ORIGINAL PAPER

Crossing the Red Line? Empirical Evidence and Useful Recommendations on Questionable Research Practices among Business Scholars

Hengky Latan[1] · Charbel Jose Chiappetta Jabbour[2,4] · Ana Beatriz Lopes de Sousa Jabbour[3,4] · Murad Ali[5]

Abstract
Academic leaders in management from all over the world—including recent calls by the Academy of Management Shaw (Academy of Management Journal 60(3): 819–822, 2017)—have urged further research into the extent and use of questionable research practices (QRPs). In order to provide empirical evidence on the topic of QRPs, this work presents two linked studies. Study 1 determines the level of use of QRPs based on self-admission rates and estimated prevalence among business scholars in Indonesia. It was determined that if the level of QRP use identified in Study 1 was quite high, Study 2 would be conducted to follow-up on this result, and this was indeed the case. Study 2 examines the factors that encourage and discourage the use of QRPs in the sample analyzed. The main research findings are as follows: (a) in Study 1, we found the self-admission rates and estimated prevalence of business scholars' involvement in QRPs to be quite high when compared with studies conducted in other countries and (b) in Study 2, we found pressure for publication from universities, fear of rejection of manuscripts, meeting the expectations of reviewers, and available rewards to be the main reasons for the use of QRPs in Indonesia, whereas (c) formal sanctions and prevention efforts are factors that discourage QRPs. Recommendations for stakeholders (in this case, reviewers, editors, funders, supervisors, chancellors and others) are also provided in order to reduce the use of QRPs.

Keywords Questionable research practices · Ethics in research · Estimated prevalence · Justifications · Formal sanctions · Prevention efforts · Perceived behavioral control

Introduction

Recently, business schools and prestigious journals all over the world have faced increasing ethical challenges related to issues surrounding potential questionable research practices (QRPs) and the 'red lines' of acceptable research ethics. These issues are related to the credibility of findings, the replicability crisis and research malpractice, such as p-hacking or hypothesizing after results are known (HARKing), have been widely witnessed and debated among scholars in recent years (Anderson et al., 2019; John et al., 2012; Linder & Farahbakhsh, 2020; O'Boyle et al., 2017; Rubin, 2017; Simmons et al., 2011; Vogel & Homberg, 2021). However, the facts indicate that scholars' involvement in these practices has grown exponentially in the last decade, with dozens of published papers being retracted due to various forms of research malpractice (Cox et al., 2018; Steen et al., 2013; Stricker & Günther, 2019). We argue that such practices have the potential to cause harmful consequences for

✉ Charbel Jose Chiappetta Jabbour
chiappettajabbour@em-lyon.com

Hengky Latan
hlatan@hlcacademicpublishing.com

Ana Beatriz Lopes de Sousa Jabbour
bsousajabbour@em-normandie.fr

Murad Ali
murad2.ali@northumbria.ac.uk

1 HLC Consulting, Jl. Kertanegara Selatan V No 5B, Semarang 50241, Indonesia

2 EMLYON Business School, 23 Av. Guy de Collongue, 69130 Écully, France

3 EM Normandie Business School, Metis Lab, 64 rue du Ranelagh, 75016 Paris, France

4 Lincoln International Business School, University of Lincoln, Brayford Pool, Lincoln, Lincolnshire LN6 7TS, UK

5 Newcastle Business School, Northumbria University, City Campus East, NE1 8ST, Newcastle Upon Tyne, UK

science by presenting the results of post hoc analysis within the realm of quantitative research, which violates ethical codes of research and the logic of hypothetical-deductive inquiry (Behfar & Okhuysen, 2018; Hollenbeck & Wright, 2017). To date, little is known about why certain researchers choose to engage in QRPs and what factors can either encourage or discourage this unethical behavior in relation to research ethics.

Theodore D. Sterling, an American statistician, conducted the first research into these practices several decades ago (Sterling, 1959) as well as the later impactful follow-up article by Kerr (1998), there has been little evidence gathered to substantiate the existence of these practices among scholars in business research. As highlighted by Leung (2011), there is a relative scarcity of literature examining the prevalence with which leading researchers engage in these practices. However, it should be recognized that these practices exist, and this cannot be ignored (Banks et al., 2016; Shaw, 2017). In addition, the self-admission rate and estimated prevalence of business scholars around the world engaging in QRPs is still unclear (Fiedler & Schwarz, 2016; John et al., 2012). To the best of our knowledge, there are no existing studies addressing this persistent gap in this context, with the self-admission rate and estimated prevalence remaining unknown among scholars in business research.

