LUNDS UNIVERSITET

DEPARTMENT of PSYCHOLOGY

Questionable Research Practices, Preregistration, and More – Exploring Self-Report Opinions of Swedish and Dutch PhD Students

Nils Arlinghaus

Master's Thesis (30 hp)
Spring 2018

Supervisor: Zoltan Kekecs, PhD

Abstract

For this exploratory study, I reached out to all 302 individuals that were listed as PhD students at the psychology departments of all Swedish and Dutch "Top 100" universities. The final 111 participants (37.9% response rate) were asked to indicate how often they encountered questionable research practices (QRPs) and replied to items asking about the impact of QRPs on themselves and their environment, thoughts of leaving academia, implications of the "replication crisis", and knowledge and use of preregistrations. The results indicated that QRPs were common, but that the type of QRP mattered. While only 3.7% of participants considered leaving academia because of fear of "replication bullies", 23.9% considered leaving academia because they were disillusioned by the problems revealed by the replication crisis. Results also indicated that the more QRPs a student encountered, the higher the chance that they had thought of leaving because of being disillusioned. Preregistration was widely known, but the majority had never preregistered a project. The study provides first insights in the situation of PhD students in Sweden and the Netherlands and can serve as a foundation for confirmatory research in the future.

Keywords: Questionable research practices, early-career researchers, preregistration, replication crisis

Questionable Research Practices, Preregistration, and More – Exploring Self-Report Opinions of Swedish and Dutch PhD Students

Early-career researchers (ECRs) in psychology face a system and a research community in transition. Since the beginning of the so-called *replication crisis*, many seemingly established psychological theories have failed to replicate, both in conceptual and procedural replication attempts (Doyen, Klein, Pichon, & Cleeremans, 2012; Open Science Collaboration, 2015; Ranehill et al., 2015; E. J. Wagenmakers et al., 2016).

As a reaction to the problems of reproducibility, there was a shift in what is considered good practice in research methods and statistical analyses. Common but problematic behaviors when analyzing and reporting data have increasingly been questioned and are suspected to play a role in the publication bias against null findings (Nelson, Simmons, & Simonsohn, 2018). Such *questionable research practices* (QRPs, explained in more detail below) are different from clearly unethical actions and include examples like deciding whether to exclude data after looking at the impact of doing so, or reporting an unexpected finding as predicted from the start. QRPs are usually seen as distinct from outright fraud and some may even be seen as ethically defensible by researchers (Sacco, Bruton, & Brown, 2017). Research on QRPs is still limited and there is some uncertainty how prevalent they are. Most findings suggested high prevalence (Agnoli, Wicherts, Veldkamp, Albiero, & Cubelli, 2017; Fanelli, 2009; Fanelli, Costas, & Larivière, 2015; John, Loewenstein, & Prelec, 2012), but some authors have contested this, suggesting lower prevalence (Fiedler & Schwarz, 2016). Either way, QRPs could be a missing piece in solving the issue of low reproducibility and are relevant to tackling the methodological crisis in the long term.

For the future of academic culture, the role of current early-career researchers (ECRs) should be considered. Their academic development is influenced by the current system and

they will become the senior researchers in the future. Since they may now still be more malleable when it comes to principles of scientific practice (Anderson et al., 2007; Fanelli et al., 2015), ECRs are a relevant research population for investigating these issues. Their opinions could provide insights in the current status of the changes in the field, and they themselves will act as agents in whether the deeply-rooted problems that led to the crisis can be overcome in the long term (Asendorpf et al., 2013).

The "Replication Crisis"

The years 2011 and 2012 marked a pivotal moment for the research community in psychology (Nelson et al., 2018). Though largely coincidental, a series of events caused a loss of trust in the credibility of psychological research, usually referred to as the "replication crisis" (Nelson et al., 2018; Pashler & Wagenmakers, 2012). One such event was that social psychology's most prestigious journal published a series of studies featuring outlandish claims and stunning flaws in methodology (Eric Jan Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). At the same time, a popular social psychologist was exposed who had successfully built an entire career on fraud without being detected (Stroebe, Postmes, & Spears, 2012), and a famous social priming effect failed to replicate (Doyen et al., 2012). In connection with these cases, some pointed out how extremely rare replication attempts had been in published research (Makel, Plucker, & Hegarty, 2012). Many researchers started to question if large parts of published research were even reliable and replicable in the first place, or if they actually constituted false-positives (Ferguson & Heene, 2012; Pashler & Wagenmakers, 2012; Simmons, Nelson, & Simonsohn, 2011; Yong, 2012). In the wake of the events and the discussions, replication attempts became more common, including the announcement of a large, independent replication project (Open Science Collaboration, 2012). However, an alarming number of replications of published research only produced smaller effects than the original studies or failed to replicate them altogether (Hagger et al.,

2016; Nelson et al., 2018; Open Science Collaboration, 2015; Simons et al., 2014; E. J. Wagenmakers et al., 2016). There were several name suggestions for it[1], but whatever one called it – the "replication crisis*"* was happening.

But how could an entire research field get to a point where large parts of published research fail to replicate? At the root of this problem was something that Nelson and colleagues (2018) described as an old paradox. While most publications in psychology describe statistically significant findings, the vast majority of published studies also lacked sufficient statistical power and should thus not even have been able to obtain statistically significant results. A common explanation for this paradox was the *file-drawer problem*, which had been described long before the beginning of the replication crisis (Begley & Ioannidis, 2015; Rosenthal, 1979; Scargle, 2000). The file-drawer problem is the situation in which studies reporting statistically significant results are strongly favored for publication by journals. As a result, non-significant studies are much less likely to be accepted for publication and consequently end up in the "file-drawer", leading to an incomplete picture of findings and effects in published literature (Greenland et al., 2016; Rosenthal, 1979). This publication bias against null results makes replications difficult and problematic, as they can only be truly meaningful if they will also be published in case of failure to replicate (Ferguson & Heene, 2012).

Even though the problems of publication bias and the file-drawer explanation had been described for decades (Ioannidis, 2005; Rosenthal, 1979; Scargle, 2000; Sterling, 1959), they were not addressed on a larger scale before the beginning of the replication crisis. Ferguson and Heene (2012) argued that the field's aversion towards null results makes replication meaningless if failed replications are less likely to be published. But they went on

[1] I would like to explicitly state that calling the situation since 2011 a "replication crisis" is not helpful, in that it might create the impression that something went wrong in 2011. The opposite is true: 2011 hopefully marked the beginning of the end of a crisis, since the problems are finally being addressed. I wholly agree with Nelson and colleagues (2018) that this process would better be described as *psychology's renaissance*, but to avoid confusion I use the most common term (which still is "replication crisis").

to also suggest that behavior on the individual level may further reduce null results, namely through questionable research practices. This fits Nelson and colleagues (2018) suggestion that the file-drawer explanation is an unrealistic depiction of researcher's behavior, as it assumes that a researcher would simply put a non-significant study in the file-drawer and start all over again. An alternative explanation that they (and others) propose are behaviors that "produce" significant results. One is the so-called *p-hacking*; by first trying different data eligibility criteria or performing several statistical procedures and then only reporting the ones that lead to significant results, null results are avoided (Head, Holman, Lanfear, Kahn, & Jennions, 2015). Similarly, *HARKing* describes the practice of hypothesizing after already knowing the results (Kerr, 1998). By failing to report this temporal order, postdiction is presented as prediction (Nosek, Ebersole, DeHaven, & Mellor, 2018). Importantly, these practices are not limited to active decisions or conscious behavior of researchers. Instead, there are many steps of conducting a research project that allow for decisions that will avoid null results (Gelman & Loken, 2014). Behaviors like p-hacking and selective reporting are some examples of what have been described above as questionable research practices and could explain the replication crisis through a high number of false-positives among published findings.

