaving badly? A survey on three areas of unethical behavior. *Economic Enquiry* **39**, 162-170 (2001).
42. KNAW; NFU; NWO; TO2-federatie; Vereniging Hogescholen; VSNU, "Netherlands code of conduct for research integrity," (The Netherlands, 2018).

Acknowledgments: The authors wish to thank the NSRI Steering Committee members (Guy Widdershoven, Herman Paul, Ineke Stoop, Joeri Tijdink, Sonja Zuijdgeest, 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.

Funding:

Netherlands Organisation for Health Research and Development (ZonMw) 20-22600-98-401

Consolidator Grant 726361 (IMPROVE) from the European Research Council (ERC, https://erc.europa.eu), awarded to J.M. Wicherts

Author contributions:

Conceptualization: GG, GtR, IS, JMW, LMB

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

Investigation: GG, JMW

Visualization: GG, GV, LMB

Funding acquisition: GtR, LMB

Project administration: GG, LMB

Supervision: GG, GtR, LMB

Writing – original draft: GG, GtR, LMB

Writing – review & editing: GG, GtR, GV, IS, JMW, LMB

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary Materials for

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

Gowri Gopalakrishna[1]*, Gerben ter Riet[2], Gerko Vink[3], Ineke Stoop[4], Jelte M. Wicherts [5], Lex M. Bouter[1,6]

Correspondence to: g.gopalakrishna@amsterdamumc.nl

This PDF file includes:

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 (WMO) was deemed not applicable by the Institutional Review Board of the Amsterdam University Medical Centers (UMCs) (Reference Number: 2020.286). The full NSRI study protocol, ethics approvals, complete data analysis plan and final dataset can be found on the Open Science Framework (17).

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 UMCs 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, 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 (18) which is an international market research company that adheres to the ICC/ESOMAR International Code of Standards (19). 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 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. 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 (17), Kantar Public sent us the anonymized dataset containing individual responses.

Survey Instrument

NSRI comprises of four components: 11 QRPs, 11 RRPs, two FFs 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 if one was doing empirical research or not (17).

All respondents obtained the same set of questions on QRPs, RRPs and FF, referring to one's behavior in the previous three years. The 11 QRPs were adapted from a recent study (20). All QRPs 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 a randomized response (RR) technique with only a yes or no answer option (21).

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 (20, 22-27). Some of the scales were incorporated into the NSRI questionnaire verbatim, others were adapted for our population or newly created (see table S6). The scales on scientific norms, peer norms, competitiveness, organizational justice, and perceived likelihood of QRP detection were piloted.

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 table S6). 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 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 (28) with 18 academics from different ranks and disciplines.

Statistical analysis

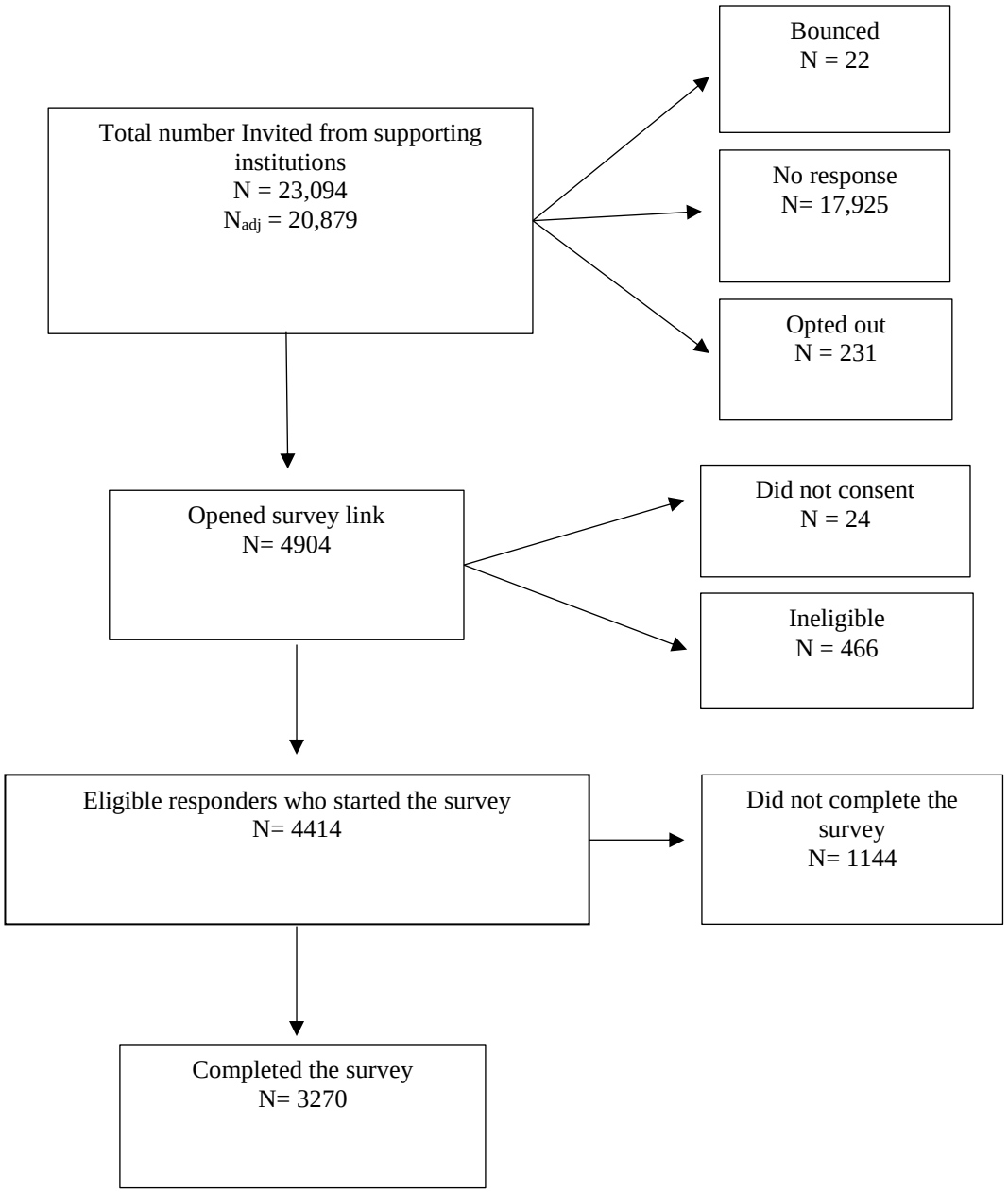
In this paper, we focus on three outcomes: (i) overall mean QRP, (ii) prevalence of any frequent QRP and (iii) any FF. Mean scores of individual QRPs only consider respondents that deemed the QRP at issue applicable. In the multiple linear regression analysis, overall mean QRP was computed as the average score on the 11 QRPs, after recoding not applicable scores to 1 (i.e. never). Prevalence was operationalized as the proportion of respondents who scored at least one QRP as 5, 6 or 7. Supplementary figures 2a to 2e show the distribution of responses for the 11 QRPs. The label 'any FF' was assigned if a respondent had admitted to at least one instance of falsification or fabrication. The associations of these three outcomes with the five background characteristics (Table S1a) and the explanatory factor scales (Table 1) were investigated with multiple (i) linear regression, (ii) binary logistic regression and (iii) ordinal logistic regression, respectively (29).