Furthermore, as noted in the existing literature, there have been research calls to conduct empirical investigations regarding the factors that affect these potential QRPs (Rajah-Kanagasabai & Roberts, 2015; Sijtsma, 2016; Wigboldus & Dotsch, 2016). As far as we are aware, there is a lack of empirical studies providing concrete evidence regarding QRPs in disciplines beyond the business management domain (Krishna & Peter, 2018).

Accordingly, we herein consider the factors that may both encourage and discourage individuals from engaging in QRPs. We examine the effects of justifications (JST), formal sanctions (FRS), prevention efforts (PRE) and perceived behavioral control (PBC) as predictors of QRPs. Following the theory of planned behavior, a person may engage in such behaviors in the presence of adequate justification (JST) and greater perceived behavioral control (PBC) (Ajzen, 2005; Latan et al., 2018; Rajah-Kanagasabai & Roberts, 2015). Consequently, based on deterrence theory, there is a possibility that prevention efforts (PRE) and formal sanctions (FRS) may have implications for reducing unacceptable research practices (Pratt et al., 2019). In this case, QRPs may occur because the sanctions applied are too lenient, because the likelihood of being caught is too low or because of a lack of ethical awareness among researchers (Hall & Martin, 2019; Sijtsma, 2016).

Hence, our current research has two main research objectives:

Objective 1: To provide preliminary evidence regarding the self-admission rate and estimated prevalence of potential QRP use among business scholars in Indonesia (Study 1).

Objective 2: To provide the first empirical evidence regarding the factors that encourage and discourage QRPs (Study 2).

To achieve our research objectives, we conducted two separate studies but sequential using a sample of business scholars affiliated with the 250 top universities in Indonesia. It was necessary to conduct Study 1 before Study 2, because Study 1 examines the level of QRP use among the sample, while Study 2 explores the factors that encourage and discourage QRP use among the same sample. In Study 1, we used the two methods proposed by John et al. (2012) and Fiedler and Schwarz (2016) to calculate two measures: self-admission rate and estimated prevalence. As Fiedler and Schwarz (2016) argue, the method proposed by John et al. (2012) does have several limitations. However, we still present both approaches in Study 1, in order to demonstrate their differences. In Study 2, we conducted empirical research based on the indicative findings of Study 1 to examine the factors that encourage (i.e., JST and PBC) and discourage (i.e., FRS and PRE) QRPs, using the same target respondents.

Our study offers several expected and unexpected contributions to theory and practice for business ethics research. First, based on our best knowledge, this is the first preliminary study addressing QRPs among business scholars by uncovering self-admission rate and estimated prevalence in relation to research misconduct. Therefore, our study extends the state-of-the-art research on QRPs into the realm of quantitative research (Harvey, 2017; Vogel & Homberg, 2021). In addition, this study reconciles debates regarding the methods proposed by John et al. (2012) and Fiedler and Schwarz (2016) by comparing the two approaches using our original sample. Second, our research is the first to investigate the factors that encourage and discourage QRPs among business scholars. As Sijtsma (2016) argues, it is more important to know which factors can hinder these unethical research practices than to identify their use. Finally, our study provides the first empirical evidence in the Indonesian context in relation to unacceptable research practices (Rabelo et al., 2020; Wigboldus & Dotsch, 2016). Although several studies on this topic have been carried out in various countries in the world – for example, the US., Italy, Germany and Brazil – no existing studies have reported on such research practices in the Asia–Pacific region, especially in the Indonesian context.

The article is structured as follows: theoretical background and research hypotheses; description of Study 1, including data collection and analysis and its results; description of Study 2, including data collection and analysis

and its results; discussion of findings of both studies; theoretical and practical implications; and finally, some recommendations to reduce the use of QRPs.

Types of Questionable Research Practice

Kerr (1998), in his seminal article on this subject, describes QRPs – which he calls hypothesizing after the results are known (HARKing) – including types, reasons for engaging in such unethical practices, and the extent of application of such practices. In short, QRPs issues then began to be discussed among scholars in various disciplines, with the field of business ethics research being no exception.