Questionable Research Practices

When conducting research projects, researchers need to continuously make decisions, particularly when unexpected situations occur. Although guidelines exist and even the official publication manual of the American Psychological Association (2010) suggests that "omitting troublesome observations from reports to present a more convincing story is […] prohibited" (p.12) (see also: Giner-Sorolla, 2016), researches may still make incorrect methodological or statistical decisions. The list of such examples is long: Selectively analyzing or reporting data, study conditions, experimental categories, and statistical tests;

adjusting hypotheses and analyses post hoc; or continued sampling when not appropriate (Sijtsma, 2016). These all fall under the umbrella of questionable research practices (QRPs), defined by Sijtsma, Veldkamp, and Wicherts (2016) as "debatable, disputable, doubtful, and problematic practices in setting up studies, collecting data, analyzing data, and reporting of methods and results" (p. 37).

QPRs had previously been described as falling somewhere between the ideal *responsible conduct of research* (RCR) and deliberate misconduct, usually referred to as *fabrication, falsification, and plagiarism* (FFP) (Steneck, 2006). Depending on the context, some QRPs might be seen as defensible by researchers (Sacco et al., 2017). But since the beginning of the replication crisis, there has been growing attention to the role they play for the publication bias against null-findings, and thus for the replication crisis at large (John et al., 2012). Just how widespread some QRPs might be in published research became evident in the findings of Franco, Malhotra, and Simonovits (2016). They investigated selective underreporting by comparing 32 registered psychology studies with their published results and found that around 40% did not report all experimental conditions and 70% did not report all outcome measures that were included. This percentage is likely to be even higher for studies that are not registered (Franco et al., 2016). Additionally, the authors found that reported effect sizes were twice as large as unreported ones, and three times as likely to be significant. This was further evidence that the file-drawer explanation cannot solely explain the publication bias and suggested that researchers may play a more active role by predominantly presenting large and significant effect sizes. Simmons and colleagues (2011) demonstrated in their paper that QRPs like "flexibility" in data collection, analysis, and reporting greatly increased the likelihood of finding support for a false hypothesis – potentially making anything statistically significant.

Prevalence of Questionable Research Practices

Fanelli (2009) reported the results of a meta-analysis (18 surveys) and a systematic review (21 surveys), which also included fields other than psychology. The results were that an average of around 33% of researchers indicated having engaged in QRPs, and an average of 72% indicated having encountered colleagues engaging in them. Though these findings were not specifically based on psychologists, they provided first evidence that suggests that QRPs are not *uncommon*. After the beginning of the replication crisis, the first published study directly investigating QRPs in a large sample of psychologists was by John and colleagues (2012). They gathered self-report data from 2155 psychologists working at major universities in the United States (36% response rate, including incomplete responses). The results showed that "approximately 35% of respondents indicated that they had doubts about the integrity of their own research on at least one occasion" (p.528). Participants admitted to some QRPs more frequently than to other QRPs, with 25% of respondents admitting to reporting an unexpected finding as predicted from the start, more than 45% admitting to selectively reporting studies that worked, and more than 60% of respondents admitting to not reporting all of a studies dependent measures, each on at least one occasion (John et al., 2012). These findings were frequently cited as evidence for the high prevalence of QRPs among psychologists. In an Italian adaption of the study, Agnoli, Wicherts, Veldkamp, Albiero, and Cubelli (2017) found similar prevalence rates for most QRPs, concluding that the practices were likely an international phenomenon. However, research from other countries is still scarce.

Fiedler and Schwarz (2016) contested the high prevalence and criticized both the phrasing of some items in John and colleagues' (2012) original study, as well as the fact that their main dependent measure was the proportion of respondents admitting having engaged in the behaviors once in their life. Fiedler and Schwarz (2016) suggested changes to the study

design, rephrased items, and included an additional prevalence estimate to the “once in their life” admission proportions used in the original study. They then collected answers to their modified survey from 1138 members of the German Psychology Association (35% response rate). The prevalence estimate produced a lower percentage compared to John and colleagues' (2012) “at least once in a lifetime” admission rates. Nevertheless, prevalence rates of researchers engaging in QRPs such as claiming to have predicted an unexpected result or failing to report all dependent measures relevant for the finding were still above 20%. Considering the likelihood of social desirability and similar biases of self-report studies, these proportions are likely to be conservative estimates. Additionally, even 20% of researchers engaging in these practices would warrant action.

Overcoming Questionable Research Practices

It is important to recognize that researchers who engage in QRPs are not necessarily aware of doing so. An example are QRPs that stem from a lack of experience with methodology or statistics (Sijtsma, 2016). If awareness or intention are not always given, how far individuals can be held accountable for QRPs might be debatable. Some authors suggested that because of this, there should be a focus on policies rather than individual behavior (Sijtsma, 2016). Fanelli and colleagues (2015) conducted a large, retrospective study based on retracted papers and linked them to bibliographic and personal information of all co-authors. They found that misconduct policies, the academic culture, and the career stage of the researchers affected their scientific integrity, but that the pressure to publish did not (Fanelli et al., 2015). Generally, it is difficult to identify individuals engaging in QRPs, particularly if they are unaware or unwilling to change. Holding them accountable is often only possible through whistle blowing (Fanelli, 2013).

A possible solution for reducing at least some of the main QRPs was described by Simmons and colleagues (2011). To avoid false positives in published literature, they

proposed requirements for authors on the one hand, and guidelines for publishers on the other hand. The requirements for authors can be summarized as full, a priori disclosure of the study design and planned analyses with appropriate power. The role of publishers would then be to enforce these requirements. While being tolerant of null results, publishers should hold authors accountable for their claims and if necessary, demand justifications and a replication. Similarly, LeBel and colleagues (2013) argued for mandatory methods disclosure statements for all psychology journals to address these and similar issues that are in the way of reliable and replicable findings. Sijtsma (2016) argued that rather than assigning blame to individuals, it might be more effective to change institutional policy to discourage QRPs. Similar to the requirements for authors that Simmons and colleagues (2011) had suggested, open data and preregistration of projects could increase transparency in the research process (Nosek & Lakens, 2014; Rouder, 2016).

Preregistration

One main way that has been suggested to lower the chance of QRPs early on in the research process is to preregister studies and analyses ahead of time (Nosek & Lakens, 2014). This would make unconscious QRPs less likely and conscious QRPs at least more difficult. A preregistration includes a priori descriptions of what exactly is being investigated, how the data will be collected and analyzed, and other relevant aspects of the study. Through this, it becomes harder for researchers to fool themselves or others by omitting experimental groups, unconfirmed hypotheses, changing the “story” of a study, or other a posteriori changes and hindsight bias (Nosek et al., 2018). Preregistration would also allow the scientific community to check analyses and conclusions more readily, increasing transparency and accountability. At the same time, a culture shift towards preregistration in combination with peer review could in fact decrease personal responsibility and increase the responsibility of the scientific community to improve studies a priori (Nelson et al., 2018; Nosek et al., 2018).

"Methodological Terrorism" and "Replication Bullies"

Related to the role of the scientific community at large, there have been discussions about the tone in the discourse since the beginning of the replication crisis. Coinciding with the increased number of failed replications, there was more criticism toward established and previously very successful researchers, whose "classics" failed to replicate (Gelman, 2014; Schimmack, 2014). These criticisms were often voiced in scientific blogs and the discussions increasingly took place online (internet blogs and social media). Since researchers were called out for bad methodology, some perceived this as personal attacks. While saying that the field was moving to the right direction methodologically, Fiske (2017) said that more personal statements constituted bullying, and that they created a "chilling, hostile work environment" (p. 653). Importantly, Fiske (2017) claimed based on anecdotal evidence that junior researchers might be afraid to speak up in online discussions or would even be driven to leave academia out of fear of bullying. However, at the time of the present paper, there were no published investigations of how junior researchers think about this.