For the multivariable analyses of the explanatory factor scales we used z-scores computed as the first principal component of the corresponding items (30). 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 (31, 32). 50 complete data sets were generated by imputing the missing values using predictive mean matching (33-35). 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 (35). All multivariable models contain the five background variables and the explanatory factor scales. The full statistical analysis plan, and statistical analysis codes were preregistered on the Open Science Framework (17).

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 (21). 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, i.e. across disciplinary fields, gender, academic rank, whether respondents conducted empirical research and were employed by an NSRI-supporting research institution.

Figure S1a: **Flowchart of supporting institutions (n=8)**



N_{adj} = total number of email addresses provided to us by the supporting institutions with three corrections:

a) Inclusion of a number of eligible researchers who were not included in the original e-mail list provided to Kantar by their institutions

b) deducting the no. of bounced e-mail addresses and

c) applying a correction factor based on the number of participations who did not fit our inclusion criteria

We defined c) as the fraction of invitees who opened the survey link and consented to participate but subsequently turned out to be ineligible i.e. did not meet the inclusion criteria (Fig 1):

466 / 4904 = 0.095

This calculation is based on the assumption that eligibility does not influence the decision to open the survey and to provide informed consent. That seems to be a reasonable assumption because the exact eligibility criteria were obscure for invitees until they had opened the survey and provided informed consent. Therefore, taken together, our best estimate of the total number of eligible invitees that fit our inclusion criteria from supporting institutes can be calculated as follows:

(23094 – 22) – (0.095 * 23094) = 20879 (N_{adj})

% Response = 4414/20879 = 21.1%

Figure S1b: Flowchart of non-supporting institutions (n=14)

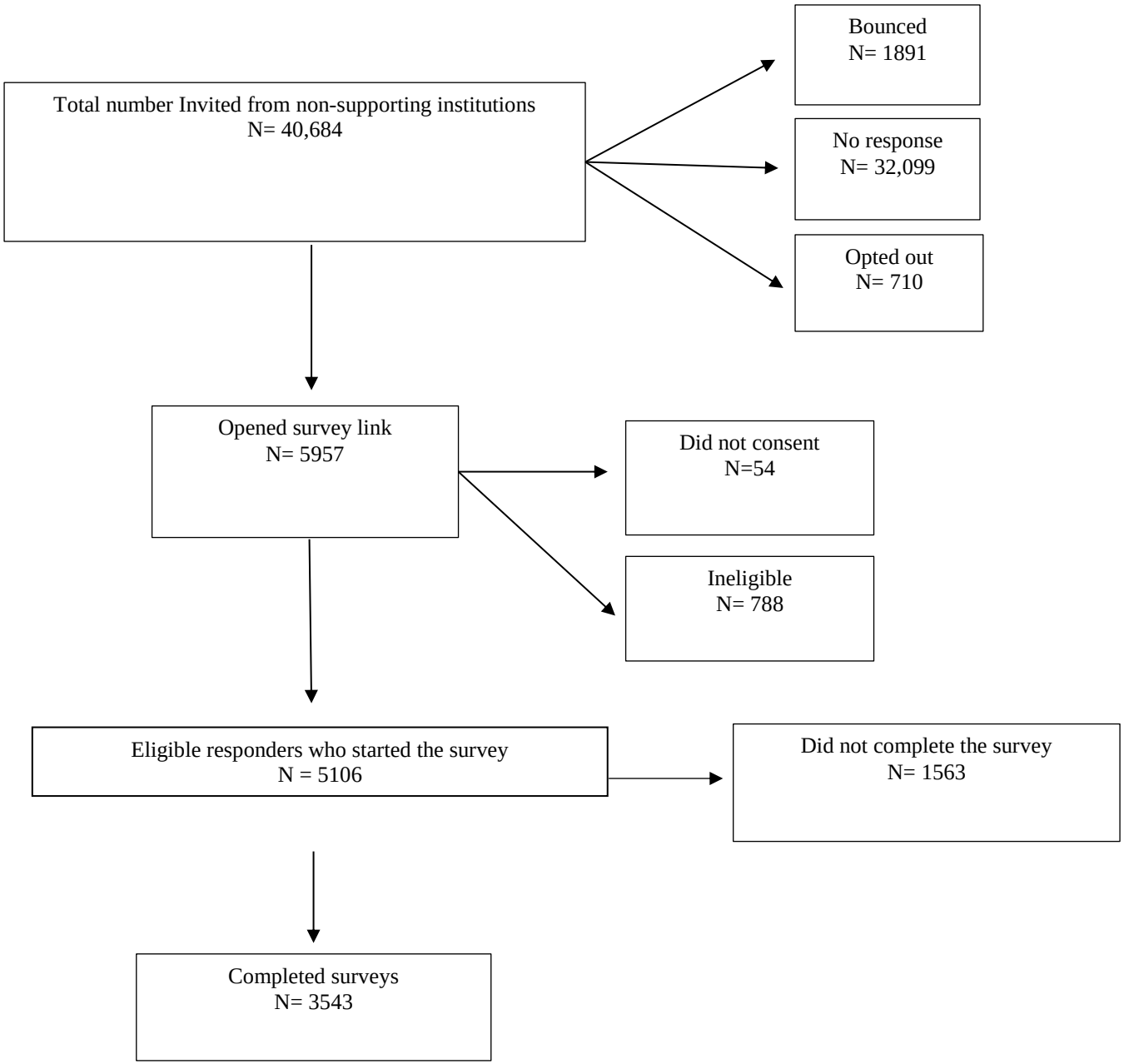


Figure S1c: Scatter plot of mean explanatory factor scale scores by disciplinary field and academic rank