We define QRPs as systematic anticipatory approaches utilized by researchers in order to meet the expectations of stakeholders (in this case, reviewers, editors, funders, supervisors, chancellors and others) by altering their initial hypotheses, research models, study design, the results of initial analysis or the original data set, in order to present and report the evidence in a more desirable way. Another definition follows Banks et al., (2016, p. 7), wherein QRPs are defined as design, analytic, or reporting practices that may be questioned because of their potential to be employed with the purpose of presenting biased evidence in favor of a particular assertion. In other words, it refers to a retroactive process of presenting findings that are a priori and/or unexpected, by adding or dropping hypotheses and then reporting statistically significant findings (Hollenbeck & Wright, 2017).

There are several types of QRPs, and the terminology used varies between scholars. For example, Leung (2011) uses the term "presenting post hoc hypotheses as a priori" (PPHA); Wigboldus and Dotsch (2016) suggest labeling such behavior simply "unacceptable research practices"; O'Boyle et al. (2017) propose the term "chrysalis effect"; Rubin (2017) uses the phrases "constructing hypotheses after the results are known" (CHARKing), "retrieving hypotheses after the results are known" (RHARKing) and "suppressing hypotheses after the results are known" (SHARKing) to explain such practices, while Hollenbeck and Wright (2017) emphasize "transparent HARKing" (THARKing). In this paper, we use these terms interchangeably.

Following the systematic logic of hypothetico-deductive inquiry, quantitative research usually begins with a description of general theories or prior knowledge related to the phenomenon to be investigated, before developing a research model. After going through this process, a researcher will then derive the hypotheses to be tested and compile research questions. In the next stage, the study design is created, and data collection is carried out. Subsequently, the data collected are analyzed in order to produce empirical findings. However, researchers often do not follow this logic, and may invert the order of the processes, i.e., analyzing the data first

to check their initial results, and then altering their research questions and hypotheses. This is generally intended to avoid reporting negative results and prevent rejection of hypotheses (Kerr, 1998; Leung, 2011). If the results of the initial analysis do not fulfill their expectations, then one or more actions from among the potential QRPs may be taken.

We illustrate the use of this logic in business research as shown in Fig. 1. In this example, we categorize these QRPs into eight actions, while previous studies have used different categorizations; for example, Crede and Harms (2019) divide these practices into seven types, Banks et al. (2016) into six types, O'Boyle et al. (2017) into five types, Murphy and Aguinis (2019) into four types, and Kerr (1998) and Rubin (2017) into three types. The differences between these categorizations arise because some scholars combine several actions into one. Below, we elaborate the types of QRPs that often occur in business research and other fields. It is worth noting that this may not be an exhaustive list, since we are only considering QRPs used in quantitative research streams, which currently remains the main approach used in various journals.

Increasing sample size (n) or analyzing only a subset of the data are the first two potential actions. These actions are usually chosen when the initial data analysis does not support the hypotheses regarding the relationships between variables. When the *p*-value produced exceeds the significance threshold of the rule of thumb (e.g., 0.05 or 0.01), researchers may add cases to the data in order to increase the power of the analysis (Schimmack, 2012). Such addition allows for changes in the initial results from insignificant to significant, to support the research questions or hypotheses. However, this practice opposes the logic of null hypothesis significance testing (NHST), misusing it and inflating false-positive rates (type I error). Conversely, reducing or excluding data is another possible method for reaching the desired threshold. This is the easiest method, when the sample size available is large enough. In practice, researchers may delete extreme data (i.e., winsorizing) or delete missing values. This method is very effective at reducing the variability of data in order to render non-significant results significant. This practice is often called cherry-picking, and is used to find the strongest possible support for hypotheses or empirical findings from the dataset (Bloomfield et al., 2018; Harvey, 2017). Regrettably, this practice has been documented by previous studies to be one of the most common QRPs among scholars, as in the findings of John et al. (2012) and Agnoli et al. (2017). Other studies show how false-positive findings can occur in data collection and analysis (Simmons et al., 2011).