Early-Career Researchers

Early-career researchers (ECRs) such as PhD students will be the next generation and are thus relevant to the changes in the field (Stürmer, Oeberst, Trötschel, & Decker, 2017). If some of them are driven away from academia because of the replication crisis, fear of bullying, or other reasons, this ought to be understood and prevented. At the same time, some authors suggested that young researchers are not only more malleable (Anderson et al., 2007; Asendorpf et al., 2013), but they were also more likely to be authors of retracted papers (Fanelli et al., 2015), suggesting that QRPs may be particularly relevant for this demographic. Additionally, current PhD students most likely started their doctoral programs after the beginning of the replication crisis, making the frequency with which they have encountered QRPs a more accurate measure of what might have already changed. Published literature that

investigates QRPs from the perspective of early-career research is scarce. In an exploratory study, Stürmer and colleagues (2017) asked 88 German "pre- and postdocs" (36% response rate) how common they believed QRPs to be. Additionally, they were asked about possible causes for researchers engaging in QRPs, and open science initiatives. The authors proposed that if ECRs perceive QRPs to be common, this could make them leave the field, or it could create the perception that it is normal and thus okay to engage in QRPs. Most respondents indicated that QRPs were moderately or highly prevalent, apart from outright fraud (which was also included among the QRPs). Furthermore, most had heard of the open science movement, and open science practices were deemed necessary by more than half of the sample. These findings were interesting, but due to the exploratory nature of the study and the German research population, the conclusions are for non-German populations are limited. Past research by Fanelli and colleagues (2015) indicated significant differences in culture when it comes to academic misconduct, and Germany in particular has an academic culture that differs from other countries (Stürmer et al., 2017).

At the time of the present study, there was no published research on attitudes about QRPs among psychologists in Sweden or the Netherlands, and thus also none on the subpopulation of early-career researchers. Due to their potential role for false-positive studies in published research, and because studies from other countries suggested at least some prevalence, these gaps needed to be addressed and possibly closed. The present study was planned as a first step towards doing so. Since there was no prior research to build on from Sweden or the Netherlands, an exploratory rather than a confirmatory design was chosen for this study. This allowed for a wide overview of possibly important factors, while providing the most reliable information for future research in relation to feasibility.

Aims of the present study

The general aim of the study was thus to gather broad information on the status quo in Sweden and the Netherlands, both in terms of QRPs and the research culture in which they may occur. Early-career researchers could provide general estimates of frequency and recency, and their opinions on a wide range of related issues could be used to investigate possible connections with QRPs. From this, the central research question of the project was thus:

Exploratory Research Question: *What is the role of the research culture that ECRs experience in relation to encountering QRPs and the other aspects related to the replication crisis?*

More precisely, the goal was to ask ECRs how they felt about the impact of the replication crisis on their career, about speaking out against senior researchers, about "methodological bullying", and about credibility problems with older publications. This included the direct assessment of how many participants had considered leaving academia, to test both the suggested risk of young researchers leaving out of fear of bullying, as well as an alternative explanation. Finally, since preregistration has been described as important for reducing QRPs, it was of interest how well-known preregistration was in general and how many research projects ECRs had preregistered in the past. With this range of exploratory information, the overarching goal was to lay a foundation for future research on how to include early-career researchers in the process of improving the credibility of psychological research.

Methods

Preregistration

Before the start of data collection, a general description of the study as well as all materials were uploaded to a project page on Open Science Framework

(https://osf.io/bmcdw/). An embargo was imposed to the project page before data collection began to prevent any changes. To ensure that participants would not be able to access any information that could influence their answers, the project page was not publicly available until May 21st, 2018.

Participants

The participants were 111 PhD students that were pursuing a PhD in psychology at one of seven large universities in Sweden and the Netherlands. They all participated voluntarily and no incentives other than helping with this research were offered.

Recruiting of participants. An exhaustive list was compiled that included all PhD students at psychology departments of all Swedish and Dutch universities ranked amongst the Top 100 of the *Times Higher Education World University Ranking 2018* (five universities from the Netherlands, two from Sweden). The manually compiled list included the contact information of all 302 individuals that were listed as PhD students, doctoral students, PhD candidates, or doctoral candidates on the public departmental staff websites in February 2018. Email invitations for participation were sent in March 2018. This initial email request and a reminder two weeks later were the only way of recruiting participants and there was purposely no contact with their departments. The only eligibility criterion to be included in the study was that participants needed to be current PhD students/candidates.

Materials

The questionnaire was created as a short online survey and started with an informed consent form. In the form, the study was explained as being aimed at understanding the culture of the psychological research community and its implications for early-career researchers. Participants were assured that all responses would be completely anonymous and neither they nor their university could be identified, and that the anonymous dataset would later be uploaded to the Open Science Framework. They were informed that there were no

known risks to participating and that they could end their participation at any time (see Appendix A or https://osf.io/4zrnj/).

Demographics. After giving their informed consent, the first page of the survey asked participants if they were currently PhD students (eligibility criterion), if it was their first PhD involving empirical research, how long they had been in their current position, and if their PhD project was an independent research project (they were the lead researcher), or whether they worked under a lead researcher (as part of a larger research effort).

QRP information. The participants were shown a short paragraph explaining that QRPs fall between *responsible conduct of research* (RCR) and *fabrication, falsification, and plagiarism* (FFP) (Appendix B). The text and an accompanying figure were based on Steneck (2006) and were purposely formulated in a neutral way to lower the chance of social desirability.

QRP frequency. Next, the participants were shown a list containing 15 QRPs that was compiled based on previous research (Fiedler & Schwarz, 2016; John et al., 2012; Sacco et al., 2017), see Table 1 (Appendix C) for a complete list. For each QRP, they had to answer the question "*I have personally encountered this during my career (e.g. by colleagues, senior researchers, or when doing so myself)*" with *Yes* or *No*. As Fiedler and Schwarz (2016) noted, questions like this fail to distinguish between respondents who encountered behaviors only once in their life, and those who do so frequently. This might lead to inflated prevalence rates if QRPs are rather rare. To explore this possibility, the participants were asked two additional questions about the general frequency in the past year. The first and main question was "*Did you encounter at least some of these practices* in the past year*?"* (*Yes/No*), to establish if QRPs are common. The second question was included as a more strongly phrased contrasting item to reduce social desirability for the first question, it was "*Did you encounter some of these practices frequently* in the past year*?"* (typographical emphasis in both items retained,

italics were the opposite). Next to its function as a contrasting item, the second question served as a direct measure of those who encountered QRPs with very high frequency and wanted to make this clear.

General questions about impact of QRPs and replication crisis. On the following page, participants were asked to indicate their agreement with six statements about the impact of QRPs, the ability and willingness to speak up against them, and worrying about being attacked for bad methodology or not being able to trust older research findings due to QRPs (for a complete list, see Table 2 in Appendix C, or https://osf.io/e3qgh/). They were asked to indicate their agreement using a slider scale from *0* to *100* with no numerical indicator, with 100 representing full agreement. Finally, they were asked to agree or disagree with the statements "*I considered leaving academia for fear of "replication bullies" (i.e. fierce criticism of research methods).*" and "*I considered leaving academia because I got disillusioned by the problems in the field revealed by the replication crisis*", answering for each with *Yes* or *No*.

Preregistration and comments. On the last page, the participants were asked whether they believed that psychological research had changed considerably since the beginning of the replication crisis (*Yes*/*No*), whether they knew what preregistration of research projects was (*Yes*/*No*), and how many research projects they had preregistered before (numerical answer field). Finally, they were asked if they had any comments, suggestions, or experiences they would like to share.

Design and Procedure

Since there is only limited and partly contradictory research published on the frequency of QRPs, this study was non-experimental and strictly descriptive and exploratory, to serve as a guide for future research. After preregistration of the project on the Open Science Framework (OSF), the data collection began on March 8th, 2018 at 09:00am

(GMT+1). In two batches, email invitations were sent to all PhD students, requesting help for a master thesis by filling in a short questionnaire about the research culture that PhD students face. A weblink in the email invitation brought participants to the online survey, hosted through the paid online survey service *SurveyMonkey*. After being reassured of their anonymity and giving their informed consent, participants answered the eligibility and background questions about their PhD position. Next, they received the information about QRPs and then the list of QRPs in random order. The following page included the general questions about the impact of QRPs, the ability of speaking up against them, and psychological research in general. Finally, they replied to the preregistration items and could leave comments. The participants were then thanked for their participation. Data collection was closed on April 10th, 2018 at 4:30pm (GMT+1).