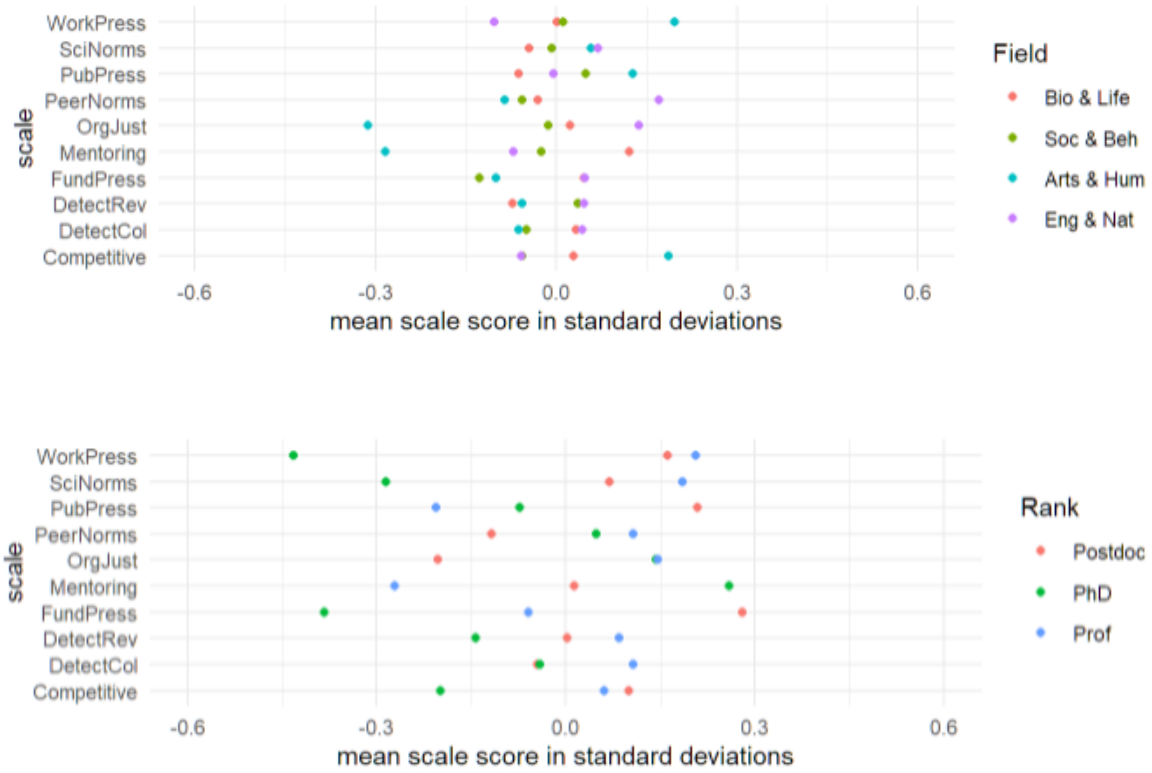


Table S1a: Characteristics of all respondents by disciplinary field, academic rank, gender, research type and institutional support

	Disciplinary field				Academic rank			Total
	Life and medical sciences (N)	**Social and behavourial sciences (N)**	**Natural and engineering sciences (N)**	**Arts and humanities (N)**	**PhD candidates and junior researchers (N)**	**Postdocs and assistant professors (N)**	**Associate and full professors (N)**	**Total sample (N)**
	2747	1965	1465	636	2013	2733	2066	6813
Female (%)	48.7	51.5	24.9	46.1	56.9	46.3	28.8	44.2
Male (%)	49.6	47.0	73.5	50.8	41.9	51.7	69.3	54.1
Undisclosed (%)	1.6	1.5	1.6	3.1	1.2	1.9	1.9	1.7
Being mainly engaged in empirical research (%)	97.6	94.0	77.7	65.1	87.9	89.6	90.2	89.3
Institutional support (%)	58.6	52.3	23.5	45.3	59.6	45.2	40.5	48.0

Table S2a: **Mean score¶¶ (95% confidence interval) of QRPs stratified by disciplinary field and academic rank**

QRP	Description (In the last three years..)	Life and medical sciences (N=2747; 40.3%)	Social and behavioural sciences (N=1965; 28.8%)	Natural and engineering sciences (N=1465; 21.5%)	Arts and humanities (N=636; 9.3%)	PhD candidates and junior researchers (N=2013; 29.5%)	Postdocs and assistant professors (N=2733; 40.1%)	Associate and full professors (N=2066; 30.3%)	Overall (N= 6813; 100%)
QRP1	Insufficient attention to the equipment, skills or expertise	2.5 (2.5,2.6)	2.5 (2.4,2.6)	2.4 (2.4,2.5)	2.6 (2.4,2.7)	2.6 (2.5,2.7)	2.5 (2.4,2.6)	2.4 (2.3,2.5)	2.5 (2.5,2.5)
QRP2	Insufficiently supervised or mentored junior co-workers	2.5 (2.5,2.6)	2.4 (2.3,2.4)	2.5 (2.4,2.6)	2.3 (2.2,2.4)	2.3 (2.3,2.4)	2.4 (2.4,2.5)	2.6 (2.5,2.7)	2.5 (2.4,2.5)
QRP3	Inadequate research designs or unsuitable measurement instruments	1.8 (1.7,1.8)	1.8 (1.7,1.8)	1.7 (1.6,1.7)	1.6 (1.5,1.6)	2 (1.9,2)	1.7 (1.7,1.7)	1.6 (1.5,1.6)	1.7 (1.7,1.8)
QRP4	Unfairly reviewed manuscripts, grant applications or colleagues	1.2 (1.1,1.2)	1.2 (1.1,1.2)	1.2 (1.1,1.2)	1.1 (1.1,1.2)	1.2 (1.2,1.2)	1.1 (1.1,1.2)	1.2 (1.2,1.2)	1.2 (1.2,1.2)
QRP5	Conclusions not sufficiently substantiated	1.9 (1.9,1.9)	1.9 (1.9,2)	1.9 (1.8,1.9)	1.9 (1.8,1.9)	2.1 (2,2.1)	1.9 (1.8,1.9)	1.7 (1.7,1.8)	1.9 (1.9,1.9)
QRP6	Improper referencing of source	1.3 (1.3,1.3)	1.3 (1.2,1.3)	1.2 (1.2,1.3)	1.3 (1.2,1.3)	1.4 (1.3,1.4)	1.3 (1.2,1.3)	1.2 (1.2,1.3)	1.3 (1.3,1.3)
QRP7	Inadequate notes of research process	2.6 (2.6,2.7)	2.6 (2.6,2.7)	2.6 (2.6,2.7)	2.5 (2.4,2.6)	2.7 (2.6,2.8)	2.6 (2.6,2.7)	2.5 (2.5,2.6)	2.6 (2.6,2.7)
QRP8	Failed to report important study details in publications	1.6 (1.6,1.6)	1.6 (1.5,1.6)	1.6 (1.5,1.6)	1.6 (1.5,1.7)	1.6 (1.6,1.7)	1.6 (1.6,1.6)	1.5 (1.5,1.6)	1.6 (1.6,1.6)
QRP9	Not submitting or resubmit valid negative studies for publication	2.3 (2.2,2.3)	2.3 (2.3,2.4)	2.8 (2.7,2.9)	2.4 (2.2,2.5)	2.3 (2.2,2.4)	2.5 (2.4,2.6)	2.3 (2.2,2.4)	2.4 (2.4,2.4)
QRP10	Insufficient inclusion of study flaws and limitations in publications	2.6 (2.5,2.7)	2.5 (2.4,2.6)	2.5 (2.4,2.6)	2.4 (2.3,2.5)	2.8 (2.7,2.9)	2.5 (2.5,2.6)	2.3 (2.2,2.4)	2.5 (2.5,2.6)
QRP11	Selectively cited references to enhance findings or convictions	2.6 (2.6,2.7)	2.4 (2.3,2.4)	2.3 (2.3,2.4)	2.3 (2.2,2.4)	2.7 (2.7,2.8)	2.5 (2.4,2.5)	2.2 (2.2,2.3)	2.5 (2.4,2.5)