Dropping or adding hypotheses and changing the relationships in the model are the third and fourth likely actions. Sometimes, adding or excluding data does not guarantee that the desired threshold will be met. Therefore, researchers may
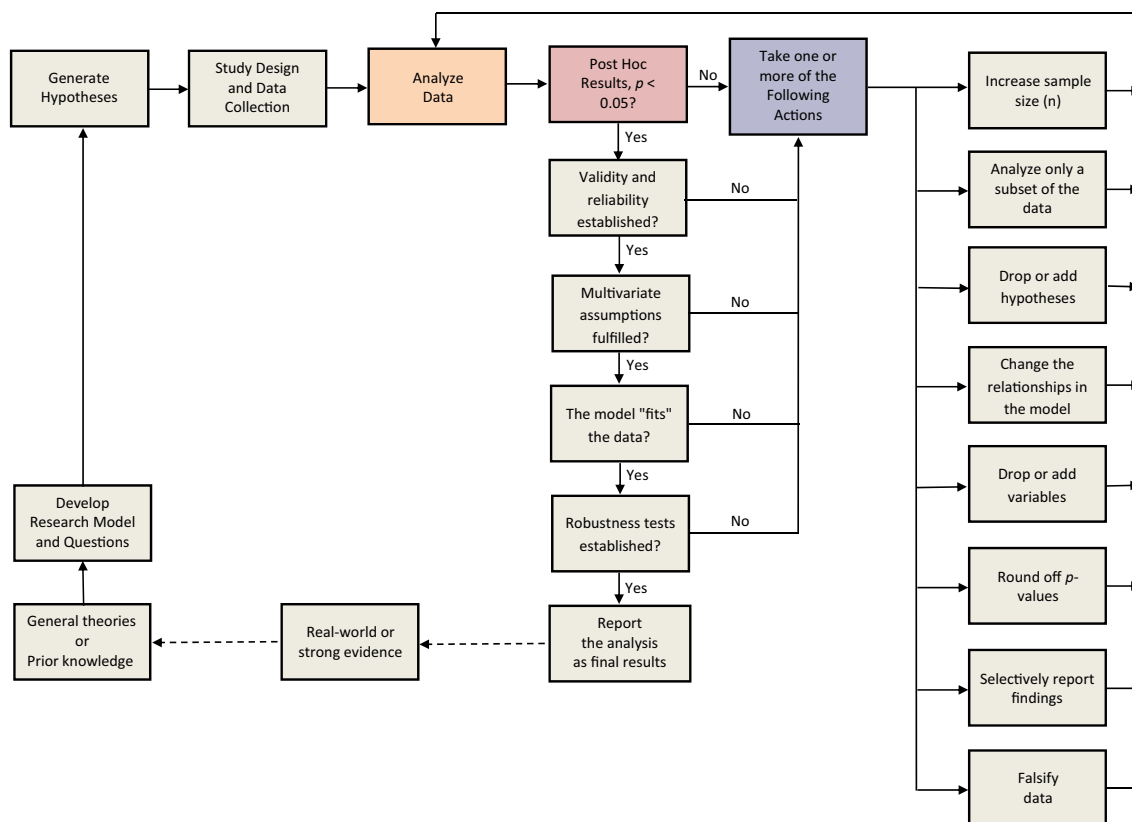
Fig. 1 Questionable research practices in business research

play with their hypotheses by adding, dropping or changing the direction of hypotheses to claim unexpected findings. This practice is often called question trolling or constructing hypotheses after the results are known (CHARKing) (Murphy & Aguinis, 2019; Rubin, 2017). Indeed, this practice does not adhere to theory-driven research, but it is based on data-driven practices and the search for causal relationships (e.g., vanishing tetrad). In its application, hypotheses that are not significantly supported will be excluded from the research questions and, instead, unexpected findings will be added. In some cases, the direction of the hypotheses found in the results is different from what was expected. To deal with such a result, researchers may change the relationships between variables in the model, for instance from independent to moderating or mediating, and vice versa. A study conducted by Murphy and Aguinis (2019) supports the existence of such practices. Other findings from Fiedler and Schwarz (2016) and Rabelo et al. (2020) indicate the existence of these practices among scholars.

Dropping or adding variables is the likely fifth action. When the above actions are not effective, researchers may take more extreme action by removing variables that show difficulty reaching a required significance level and/or adding new variables into the model. In other words, the research model is changed to help the proposed theoretical

model fit with the data. However, removing or adding variables from the model increases the effect of chance and is not in line with the principles of theory-driven research. Kline (2016) argues that changing a research model without adequate justification raises the level of bias in analysis and findings. Although the structural equation modeling approach supports this practice through modification indices, in general this is not recommended.