Results

The general aim of this exploratory study was to provide information on what Swedish and Dutch PhD students think about different topics related to the replication crisis. The results primarily contain frequency information on the responses of the participants, as well as the attempt to find possible relationships between variables based on their correlations and a logistic regression model. The dataset, scripts, and other information about the analyses will be available on the Open Science Framework (see https://osf.io/bmcdw/).

Participants

Response rate and dropouts. Invitations for participating in the survey were sent to 302 email addresses of PhD students. Eight individuals indicated through personal or automated replies that they were on parental leave or out of office until after the end of data collection, or that they were no longer PhD students. One individual's email address was not valid. Out of the remaining 293 potential participants, 124 participants followed the email invitation and started the online survey. Eight participants were excluded because they

indicated that they were not currently PhD students/candidates. Another five respondents did not proceed past the demographics page. Thus, the final sample of participants contained the responses of 111 participants, which corresponds to a response rate of 37.88%. Two of those participants did not finish the later section of the survey, thus only their answers of the first section were included where possible.

Type of PhD project. Since the motivations and experiences of PhD students could differ depending on their autonomy in conducting research, they were asked about the nature of their PhD position in terms of being the lead researchers themselves or working under a lead researcher. The proportion of both groups was roughly equal, with slightly more respondents indicating that they were lead researchers themselves (53.2%). Participants were also asked if this was their first PhD involving empirical research. Out of all participants, only one participant (0.9%) indicated that this was not the case, suggesting that the sample indeed consisted of early-career researchers.

Questionable Research Practices

Only after the end of the data collection, I noticed that two of the 15 items on the list of QRPs were in fact almost exact duplicates ("*Reporting an unexpected finding as having been predicted from the start.*" and "*Reporting an unexpected result as having been hypothesized from the start.*"). To avoid inflating the total number of encountered QRPs, the item with the higher frequency (40.54% as opposed to 39.64%) was removed, leaving a more conservative estimate. For the majority of participants (90%), the responses to these two items had been identical. Figure 1 shows the percentages of respondents having encountered each of the remaining 14 items that were included in the following analysis. The two QRPs that most respondents had personally encountered before in their career were "*Drawing strong inferences from underpowered, but statistically significant results.*" (45.05%), and "*Publishing results of a single study as several articles, to increase the number of*

publications derived from the research." (41.44%). The two least frequently encountered QRPs were "*Changing the design, methodology, or results of a study to please a sponsor.*", which only two respondents had encountered (1.80%), and "*Not reporting some potentially relevant conflicts of interest.*", which none of the respondents had ever encountered. These two least frequent QRPS were the only items related to conflicts of interest, suggesting that this is not common for most PhD students, or that these events are more concealed from them.

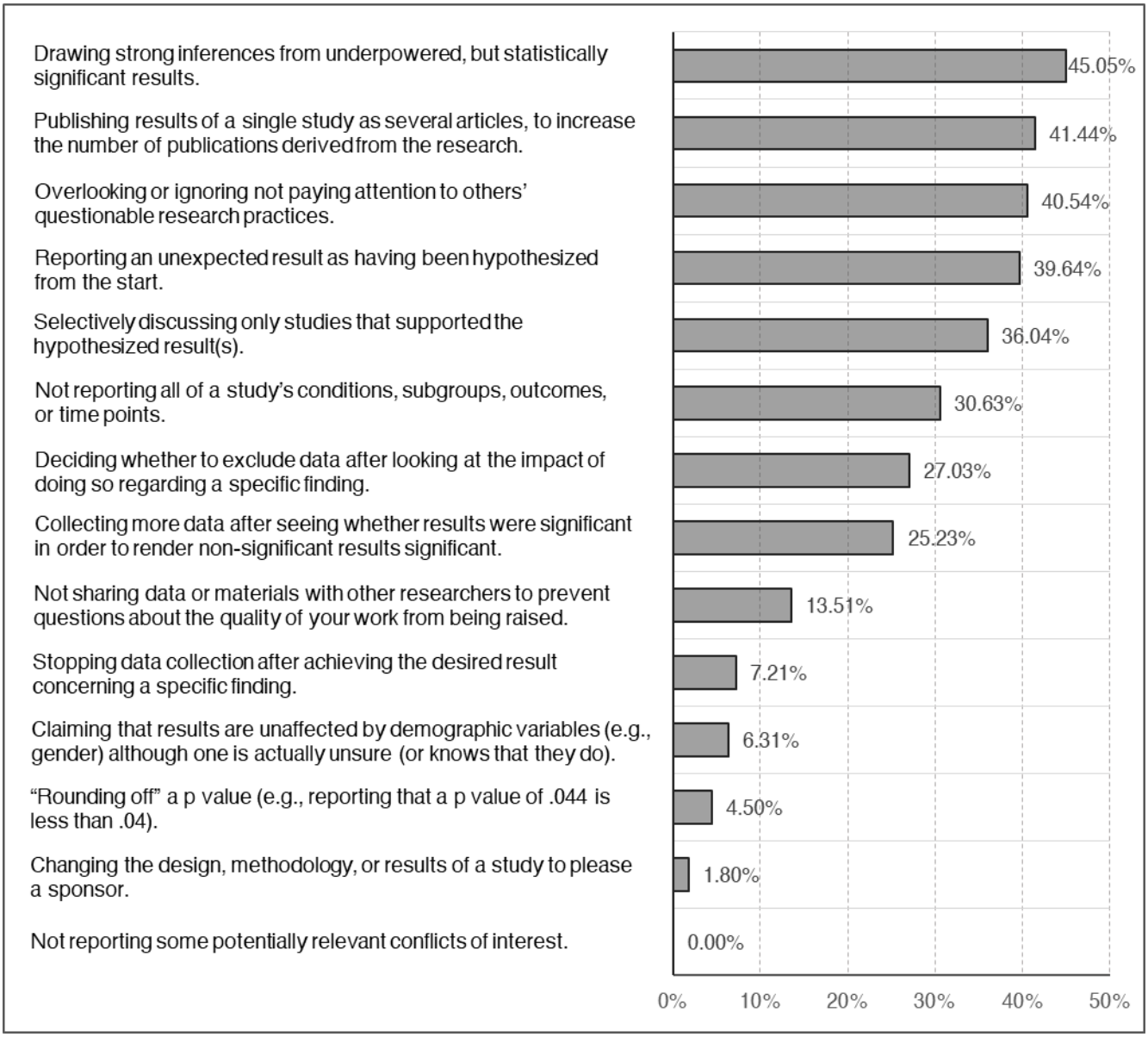


Figure 1. Percentages of respondents who personally encountered each questionable research practice in their career (by colleagues, senior researchers, or themselves).

Past year. Since the list of QRPs merely asked whether participants had personally encountered them in their career, the two items following the list were aimed at providing more information about the recency and magnitude of QRPs. More than half (63.96%) had encountered at least some of the listed QRPs in the past year, highlighting that QRPs are more common than just "one off" experiences. However, only few respondents (10.81%) indicated that they had encountered QRPs frequently in the past year. Table 3 shows the percentages for the two "past year" questions and the average total sum of encountered QRPs, including subgroups for respondents being the lead researchers or not. In the group of those who were lead researchers themselves, slightly more respondents indicated having encountered QRPs in the past year and frequently in the past year. This was also reflected in a higher number of total QRPs encountered, as can be seen by comparing the means in Table 3.

Table 3

Responses to two QRPs frequency items and total encountered QRPs, subgroups for those who were lead researchers themselves and those who work under a lead researcher.