¶¶*Scores range from 1=never to 7=always; Mean scores of individual QRPs only consider respondents that deemed the QRP at issue applicable.*

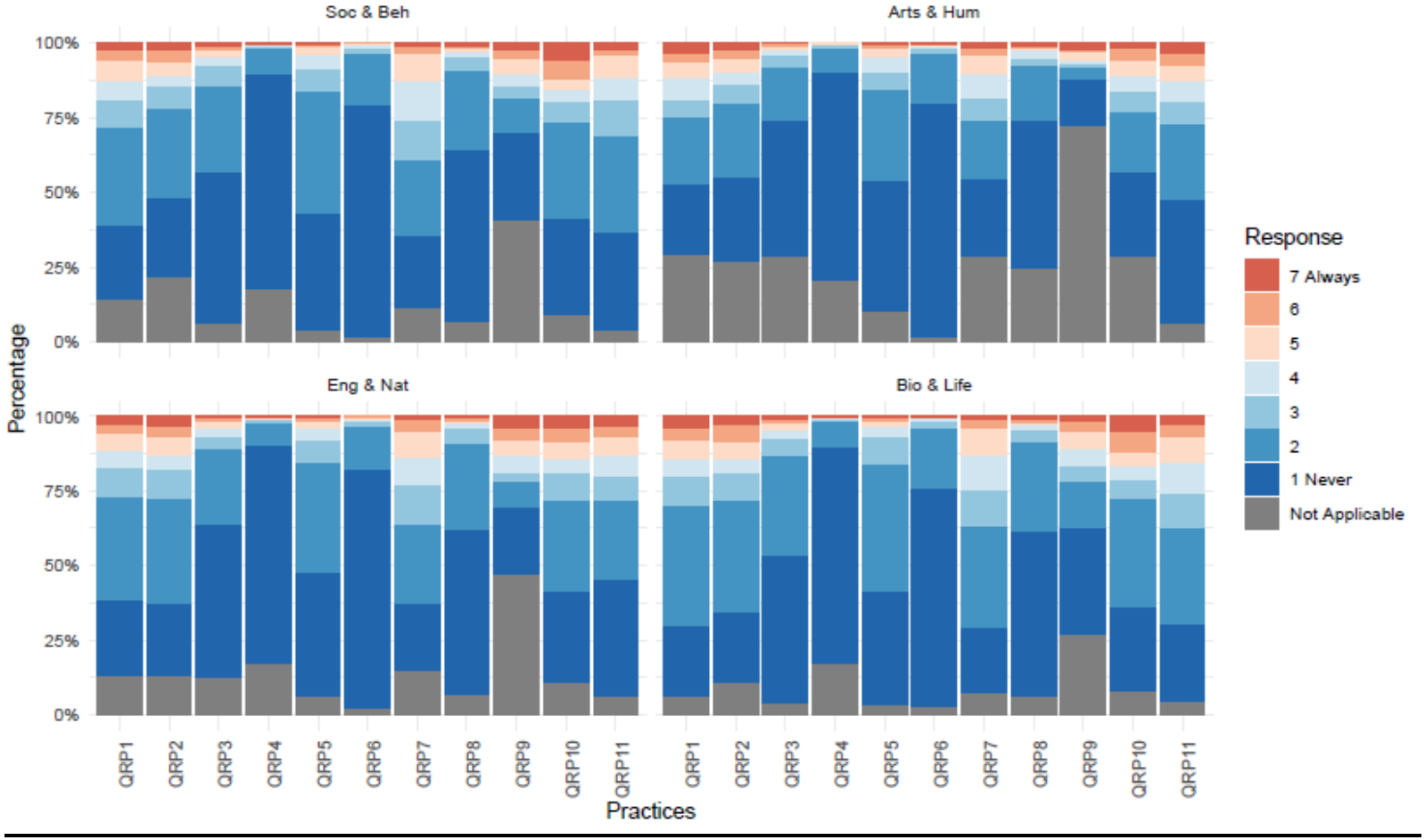
Table S2b: Mean score¶¶ (95% confidence interval) and prevalence+ (95% confidence interval) of QRPs stratified by gender, research type and institutional support

QRP	Description (In the last three years..)	Gender			Being mainly engaged in empirical research	Institutional support
		Male	Female	Undisclosed		
QRP1	Insufficient attention to the equipment, skills or expertise essential to perform my studies	2.5 (2.4,2.5)	2.5 (2.5,2.6)	2.5 (2.2,2.8)	2.5 (2.5,2.6)	2.5 (2.5,2.6)
		14.6 (13.4,15.9)	14.8 (13.5,16.2)	15.6 (9.3,24.8)	14.8 (13.9,15.7)	14.3 (13,15.6)
QRP2	Supervised or mentored junior co-workers	2.6 (2.5,2.6)	2.4 (2.3,2.4)	2.8 (2.5,3.2)	2.5 (2.4,2.5)	2.4 (2.4,2.5)
		15.9 (14.7,17.3)	13.4 (12.1,14.9)	22.3 (14.7,32.3)	14.9 (14,15.9)	14.9 (13.6,16.3)
QRP3	Inadequate research designs or used evidently unsuitable measurement instruments for my studies	1.7 (1.7,1.7)	1.8 (1.8,1.8)	1.5 (1.4,1.6)	1.7 (1.7,1.8)	1.8 (1.7,1.8)
		4.3 (3.6,5)	4.6 (3.8,5.4)	none	4.4 (3.9,5)	4.5 (3.8,5.4)
QRP4	Unfairly reviewed manuscripts, grant applications or colleagues applying for promotion.	1.2 (1.2,1.2)	1.2 (1.1,1.2)	1.1 (1.1,1.2)	1.2 (1.2,1.2)	1.2 (1.2,1.2)
		0.8 (0.6,1.2)	0.8 (0.5,1.3)	none	0.8 (0.6,1.1)	0.7 (0.4,1.1)
QRP5	Drew conclusions that were not sufficiently substantiated by my studies	1.9 (1.8,1.9)	1.9 (1.9,1.9)	1.8 (1.6,1.9)	1.9 (1.9,1.9)	1.9 (1.9,1.9)
		4.1 (3.5,4.9)	3.9 (3.3,4.7)	2.8 (0.7,8.4)	3.8 (3.4,4.4)	4 (3.4,4.8)
QRP6	Used published or unpublished ideas or phrases from others without properly referencing its source	1.2 (1.2,1.3)	1.3 (1.3,1.4)	1.2 (1.1,1.3)	1.3 (1.3,1.3)	1.3 (1.3,1.3)
		0.6 (0.3,0.9)	0.8 (0.5,1.2)	none	0.6 (0.5,0.9)	0.5 (0.3,0.8)
QRP7	Kept inadequate notes of my research process in a project	2.7 (2.6,2.7)	2.6 (2.5,2.6)	2.6 (2.4,2.9)	2.6 (2.6,2.7)	2.6 (2.6,2.7)
		15.8 (14.6,17.2)	13 (11.8,14.4)	12.6 (7,21.4)	14.5 (13.6,15.4)	14.1 (12.9,15.4)
QRP8	Did not mention clearly important details of my study method in my publications	1.6 (1.6,1.6)	1.6 (1.6,1.6)	1.4 (1.3,1.6)	1.6 (1.6,1.6)	1.6 (1.5,1.6)