Rounding off p values is the potential sixth action. This is a common practice and is often carried out by researchers, as noted by Pritschet et al. (2016), Olsson-Collentine et al. (2019), Khan and Trønnes (2019) and Vogel and Homberg (2021). In this practice, researchers straightforwardly change the results of their analysis without having to go through a lengthy process. Usually, rounding off the p values occurs because the true results found are close to a required threshold. In addition, the results of such analysis may then be reported as "marginally significant". Another possible way for researchers to do this is by changing the significance level, i.e., from 0.05 to 0.10, or using a one-tailed rather than a two-tailed test. However, rounding off p values can produce bias by reporting findings which were initially not significant, leading to inaccurate inferences. Therefore, some scholars suggest requiring a good understanding of using p

values in quantitative research in order to prevent publication bias (Miller, 2017; Savalei & Dunn, 2015).

Selectively reporting findings is the likely seventh action. In his initial study on this topic, Sterling (1959) illustrates how the practice of selective reporting of findings (i.e., only reporting statistically significant results) leads to incorrect conclusions. A few decades later, a second study by Sterling et al. (1995) revisited these practices, and found no difference after 30 years. At this point, Sterling et al. (1995) noted the existence of a myth among scholars – that a good scientific result means that the hypotheses tested should be supported by empirical findings, and that insignificant results cannot be published. This principle has encouraged scholars to report only significant findings, leading them to become involved with *p* hacking and to reject negative results (Harvey, 2017; Khan & Trønnes, 2019). In addition, there are concerns that reviewers and editors do not like negative results because these results are considered to make no incremental contributions to theory and the body of knowledge (Hill et al., 2020). In fact, these negative results can lead to seeking justification in prior findings. This practice is often called retrieving hypotheses after the results are known (RHARKing) (Rubin, 2017). Several studies have indicated the involvement of scholars in such practices (Fanelli, 2010; Khan & Trønnes, 2019; O'Boyle et al., 2017).

Finally, falsifying data is the likely eighth action. This is the last and most extreme option that may be utilized by researchers when the previous methods fail to obtain the desired results. In this situation, data are created according to the results desired by researchers, either by changing, replacing or fabricating respondents' answers, as if these were the original data. In line with this characterization, Bedeian et al. (2010) define falsification as the manipulation of research data or processes or omission of critical data or results such that the research is not accurately represented in the research results. However, falsifying data is more difficult to conduct than the previous methods, because it requires skills and sophistication in calculating the scales, measurements and algorithms used in statistical analysis. However, this is not an impossible method to be performed, and a number of scholars have been involved in this practice (Bergh et al., 2017; Stricker & Günther, 2019).

Factors that Encourage and Discourage Individuals from Engaging in QRPs

Justifications for Employing QRPs

The theory of planned behavior (Ajzen, 2005) explains that a person's behavior will depend on several traits that can act as an impetus for action. Following this logic, someone who engages in unethical behavior usually has a strong sense of justification (JST) before acting. As illustrated by Schwartz

(2016), JST plays an important role in a person's decision to act. We define JST as the process of internal rationalization that underlies a scholar's decision to engage in QRPs (Latan et al., 2021). This is a cognitive process used to justify an action. As Fiedler and Schwarz (2016) argue, JST is an important factor in assessing why some scholars choose to engage in these research practices, while others do not. A report by Necker (2014) indicates a positive relationship between justifiability and prevalence of engaging in unacceptable research practices and misconduct. In addition, previous studies conducted by Stone et al. (2009) identify a positive effect of JST on student academic misconduct. Another study from Rajah-Kanagasabai and Roberts (2015) finds a positive and significant correlation between JST and research misconduct among university students. Based on the above discussion, our first hypothesis is:

H1 Justification has a positive effect on the prevalence of QRPs.