Item		Total %		Nature of PhD project			
				Lead researchers themselves		Under a lead researcher	
		No	Yes	No	Yes	No	Yes
I encountered at least some of these practices in the past year.		36.04%	63.96%	32.20%	67.80%	40.38%	59.62%
		n = 40	*n* = 71	*n* = 19	*n* = 40	*n* = 21	*n* = 31
I encountered some of these practices **frequently** in the past year.		89.19%	10.81%	88.14%	11.86%	90.38%	9.62%
		n = 99	*n* = 12	*n* = 52	*n* = 7	*n* = 47	*n* = 5
Total of encountered QRPs[a]	*M*	3.19		3.63		2.69	
	SD	(2.51)		(2.66)		(2.24)	

Note. *N* = 111.
[a] This total of QRPs still includes two participants that were not included in the later total.

Total number of QRPs. As a general informative measure and for analyses described below, the sum score of all encountered QRPs was computed for all participants that had

completed the entire questionnaire (N = 109, see Figure 2). Less than a quarter of participants (18.35%) indicated that they had never encountered any of the listed QRPs and the majority (81.65%) did encounter at least one QRP at some point of their career. The sum scores ranged from 0 to 10 encountered QRPs, with a median sum of three QRPs (M = 3.21, SD = 2.52).

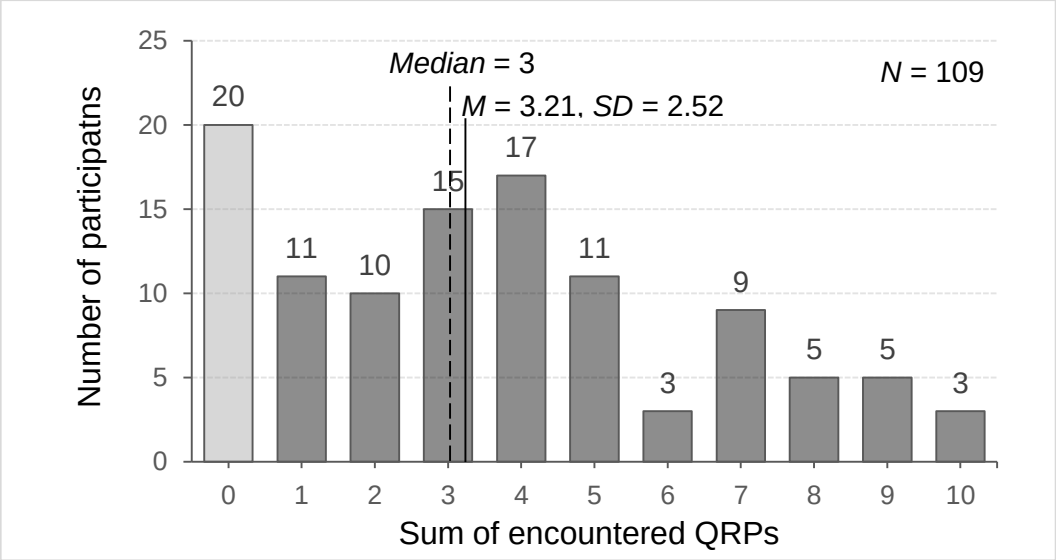


Figure 2. Number of participants that encountered each sum of QRPs.

One possible concern with creating such a single frequency score using the sum of QRPs is that it could misrepresent those who encounter only a few QRPs, but very frequently. Thus, their total QRP score would be very low, even though they are actually encountering QRPs very frequently. The past year frequency questions can be used to diagnose this potential validity issue, Table 4 shows the crosstabulation of the responses of the past year frequently question and the total QRP score. Among those who had encountered a total of less than three

Table 4

Crosstabulation of responses of the frequently in the past year question and total QRPs

	Number of total QRPs						
Encountered QRPs frequently?	0	1	2	3	4	5	>5
No	20	13	11	19	15	6	14
Yes	0	0	1	0	1	1	8

Note. N = 109

different QRPs, there was only one respondent who reported that they encountered QRPs frequently in the past year. This indicated that there was only a single person who encountered few different QRPs, but those frequently.

Correlates of QRPs. To inspect how encountering questionable research practices relates to factors of academic culture and to provide directions for future research, a Pearson correlation table for the total number of encountered QRPs and all variables of culture was computed (Table 5). There was a positive relationship between encountering QRPs and indicating that QRPs affect the own career (r = .43, n = 109) and between encountering QRPs and worrying about trusting old research findings (r = .37, n = 109). An increase of encountered QRPs was thus correlated with an increased feeling that they affected one's career, and increased worrying about the credibility of old research. It is probable that in a confirmatory model, these variables would provide predictive information.

Table 5

Correlations between total number of encountered QRPs and all items about culture.

		1.	2.	3.	4.	5.	6.
1.	Total encountered QRPs	--					
2.	QRPs are a problem	.27	--				
3.	QRPs affect own career	.43	.54	--			
4.	Would **not** speak out	.17	.15	.21	--		
5.	Peers would speak out	-.20	-.07	.06	-.33	--	
6.	Worried about bullying	.21	.01	.20	.04	.06	--
7.	Worried about trusting old findings	.37	.35	.41	.08	-.10	.03

Note. N = 109

As to be expected, some factors of culture correlated positively with another, emerging in clusters. Reporting that QRPs were a problem correlated positively with QRPs affecting one's career (r = .54, n = 109) and with being worried about trusting old findings (r = .35, n = 109). The latter two were correlated with each other as well (r = .41, n = 109). Interestingly, the correlations were very low between worrying about being bullied and

seeing QRPs as a problem (r = .01, n = 109), not speaking out against senior researchers in their department (r = .04, n = 109), peers speaking out (r = .06, n = 109), and being worried about trusting old findings (r = .03, n = 109). The mean agreement scores as well as additional descriptive statistics for all these variables of culture can be found in Appendix C.

Preregistration

With only seven participants indicating otherwise, the clear majority (93.58%) of respondents knew what preregistration was. However, when asked about the number of projects they had preregistered before, most participants answered that they had never done so (68.81%) and the overall average of preregistrations was thus very low (M = 0.58, SD = 1.17). Figure 3 shows the distribution of answers counts of participants.

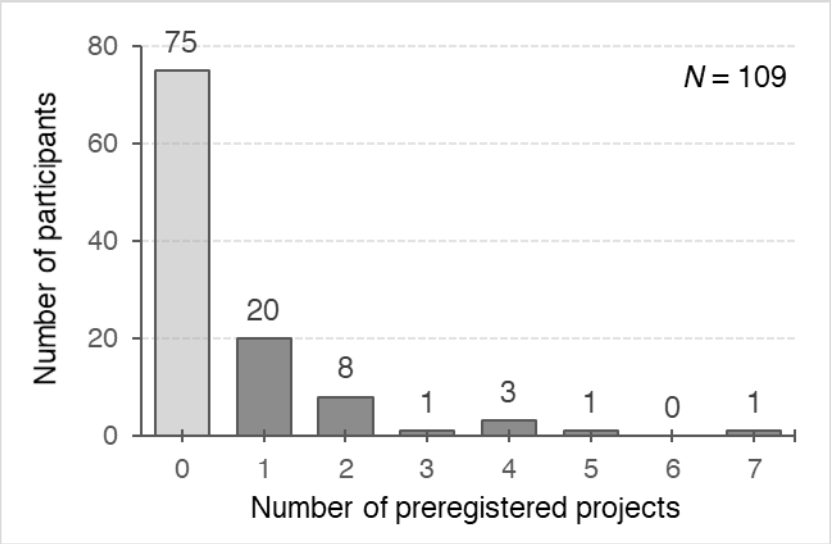


Figure 3. Number of projects that participants had preregistered.

Opinions About the Replication Crisis

In relation to the question whether the replication crisis, the changes in the field, or the tone of discussions had the respondents consider leaving academia, the answers were quite clear. Only a handful of respondents indicated having considered leaving academia for fear of "replication bullies" (3.67%). On the other hand, almost a quarter of respondents indicated having considered leaving because of being disillusioned by the problems in the

field (23.85%). Only two participants had considered leaving for both reasons. Compared to respondents working under a lead researcher, fewer of those who were lead researchers themselves considered leaving because of "replication bullies" (1.72%) and more considered leaving because of being disillusioned (27.59%). Table 6 provides an overview of the count and frequency information for the subgroups. Out of all respondents, less than half agreed with the statement that psychological research had changed considerably since the beginning of the replication crisis (41.28%). This percentage was only slightly higher for respondents working under a lead researcher (43.14%).