		2.6 (2.1,3.2)	3.2 (2.6,4)	1 (0,6)	2.9 (2.5,3.4)	2.9 (2.3,3.5)
QRP9	Chose not to submit or resubmit valid negative studies for publication	2.5 (2.5,2.6) 19.8 (18.2,21.5)	2.2 (2.2,2.3) 14.5 (12.9,16.2)	2.8 (2.5,3.2) 20.3 (11.9,32)	2.4 (2.3,2.4) 17.1 (15.9,18.3)	2.3 (2.3,2.4) 16.9 (15.2,18.6)
QRP10	Insufficiently mentioned study flaws and limitations in my publications	2.5 (2.5,2.6) 17 (15.8,18.4)	2.5 (2.5,2.6) 17.2 (15.8,18.8)	2.3 (2,2.6) 10 (5.2,18)	2.5 (2.5,2.6) 17.1 (16.1,18.1)	2.6 (2.5,2.6) 17.1 (15.8,18.5)
QRP11	Selectively cited references to enhance my own findings or convictions	2.4 (2.4,2.5) 13.8 (12.7,15)	2.5 (2.5,2.6) 14.4 (13.2,15.8)	2.4 (2.1,2.7) 12.3 (7,20.4)	2.5 (2.4,2.5) 14 (13.1,14.9)	2.5 (2.5,2.6) 14.4 (13.2,15.7)
Any Frequent QRP	Score 5, 6 or 7 on at least 1 of the 11 QRPs	52.7 (51.0, 54.3)	49.8 (4.8, 51.6)	48.3 (39, 57.7)	52.5 (51.2, 53.7)	50.9 (49.2, 52.6)
Falsification	Making up of data or results	3.7 (1.8, 5.7)	5.1 (2.9, 7.2)	0.9 (0, 11.2)	4.4 (2.9, 5.9)	4.4 (2.4, 6.5)
Fabrication	Manipulating research materials, data or results	3.7 (1.8, 5.6)	4.9 (2.8, 7.1)	3.4 (0, 14.1)	4.6 (3.1, 6.1)	3.9 (1.8, 5.9)
Any FF	Fabrication and/or Falsification	7.0 (4.2, 9.9)	9.8 (6.6, 12.9)	7.8 (0, 10.8)	8.7 (6.5, 10.9)	8.0 (5.0, 11.0)

¶¶*Scores range from 1=never to 7=always;* [+]*Prevalence is based on a Likert score of 5, 6 or 7 among respondents that deemed the QRP at issue applicable; Mean scores of individual QRPs only consider respondents that deemed the QRP at issue applicable.*

Figure S2a: Percentage of observed answer categories of QRPs stratified by disciplinary field



5

Figure S2b: Percentage of observed answer categories of QRPs stratified by academic rank

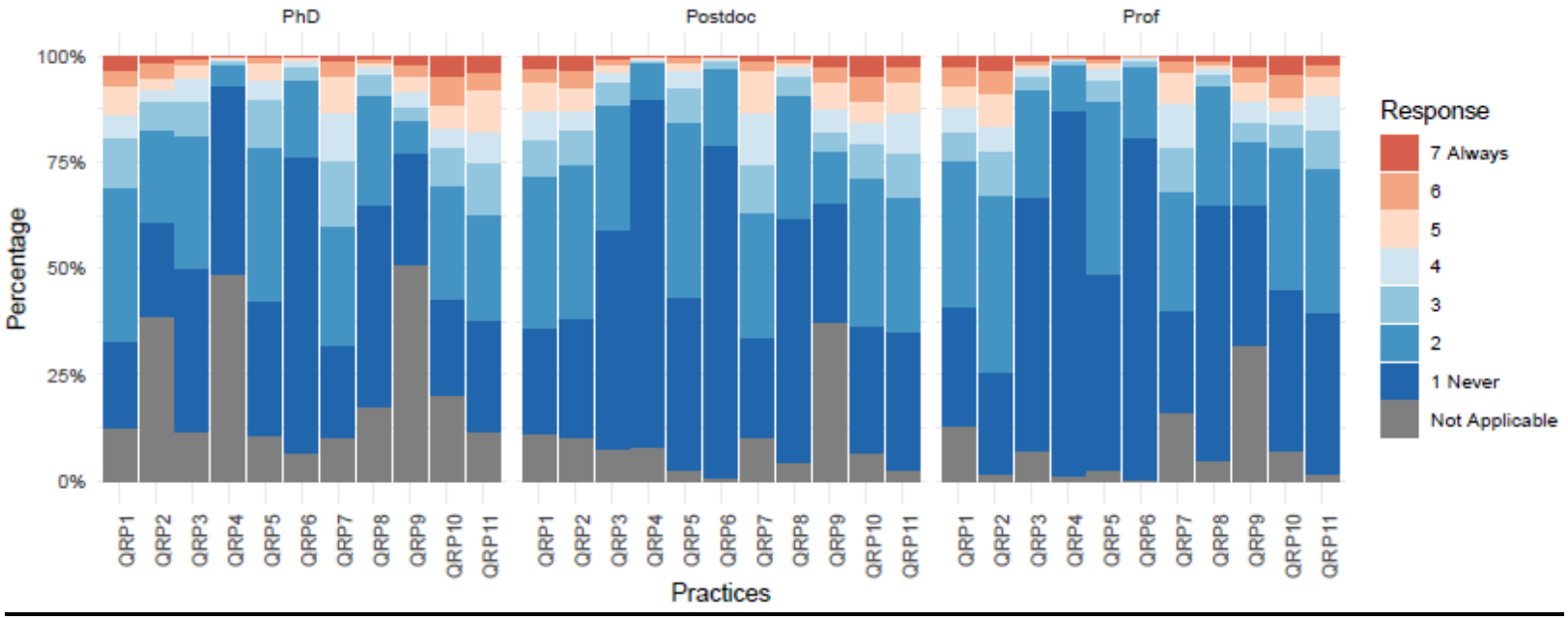


Figure S2c: Percentage of observed answer categories of QRPs stratified by gender