Formal Sanctions for Employing QRPs

Several scholars have asserted that one way to reduce potential QRPs is to impose strict sanctions on perpetrators (Butler et al., 2017; Hall & Martin, 2019). Such sanctions can be divided into formal sanctions, which are supported by legal and professional bodies, and informal sanctions, which occur among peers. In this paper, we consider formal sanctions (FRS) as a proxy for all sanctions, because they usually present clear consequences and penalties for unethical behavior. In addition, most institutions tend to have FRS in place for their faculty members if they are discovered to be violating the research code of ethics. Meanwhile, informal sanctions usually depend on the opinions of the research community and colleagues about whether or not an action violates ethical research norms. However, informal sanctions often have mild consequences; for example, peers refusing to cite the presumed wrongdoer's works (Pratt et al., 2019). Conversely, FRSs have a high deterrent effect, which has a negative impact on the personal and professional life of the wrongdoers. The range of possible FRSs includes revocation and refunding of research funding, being blacklisted from journals, risk of being fired by the institution or university, and perhaps even criminal proceedings. Several previous studies indicate that FRSs have a negative effect on unacceptable research practices (Honig et al., 2014; Pratt et al., 2019). Based on the above discussion, our second hypothesis is:

H2 Formal sanctions have a negative effect on the prevalence of QRPs.

Prevention Efforts and QRPs

In recent years, preventative efforts to combat QRPs have been added into the author guidelines of several journals, including making data and research material publicly available, submitting pre-registration reports or encouraging authors to be transparent in reporting their results (THARK-ing) related to QRPs (Anderson et al., 2019; Chambers, 2013; Chen, 2018; Grand et al., 2018; Hollenbeck & Wright, 2017). In addition, there have also been prevention efforts (PRE) against QRPs from institutions or universities. For example, mentoring programs for junior researchers, workshops on ethical research practices, publication requirements in journals that have ethical guidelines and inclusion of members on the committee of publication ethics (COPE). In this study, we consider the PREs undertaken in the workplaces of researchers, rather than author guidelines from journals. We consider these actions to be early preventative efforts, before further PREs are made by editors and reviewers. We argue that early PREs may play an important role in combatting unacceptable research practices. Through their use, researchers are taught about the potential dangers of this kind of practice from the beginning of their careers. Chambers (2014) argues that these PREs are effective in reducing QRPs. In a study conducted by Pratt et al. (2019), a significant negative relationship was found between PREs and research misconduct among scholars. Based on the above discussion, our third hypothesis is:

H3 Prevention efforts have a negative effect on the prevalence of QRPs.

Perceived Behavioral Control and QRPs

PBC represents the perceived ease or difficulty of performing certain behaviors, depending on the opportunities that exist, with greater difficulty reducing both intention to engage in a specific behavior and actual behavior, and vice versa (Ajzen, 2005; Rajah-Kanagasabai & Roberts, 2015). In other words, PBC will encourage someone to take a particular action when the action is perceived as being easy to perform. Ajzen (2005) argues that PBC is a direct predictor or antecedent of actual behavior and that the greater the level of PBC an individual experiences, the stronger their involvement in actual behavior. In this regard, prior to taking action to engage in QRPs, a scholar will likely assess how much control he/she has over the situation and the likelihood of being caught by triggering this action. Hence, when a researcher feels that the likelihood of being caught is very small, or non-existent, their intention to engage in unacceptable research practices will be higher. Several previous studies have shown a positive and significant relationship between PBC and research misconduct among students (Rajah-Kanagasabai & Roberts, 2015; Stone et al., 2009). Based on the above discussion, the fourth hypothesis is:

H4 Perceived behavioral control has a positive effect on the prevalence of QRPs.

Figure 2 portrays the QRP research framework which is empirically tested in this work.