Table 6

Responses to items regarding the replication crisis and culture, subgroups for those who were lead researchers themselves and those who work under a lead researcher

Item	Total %		Nature of PhD project			
			Lead researchers themselves		Under a lead researcher	
	No	Yes	No	Yes	No	Yes
I considered leaving academia for fear of "replication bullies" (i.e. fierce criticism of research methods).	96.33%	3.67%	98.28%	1.72%	94.12%	5.88%
	n = 105	*n* = 4	*n* = 57	*n* = 1	*n* = 48	*n* = 3
I considered leaving academia because I got disillusioned by the problems in the field revealed by the replication crisis.	76.15%	23.85%	72.41%	27.59%	80.39%	19.61%
	n = 83	*n* = 26	*n* = 42	*n* = 16	*n* = 41	*n* = 10
Psychological research has changed considerably since the beginning of the replication crisis.	58.72%	41.28%	60.34%	39.66%	56.86%	43.14%
	n = 64	*n* = 45	*n* = 35	*n* = 23	*n* = 29	*n* = 22

Note. *N* = 109 due to two participants dropping out.

Predicting thoughts of leaving academia. A hierarchical logistic regression analysis was performed to explore if encountering QRPs and the ability of speaking out against them related to whether participants considered leaving academia due to feeling disillusioned about the field. First, a binary logistic regression model that only included the total QRPs as a

predictor was computed. The intercept of the model was b_0 = -2.47, 95% CI [-3.49, -1.60], and the unstandardized coefficient of total_QRP was b_1 = 0.35, 95% CI [0.17, 0.56]. This first model had a smaller Akaike information criterion (AIC = 109.3, *df* = 2) compared to a null model that only included the intercept (b_0 = -1.16, 95% CI [-1.62, -0.74]) with AIC = 121.8 (*df* = 1). This suggests that the first model was a better model to describe the data than a null model. To see if variables of speaking out would add predictive value after controlling for the number of encountered QRPs, a full model was created that included the two additional predictors concerning speaking out. One predictor was the probability of the participants remaining silent if they would see a senior researcher engaging in QRPs (speakout_self_rc). The other one was the perceived ability of PhD students in participants' departments to speak up against senior researchers engaging in QRPs (speakout_other). The coefficients of the full model were b_0 = -2.97, 95% CI [-4.68, -1.50] for the intercept; b_1 = 0.36, 95% CI [0.17, 0.57] for total_QRP; b_2 = 0.003, 95% CI [-0.01, 0.02] for speakout_self_rc; and b_3 = 0.007, 95% CI [-0.01, 0.02] for speakout_peers. However, the full model including all three predictors had a larger Akaike information criterion (AIC = 112.6, *df* = 4) than the first model (AIC = 109.3, *df* = 2), suggesting that it was not a better model to describe the data. This exploratory result indicates that the total QRP score could be a meaningful predictor for the thought of leaving academia, while the questions about speaking out that were used in this study are less likely to provide additional unique explanatory value if we already consider total QRP score. As mentioned above, the unstandardized regression coefficient of total_QRP in the first model was *b* = 0.35, 95% CI [0.17, 0.56]. In other words, with each observed QRP, the odds of a respondent considering leaving increased by a factor of 1.42. This means that a person who encountered five QRPs is predicted by this model to be seven times more likely to having considered leaving, compared to somebody who did not encounter any QRPs. However, since no criteria for model selection were preregistered,

definite model selection is not possible, and these results should merely be seen as information for planning future research.

Discussion

The results of this exploratory study of Swedish and Dutch PhD students provide novel insights to their experience of psychological research after the beginning of the replication crisis. The main finding about prevalence of QRPs was that the majority had encountered at least some of them in the last year, adding to the existing evidence that QRPs are pervasive enough to warrant action (Agnoli et al., 2017; Fiedler & Schwarz, 2016; John et al., 2012). The model including only the total amount of encountered QRPs as a predictor was indicative to be the most efficient in predicting whether respondents considered leaving academia because of being disillusioned, which almost a quarter of them had thought about. Interestingly, only a handful of respondents considered leaving due to fear of fierce criticism of methods. While almost everyone knew what preregistration was, only a minority preregistered a project before. This suggests that the currently low number of preregistrations in this sample was not due to a lack of knowledge thereof. Though the findings were exploratory, and no definite inferences should be made based on them, they address important issues and can guide future research. Practical and theoretical implications, as well as limitations, will be discussed below.

On the Prevalence of Questionable Research Practices

Because of the difference in measures and phrasing of the items, a direct comparison with past research is not possible. In addition to these differences, six years had passed since the publication of John and colleagues' (2012) original study. In this time, there have been a number of changes in research practices (Motyl et al., 2017) and most of the PhD students in this study started their career after the 2012 study. However, general observations can be made in relation to the previous studies. For some of the similar items, the prevalence rates

seem to fall in between those of John and colleagues (2012) and those of Fiedler and Schwarz (2016). This indicates that the patterns of encountering QRPs could be comparable among Swedish and Netherland PhD students to the populations used in these two other studies, so hypotheses derived from these previous studies could be tested in the population used in this study as well in the future. It also is in line with Agnoli and colleagues' (2017) conclusion that QRPs might be an international phenomenon.

The overall prevalence found in the present study was lower than in the original study by John and colleagues (2012), but the two QRP recency measures concerning the past year provided additional information. A majority in the present sample had encountered at least some QRPs in the past year, which suggests that future studies are still likely to find a considerable prevalence of QRPs among ECRs in Sweden and the Netherlands. It is worth pointing out that these are the first findings of this kind concerning Swedish and Dutch psychology departments from the perspective of ECRs. Since engaging in QRPs can so easily produce false-positives (Simmons et al., 2011), these results seem to warrant further investigation. When interpreting the prevalence reported in the present study, one should also keep in mind that throughout the entire project, measures were taken to keep the estimates as conservative as possible. The aim of reducing social desirability was addressed both while contacting participants and in the survey itself. The invitation email was phrased neutrally and the explanations about QRPs were phrased in a neutral and deliberately vague fashion to not influence responses in one or the other way.

Types of Questionable Research Practices

It is also interesting which specific QRPs were common and which ones were rarely and never encountered. Three of the four most common QRPs (each encountered by at least 40% of participants) were behaviors that would give readers an incorrect impression about the strength and relevance of findings, likely skewing the overall body of research.

Additionally, several items related to p-hacking had been encountered by more than a quarter of the participants, and the item related to HARKing (hypothesizing after knowing the results) was encountered by even more participants (40%). All these common QRPs have been described as likely being closely connected to false-positives in published research (Forstmeier, Wagenmakers, & Parker, 2017; Nelson et al., 2018). On the other side, the two least encountered QRPs were both related to conflicts of interest. None of the participants indicated having ever encountered a failure to report conflicts of interest, and only two participants had encountered changing aspects of a study to please a sponsor. Thus, the respondents to this survey did not seem to see conflicts of interest as a common occurrence. The exclusion of these items could be considered in future studies in the same population.

Leaving Academia

Almost a quarter of the PhD students who participated had considered leaving academia due to being disillusioned by the problems in the field. Additionally, an exploratory analysis found that the number of encountered QRPs had substantive predictive value for having considered to leave for this reason. Based on this exploratory finding, one could speculate that working in an environment where QRPs are more common could have a negative effect on how academia is seen. This hypothesis could be assessed with future studies. How likely the respondents themselves were to remain quiet when observing QRPs of senior researchers did not seem to have an important predictive value for being disillusioned, and neither did the likelihood of peers speaking up.

Interestingly, only a handful of respondents considered leaving because of fear of fierce criticism of methods. This is the first empirical assessment of the claim that rigorous measures against bad practices discourage young researchers (Fiske, 2017). The fear that a hostile tone or fierce methodological criticism could drive young researchers to leave

promising careers seemed not to be warranted for this sample, but confirmatory studies would be needed to make claims beyond the present sample.