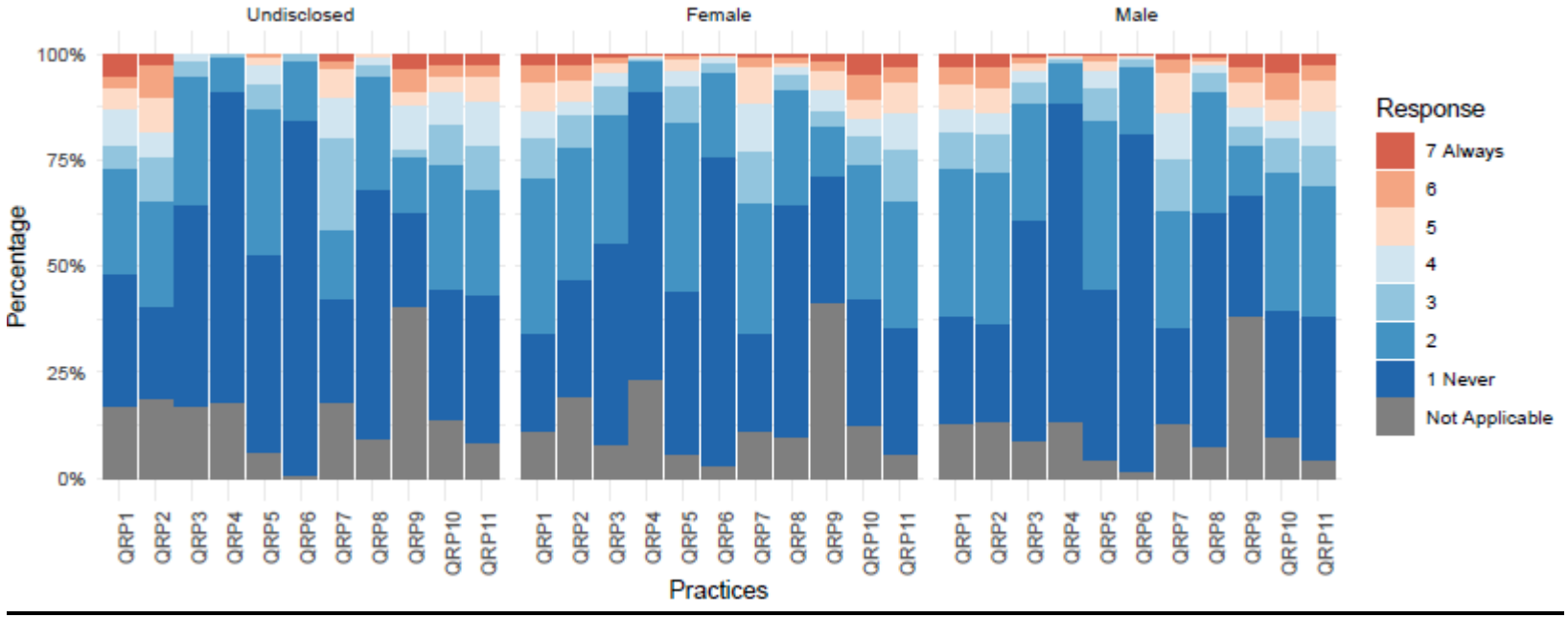


Figure S2d: Percentage of observed answer categories of QRPs stratified by research type (empirical Y/N)

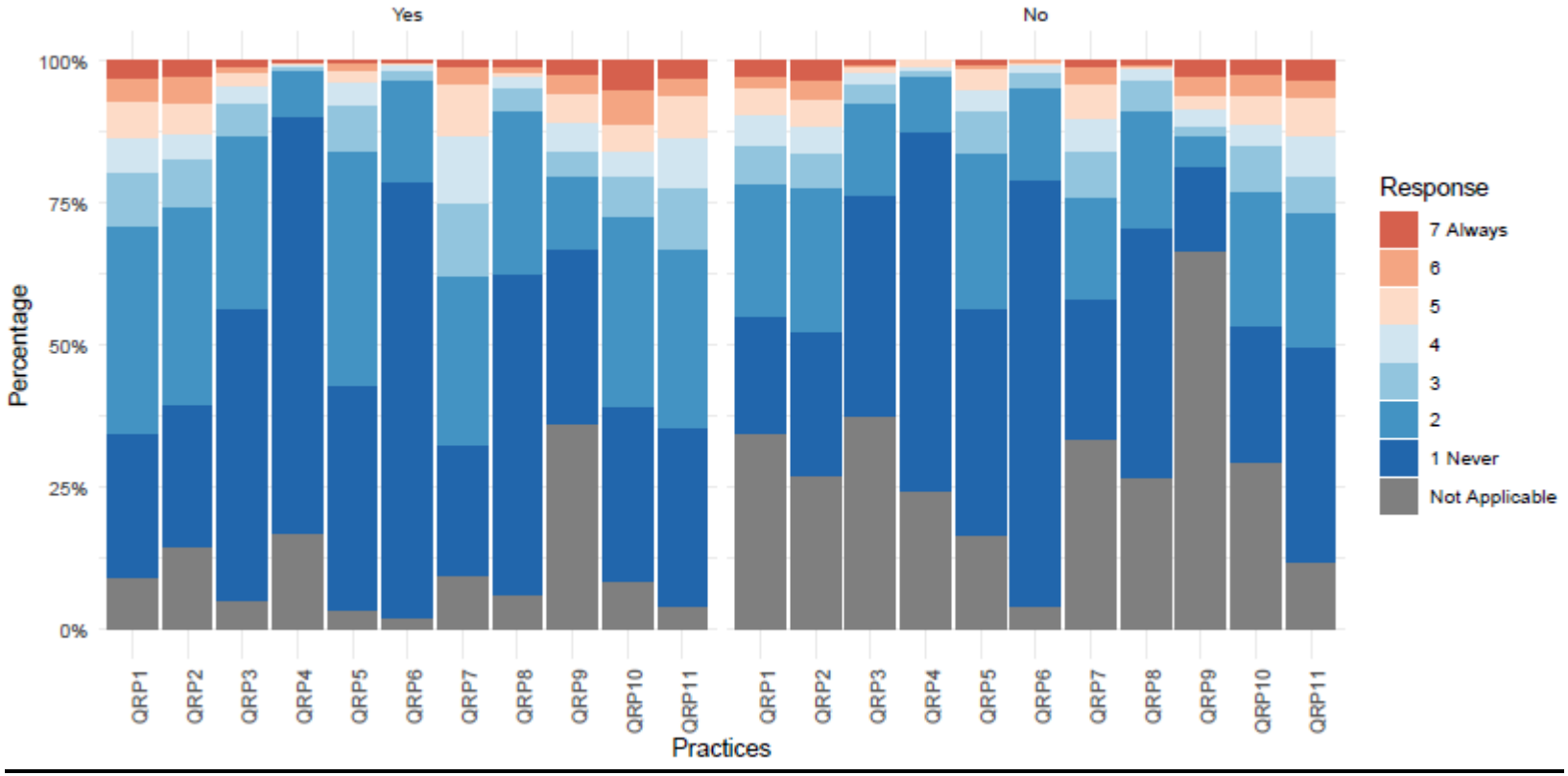


Figure S2e: Percentage of observed answer categories QRPs stratified by institutional support (Y/N)

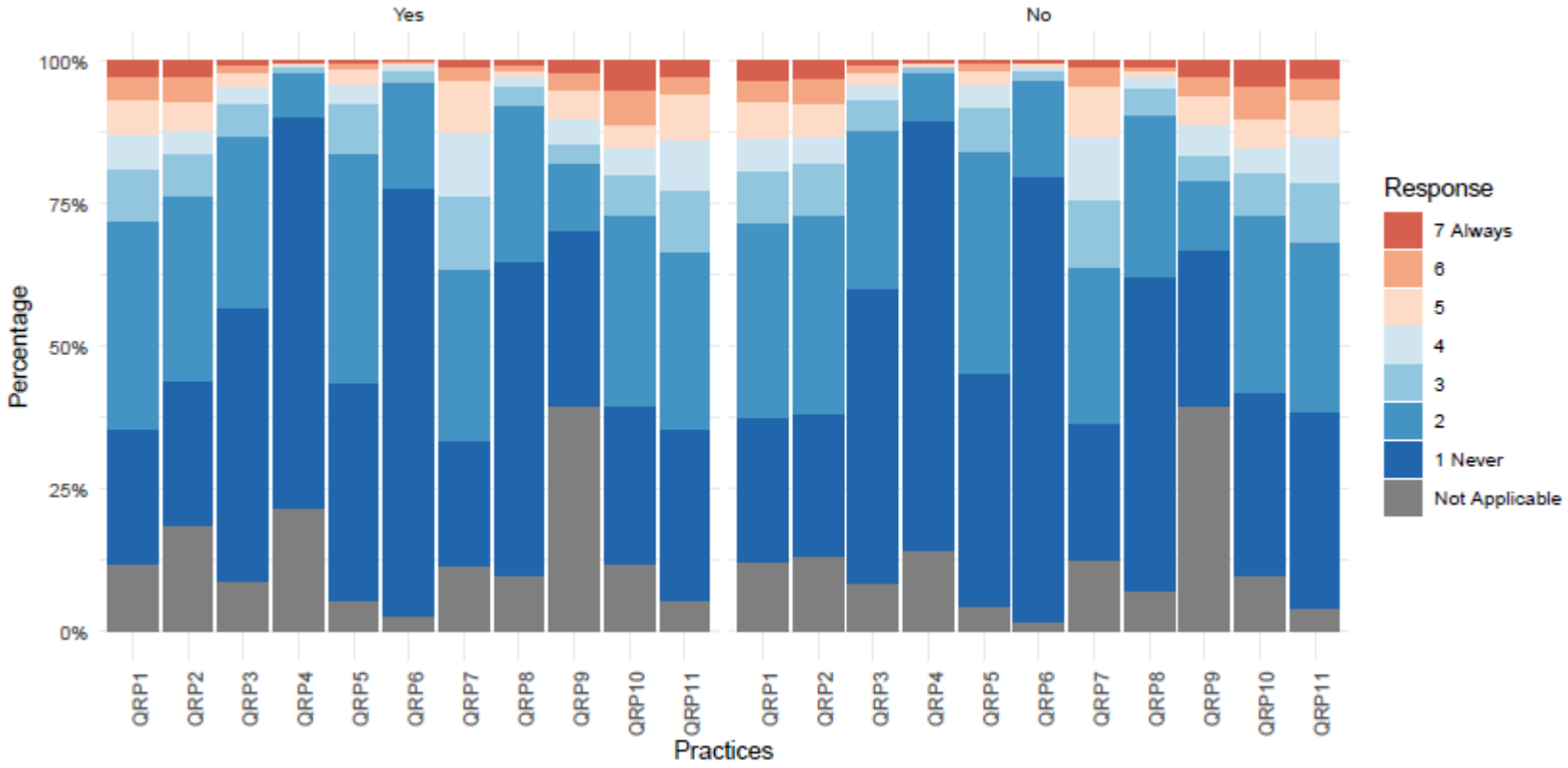


Table S4: Correlation matrix of the z scores of the principal component analysis of the explanatory factor scales

Explanatory factor scale	Scientific norms	Peer norms	Work pressure	Publication pressure	Funding pressure	Competitiveness	Likelihood of detection (collaborators)	Likelihood of detection (reviewers)	Mentoring	Mentoring (survival)	Mentoring (responsible)	Organizational justice	Organizational justice (distributional)	Organizational justice (procedural)
Scientific norms	1													
Peer norms	0.12	1												
Work pressure	0.12	-0.21	1											
Publication pressure	0.09	-0.29	0.39	1										
Funding pressure	0.09	-0.17	0.31	0.38	1									
Competitiveness	0.06	-0.41	0.28	0.44	0.27	1								
Likelihood of detection (collaborators)	0.04	0.11	-0.03	-0.03	-0.01	-0.04	1							
Likelihood of detection (reviewers)	0.07	0.06	0.02	0.01	0.06	-0.02	0.46	1						
Mentoring	-0.04	0.23	-0.16	-0.16	-0.04	-0.29	0.11	0.06	1					
Mentoring (survival)	-0.04	0.17	-0.11	-0.12	0	-0.24	0.08	0.06	0.91	1				
Mentoring (responsible)	-0.03	0.25	-0.19	-0.18	-0.08	-0.29	0.12	0.04	0.92	0.68	1			
Organizational justice	0.02	0.43	-0.33	-0.39	-0.31	-0.43	0.08	0.04	0.35	0.30	0.33	1		
Organizational justice (distributional)	0.02	0.42	-0.33	-0.39	-0.30	-0.43	0.07	0.04	0.33	0.30	0.31	0.95	1	
Organizational justice (procedural)	0.02	0.41	-0.29	-0.35	-0.28	-0.39	0.07	0.03	0.33	0.28	0.33	0.94	0.80	1

Table S5: Prevalence (%) of the "not applicable" answers stratified by disciplinary field and academic rank