Preregistration

Almost every participant knew what preregistration was. However, only a minority actually preregistered a project before, which hints at a gap between knowledge and action. Nosek and colleagues (2018) suggested that increasing the number of preregistrations does not have to be difficult. They point out that many grant applications already demand prespecifying data collection and analysis to an extent similar to preregistration. Research following up on the discrepancy between knowledge and number of preregistrations could investigate if PhD students are aware of this. Some researchers argue that the gap between knowledge and action related to preregistration could be addressed by policy changes and by providing more incentives for preregistration (Nosek et al., 2018). As Simmons, Nelson, and Simonsohn, (2018) point out in a recent article published at the time of data collection for the present study, "none of the top American Psychological Association journals have implemented disclosure requirements" (p.256). This lack of appropriate policies about preregistration might also explain why so few PhD students preregistered projects before. Recent findings by Washburn and colleagues (2018) connect to this. They, too, found that only a minority of researchers in their sample had preregistered projects before. However, Washburn and colleagues (2018) additionally reported that a majority of researchers thought it was acceptable to not preregister, or that they were not sure if it was not. This is a further indication that there is still a need for stricter policies in support of preregistration. If the publication bias for positive results and the problems with false-positives and failed replications are to be overcome, preregistration should be an important tool for the scientific community in achieving these goals. In medical research, preregistration has already been found to correlate with increased publication of null-findings (Kaplan & Irvin, 2015).

Limitations of the Present Study

Due to the broad scope of this exploratory project, there are limitations to the results of the study. In terms of the list of encountered QRPs, a limitation was that the measure asked about encountering them in senior researchers, colleagues, or oneself, without making a difference between these different possibilities. The decision to not include a differentiation between who engaged in QRPs was made consciously to prevent participants feeling judged and to reduce social desirability bias. For the purposes of exploring the overall frequency of QRPs, it does not matter who engages in QRPs, it matters that someone engages in them. In practice, however, it could be useful to have the distinction to link prevalence rates with the items of culture. For example, PhD students who engage in QRPs themselves are likely to give systematically different replies when it comes to speaking out against them. A solution to this in future research could be to at least distinguish between senior researchers on the one side, and respondents and their PhD student peers on the other side. That way, admitting to engaging in QRPs would be possible indirectly, and it would prevent losing information about potential differences between generations of researchers.

Another limitation of this study is the broadness of some of the statements. Most cultural factors were assessed by relying on a single item rather than measures containing several items, thus putting more emphasis on phrasing and wording rather than the underlying factors. This introduces additional variance to the replies of respondents, since they might relate their answer to different aspects of the item. This limitation was considered in the planning phase of the study, but the single items were deemed more useful for the goal of making the survey concise and participation easy, thus achieving a higher response rate. Nevertheless, follow-up studies could address this limitation by creating measures with several items for each factor.

Lastly, the limited information that was collected about participants could be considered a limitation by some. To ensure absolute anonymity and guarantee this to the participants both explicitly and implicitly, almost no demographic data was collected. Gender did not affect scientific integrity in a past study (Fanelli, Costas, & Larivière, 2015), and there was no evidence suggesting that age should matter. Nevertheless, collecting information like which area of psychology the participants work in could have been useful to identify potential differences. Much of the research related to the replication crisis focuses on social and cognitive psychology, with less attention to challenges and practices in other subfields (Hamlin, 2017; Tackett et al., 2017). The same applies to information about the country that PhD students were from. No data were collected to guarantee anonymity, so it is possible that there were very different response patterns in Sweden and the Netherlands, which couldn't be identified. Anonymity should still be prioritized, but similar studies in the future should build on the general findings from this study and check if QRPs differ between subfields and between countries or investigate populations separately in the first place. Finally, because of the exploratory nature of the study, the conclusions are limited to advising future research and conclusions cannot be extended to the target population.

Conclusion

In summary, this study provided several novel insights into the opinions of PhD students from top universities in both Sweden and the Netherlands. Due to the exploratory nature of the study, first steps into investigating a broad range of issues were made. Most PhD students have encountered QRPs, and most did so in the past year, suggesting that they are a relevant issue. This study was also the first to report actual data about PhD students' reactions to the methodological crisis. As opposed to what some in the field suggested based on anecdotal evidence (Fiske, 2017), the majority of early-career researchers in this sample did not fear bullying. On the other hand, almost a quarter of respondents having considered

leaving academia due to being disillusioned by the problems revealed by the replication crisis. The results of this study can serve as a useful guide to future studies in this area, moving closer to the goal of increasing the credibility of psychological research.

References

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS ONE*, *12*(3), 1–17. https://doi.org/10.1371/journal.pone.0172792

American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Anderson, M. S., Horn, A. S., Risbey, K. R., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). 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*(9), 853–860. https://doi.org/10.1097/ACM.0b013e31812f764c

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., … Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, *116*(1), 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE*, *7*(1). https://doi.org/10.1371/journal.pone.0029081

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, *4*(5). https://doi.org/10.1371/journal.pone.0005738

Fanelli, D. (2013). Redefine misconduct as distorted reporting. *Nature*, *494*(7436), 149.

https://doi.org/10.1038/494149a

Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. *PLoS ONE*, *10*(6), 1–18. https://doi.org/10.1371/journal.pone.0127556

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Fiske, S. T. (2017). Going in Many Right Directions, All at Once. *Perspectives on Psychological Science*, *12*(4), 652–655. https://doi.org/10.1177/1745691617706506

Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews*, *92*(4), 1941–1968. https://doi.org/10.1111/brv.12315

Franco, A., Malhotra, N., & Simonovits, G. (2016). Underreporting in Psychology Experiments: Evidence From a Study Registry. *Social Psychological and Personality Science*, *7*(1), 8–12. https://doi.org/10.1177/1948550615598377

Gelman, A. (2014). How much time (if any) should we spend criticizing research that's fraudulent, crappy, or just plain pointless? Retrieved April 1, 2018, from http://andrewgelman.com/2014/03/06/much-time-spend-criticizing-research-thats-fraudulent-crappy-just-plain-pointless/

Gelman, A., & Loken, E. (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Psychological Bulletin*, *140*(5), 1272–

1280. https://doi.org/dx.doi.org/10.1037/a0037714

Giner-Sorolla, R. (2016). APA Ethics vs. the File Drawer. Retrieved April 17, 2018, from https://approachingblog.wordpress.com/2016/06/21/apa-ethics-vs-the-file-drawer/

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science*, *11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The Extent and Consequences of P-Hacking in Science. *PLoS Biology*, *13*(3), 1–16. https://doi.org/10.1371/journal.pbio.1002106

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS ONE*, *10*(8), 1–13. https://doi.org/10.1371/journal.pone.0132382

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203

LeBel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure.org: Grassroots Support for Reforming Reporting

Standards in Psychology. *Perspectives on Psychological Science, 8*(4), 424–432. https://doi.org/10.1177/1745691613491437

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research: How Often Do They Really Occur? *Perspectives on Psychological Science*, *7*(6), 537–542. https://doi.org/10.1177/1745691612460688

Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Allison, B., … Skitka, L. J. (2017). Journal of Personality and Social Psychology The State of Social and Personality Science : Rotten to the Core , Not So Bad , Getting Better , or Getting Worse ? The State of Social and Personality Science : Rotten to the Core , Not So Bad , Getting Better, *113*(1), 34–58.

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annual Review of Psychology*, *69*(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7*(6), 657–660. https://doi.org/10.1177/1745691612462588

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716-aac4716. https://doi.org/10.1126/science.aac4716

Pashler, H., & Wagenmakers, E. J. (2012). Editors' Introduction to the Special Section on

Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. A. (2015). Assessing the Robustness of Power Posing: No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women. *Psychological Science*, *26*(5), 653–656. https://doi.org/10.1177/0956797614553946

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavior Research Methods*, *48*(3), 1062–1069. https://doi.org/10.3758/s13428-015-0630-z

Sacco, D. F., Bruton, S. V., & Brown, M. (2017). In Defense of the Questionable: Defining the Basis of Research Scientists' Engagement in Questionable Research Practices. *Journal of Empirical Research on Human Research Ethics*, *13*(1), 101–110. https://doi.org/10.1177/1556264617743834

Scargle, J. D. (2000). Publication Bias : The "File-Drawer" Problem in Scientific Inference. *Journal of Scientific Exploration*, *14*(1), 91–106.