		Disciplinary field				Academic rank		
QRP	**Description (In the last three years..)**	**Life and medical sciences**	**Social and behavioural sciences**	**Natural and engineering sciences**	**Arts and humanities**	**PhD candidates and junior researchers**	**Postdocs and assistant professors**	**Associate and full professors**
QRP1	Insufficient attention to the equipment, skills or expertise	6.2	14.2	13.1	29.1	12.6	11.1	13.1
QRP2	Insufficiently supervised or mentored junior co-workers	10.8	21.7	13.3	27.2	38.7	10.1	1.7
QRP3	Inadequate research designs or unsuitable measurement instruments	3.9	6.2	12.5	28.6	11.8	7.7	7
QRP4	Unfairly reviewed manuscripts, grant applications or colleagues	17.5	18.1	17.4	20.4	48.7	7.9	1.2
QRP5	Conclusions not sufficiently substantiated	3.7	4.2	6.3	10.4	10.7	2.6	2.7
QRP6	Improper referencing of source	3	1.6	2.3	1.6	6.5	0.6	0.5
QRP7	Inadequate notes of research process	7.4	11.5	14.8	28.9	10.4	10.5	16.1
QRP8	Failed to report important study details in publications	6.6	6.8	7.2	24.7	17.4	4.6	4.8
QRP9	Not submitting or resubmit valid negative studies for publication	27.4	40.6	47.3	72.3	50.9	37.2	31.9
QRP10	Insufficient inclusion of study flaws and limitations in publications	8.3	9.1	10.8	28.8	20.4	6.8	7.3
QRP11	Selectively cited references to enhance findings or convictions	4.4	4.2	6.3	6.4	11.6	2.6	1.5

Table S6: Full list of the explanatory factor scales and their corresponding items showing which were adapted, newly created or piloted

Explanatory factor scale	Explanatory factor scale item	Reference
Scientific norm description and normative behavior of peers (piloted)	1 and 10. Researchers evaluate research only on its merit.	(*23*)
	2 and 11. Researchers judge each other's contributions primarily on the basis of quality.	(*25*)
	3 and 12. The acceptance or rejection of claims entering the scholarly domain does not depend on the personal or social characteristics of researchers. 7 and 16. No researchers' contribution to knowledge can be accepted without careful scrutiny.	
	4 and 13. Researchers consider all new evidence, hypotheses, theories, and innovations, even those that challenge or contradict their own work. 5 and 14. Researchers are motivated by the desire for knowledge and discovery, and not by the possibility of personal gain.	(*23*)
	6 and 15. Researchers are clear about what data their work is based on, and how results were achieved.	(*42*)
	8 and 17. Researchers put their work in the public domain to be read and used by other scientists and the general public. (n*ew item)* 9 and 18. Researchers derive satisfaction from the mere act of doing research. *(new item)*	
Perceived work pressure	19. How often does it occur that you have enough time to do all the tasks demanded of you? *(adapted item)* 20. How often are you assigned too much work to do in a limited time? *(adapted item)* 21. How often does an excess of work prevent you from having time to rest? *(adapted item)*	(*26*)

Publication pressure	22. I feel a pressure to publish. *(new item)* 23. I experience stress at the thought of my colleagues' assessment of my publications output. 24. I have the feeling that my colleagues judge me mainly on the basis of my publications. *(adapted item)* 25. Publication pressure harms my ability to do good research. *(new item)* 26. The current publication climate puts pressure on relationships with fellow-researcher. 27. Publication pressure sometimes leads me to cut corners. *(adapted item)*	(*22*)
Pressure due to dependence on funding	28. Judgements of my academic performance do not depend on my successful grant applications. *(new item)* 29. My job security depends strongly on research grants I receive. *(new item)* 30. My prospects for promotion depend on me obtaining funding. *(new item)* 31. The continuation of my research depends on obtaining my own funding. *(new item)* 32. I would be able to do my research without obtaining my own funding. *(new item)* 33. Obtaining my own research funding is crucial for my academic career. *(new item)*	
Mentoring (survival)	34. How often have your most important academic mentor(s) provided you with help in learning the art of survival in your field? *(adapted item)* 35. How often have your most important academic mentor(s) helped you in developing professional relationships with others in your field? *(adapted item)* 36. How often have your most important academic mentor(s) provided you with guidance in writing grant and contract proposals? *(adapted item)*	(*25*)
Mentoring (survival)	37. How often have your most important academic mentor(s) coached you in career advancement? *(new item)* 38. How often have your most important academic mentor(s) given you guidance on how to seize career opportunities? *(new item)* 39. How often have your most important academic mentor(s) advised you on how to get your research published? *(new item)*	
	40. My scientific field functions largely as a community of researchers.	(*24*)

Competitiveness of research field *(piloted)*	41. Many scientists in my field are afraid of being scooped by their peers. 42. Many scientists in my field are unhappy when their peers obtain a major award or recognition. 43. Rivalry between researchers is common in my field. 44. Researchers in my field working on similar topics are inclined to collaborate with each other. 45. Most scientists in my field consider their own work to be part of a larger collaborative effort.	
Mentoring (Responsible)	46. How often have your most important academic mentor(s) helped you in presenting the limitations of your research? *(new item)* 47. How often have your most important academic mentors given you feedback on how to select the most robust research methods? *(new item)* 48. How often have your most important academic mentor(s) advised you on making your work as transparent as possible? *(new item)* 49. How often have your most important academic mentor(s) coached you on how to deal with conflicts of interest in your work? *(new item)* 50. How often have your most important academic mentor(s) provided you with insights in the ethical aspects of a research design? *(new item)*	
Mentoring (responsible)	51. How often have your most important academic mentor(s) provided you with guidance on good research practices? *(adapted item)*	(*25*)
Distributional organizational justice *(piloted)*	52. The allocation of resources at my department is fair. *(adapted item)*	(*27*)
Distributional organizational justice *(piloted)*	53. The allocation of responsibilities at my institution is biased. *(new item)* 54. Tenure decisions at my organization are often biased. *(new item)* 55. Decisions about promotion at my department are reasonable. *(new item)* 56. The management at my organization makes reasonable decisions. *(new item)* 57. The assessment of my academic performance is fair. *(new item)*	

Procedural organizational justice *(piloted)*	58. The process of allocating resources in my department is poorly managed. *(adapted item)* 59. The process of allocating responsibilities in my department is ethical. *(adapted item)* 61. The process for promotion at my department is poor. *(adapted item)* 63. At my department, my academic performance is assessed objectively. *(adapted item)*	(*27*)
Procedural organizational justice	60. The criteria for tenure at my department are applied consistently. *(new item)* 62. The management at my department is transparent about their decisions. *(new item)*	
Likelihood of detection by (collaborators/reviewers) *(Piloted)* *(Items are adapted to likelihood of detection format)*	How likely is it that a ***collaborator*** detects that a researcher in your field… 64. Provides insufficient supervision or mentoring to junior co-workers. 65. Does not submit (or resubmit) for publication a valid negative study. (*adapted item)* 66. Keeps inadequate notes of their research process in their project. 67. Uses published or unpublished ideas or phrases without properly referencing the originating source. (*adapted item)* 68. Unfairly reviews papers, grant applications, or colleagues applying for promotion. 69. Fabricates data in their research. How likely is it that a ***reviewer*** detects that a researcher in your field … 70. Draws conclusions that were not sufficiently substantiated by their study. (*adapted item)* 71. Chooses an inadequate research design or uses evidently unsuitable measurement instruments for their study. 72. Gives insufficient attention to the equipment, skills or expertise essential to perform their study. 73. Fails to report clearly relevant details of the study method. 74. Insufficiently reports study flaws and limitations. 75. Selectively cites references to enhance their own findings or convictions.	(*20*)