Schimmack, U. (2014). The Replicability-Index (R-Index): Quantifying Research Integrity. Retrieved April 1, 2018, from https://replicationindex.wordpress.com/2014/12/01/quantifying-statistical-research-integrity-r-index/

Sijtsma, K. (2016). Playing with Data—Or How to Discourage Questionable Research Practices and Stimulate Researchers to Do Things Right. *Psychometrika*, *81*(1), 1–15. https://doi.org/10.1007/s11336-015-9446-0

Sijtsma, K., Veldkamp, C. L. S., & Wicherts, J. M. (2016). Improving the Conduct and Reporting of Statistical Analysis in Psychology. *Psychometrika*, *81*(1), 33–38.

https://doi.org/10.1007/s11336-015-9444-2

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-Positive Citations. *Perspectives on Psychological Science*, *13*(2), 255–259. https://doi.org/10.1177/1745691617698146

Simons, D. J., Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Birch, S., … Zwaan, R. A. (2014). Registered Replication Report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*(5), 556–578. https://doi.org/10.1177/1745691614545653

Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics*, *12*(1), 53–74.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*, 30–34.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science*, *7*(6), 670–688. https://doi.org/10.1177/1745691612460687

Stürmer, S., Oeberst, A., Trötschel, R., & Decker, O. (2017). Early-Career Researchers' Perceptions of the Prevalence of Questionable Research Practices, Potential Causes, and Open Science. *Social Psychology*, *48*(6), 365–371. https://doi.org/10.1027/1864-9335/a000324

Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., …

Zwaan, R. A. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, *11*(6), 917–928. https://doi.org/10.1177/1745691616674458

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., … Carsel, T. S. (2018). Why Do Some Psychology Researchers Resist Adopting Proposed Reforms to Research Practices? A Description of Researchers' Rationales. *Advances in Methods and Practices in Psychological Science*, 251524591875742. https://doi.org/10.1177/2515245918757427

Yong, E. (2012). Replication studies: Bad copy. *Nature*, *485*, 298–300. https://doi.org/10.1038/485298a

Appendix A

Informed consent form from the online survey.

Hi there!

I really appreciate your time and promise that it won't take long. During pilot testing, completing the survey never took longer than **10 minutes** and I will do my best to make it worthwhile! I kindly ask for your full attention when reading the instructions and the items in order to get meaningful data.

I am conducting this exploratory study to better understand the culture of the psychological research community and its implications for early-career researchers like you. I will use the collected data for my master thesis, supervised by Zoltan Kekecs at Lund University in Sweden. After the data collection, the fully anonymous dataset will be uploaded to the Open Science Framework.

I would like to stress that all your responses will be **completely anonymous**, and I will treat any potential communication with **highest confidentiality**. There will be no saved dataset that could allow for identifying you or which university you are from!

There are no known risks to participating in this study and you are free to refuse to answer and end your participation at any time.

If you have any questions or comments now or after your participation, please don't hesitate to contact me via email. Otherwise, you will get the chance to leave a comment at the end of the survey.

Kind regards,

Nils Arlinghaus

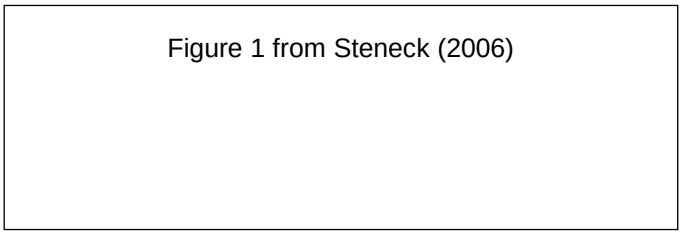
Appendix B

Explanation of questionable research practices as provided to the participants.

Questionable Research practices explanation

The term “questionable research practice” has received some growing attention in the last years. It describes research behaviors that fall in between what has been called *responsible conduct of research* (RCR) and *fabrication, falsification, and plagiarism* (FFP). In other words, questionable research practices (QRPs) are in the grey area somewhere between ideal practice and outright fraud.

Researchers don’t always engage in questionable research practices on purpose, and some say that they are not serious enough to warrant regulatory action (Steneck, 2006).



(Figure from Steneck, N. H. (2006): *Fostering integrity in research: Definitions, current knowledge, and future directions.*)

Appendix C

Table 1

List of Questionable Research Practices

Item	Original item (if modified)	Source
"Rounding off" a p value (e.g., reporting that a p value of .**044** is less than .**04**)	In a paper, "rounding off" a p value (e.g., reporting that a p value of .054 is less than .05)	John et al. (2012)
(Reporting an unexpected finding as having been predicted from the start.) [a]		John et al. (2012)
Claiming that results are unaffected by demographic variables (e.g., gender) although one is actually unsure (or knows that they do).		Fiedler & Schwarz (2016)
Collecting more data after seeing whether results were significant in order to render non-significant results significant.		Fiedler & Schwarz (2016)
Stopping data collection after achieving the desired result concerning a specific finding.		Fiedler & Schwarz (2016)
Deciding whether to exclude data after looking at the impact of doing so regarding a specific finding		Fiedler & Schwarz (2016)
Changing the design, methodology, or results of a study to please a sponsor.		Sacco et al. (2017)
Publishing results of a single study as several articles, to increase the number of publications derived from the research.	Publishing results of a single study as several articles simply to increase the number of publications derived from the research (the so-called "salami slicing" problem).	Sacco et al. (2017)
Not reporting **all of a study's conditions**, subgroups, outcomes, **or** time points.	Selective reporting of subgroups, outcomes, and time points.	Sacco et al. (2017)
Selectively discussing only studies that supported the hypothesized result(s).		Sacco et al. (2017)
Reporting an unexpected result as having been hypothesized from the start.		Sacco et al. (2017)
Drawing strong inferences from **underpowered, but statistically significant results.**	Drawing strong inferences from statistically significant but underpowered results.	Sacco et al. (2017)
Not reporting some potentially relevant conflicts of interest.	Failing to disclose all potentially relevant conflicts of interest.	Sacco et al. (2017)
Not sharing data or materials with other researchers to prevent questions about the quality of your work from being raised.	Refusing to share data or materials with other researchers to prevent questions about the quality of your work from being raised.	Sacco et al. (2017)
Overlooking or **not paying attention to** others' questionable research practices.	Overlooking or ignoring others' questionable research practices.	Sacco et al. (2017)

[a] Removed during the analysis due to being a duplicate

Table 2

General agreement questions about impact of QRPs and replication crisis.

Item
I think that questionable research practices are a problem that affects published research to a large extend
I think that questionable research practices affect my academic life.
If I ever saw a senior researcher in my department engaging in QRPs, I probably wouldn't mention it
If individual PhD students in my department saw senior researchers engaging in QRPs, they wouldn't have to hesitate to raise their concerns
I have been worried about being attacked by other researchers for bad methodology
I have been worried about whether I can trust older research findings (because of the possibilities of QRPs)

Appendix D

Descriptive statistics for agreement [0-100] with statements about culture

Item	*M*	*SD*	Median	Skewness	Kurtosis
QRPs are a problem	66.65	25.84	71	-0.60	-0.57
QRPs affect own career	39.80	31.22	37	0.41	-1.13
Would **not** speak out	40.62	28.13	46	0.11	-1.23
Peers would speak out	42.33	30.89	42	0.20	-1.27
Worried about bullying	25.03	27.49	14	1.17	0.29
Worried about trusting old findings	49.62	30.18	52	-0.02	-1.12

Note. $N = 109$