Fig. 2 Conceptual model and relationships among variables
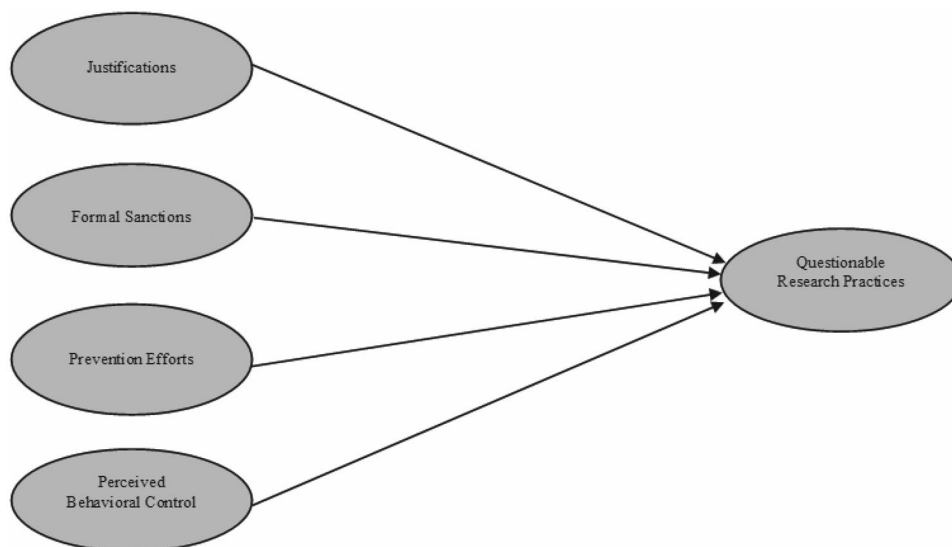
Study #1

The objective of Study 1 was to provide preliminary evidence regarding the self-admission rate and estimated prevalence of potential QRPs among business scholars in Indonesia. The results from Study 1 may provide an indication whether QRP use is either quite low or widely used among the sample analyzed. If the results of Study 1 indicate that the use of QRPs among the sample analyzed is quite high, the findings of Study 1 will provide justification for conducting Study 2.

To achieve the objective of Study 1, two different methods were used to calculate self-admission rate and estimated prevalence (Fiedler & Schwarz, 2016; John et al., 2012). In addition, the results of the self-admission rate and estimated prevalence were compared with previous studies conducted in the US, Germany, Italy and Brazil. A sample was collected through the use of a survey questionnaire to calculate these two measures.

Sample and Data Collection

In Study 1, we chose business scholars affiliated with the top 250 universities in Indonesia as our sample. We opted for business scholars rather than students because scholars have more experience with the process of research and publication. This position is aligned with previous studies conducted by Bedeian et al. (2010), John et al. (2012), Braun and Roussos (2012), Banks et al. (2016) and Motyl et al. (2017). For details on the survey and the sampling frame, see Appendix A in the supplementary material, available online.

After exceeding the specified cut-off date, we received 472 responses, giving a total response percentage of 39.20%. This response rate is generally consistent with – or even slightly higher than – previous relevant studies (Banks et al., 2016; Bedeian et al., 2010; Braun & Roussos, 2012; John et al., 2012). We analyzed our data to detect non-response bias (Fulton, 2018) by comparing the two response waves using the independent *t*-test. We found no significant differences between the two response waves regarding the QRP items. Finally, we assessed social desirability bias (SDB), which often occurs when researching sensitive issues, such as QRPs (Nunnally & Bernstein, 1994). We used an indirect question approach and compared the two responses from a personal and third-person perspective (Bedeian et al., 2010). We found no significant differences between the two responses. As noted by Fiedler and Schwarz (2016), other biases, such as memory bias, pose a major challenge when studying QRPs. Since there are no statistical tests that can account for memory bias, we conducted an online interview via Skype with 10 randomly selected respondents. The results obtained are quite satisfactory, showing that the respondents still had an accurate memory of their responses. However, we acknowledge that these biases may still exist, even though we did not detect them at this time.

Measurement Items and Scales

We adopted the items used to measure QRPs in this study from Fiedler and Schwarz (2016), which is an enhanced version of a previous study conducted by John et al. (2012). Fiedler and Schwarz (2016) argue that the QRP items used by John et al. (2012) are ambiguous, unclear and have weaknesses in the response format. However, several previous studies have used the QRP items from John et al. (Agnoli et al., 2017; Banks et al., 2016; Rabelo et al., 2020) without any modification. We argue that the enhanced version proposed by Fiedler and Schwarz (2016) is better at capturing the essence of QRPs. As far as we know, there have been no further previous studies using this enhanced version of the QRP items. A complete list of the QRP items used is presented in Table 1 (see Table 1A in Appendix A in the supplementary material for the actual wording of the questionnaire used).

These QRP items were measured using various scales. First, to calculate the self-admission rate and estimated prevalence, we followed the approach offered by John et al. (2012) and Fiedler and Schwarz (2016), where the respondents were asked to provide impersonal estimates for each QRP item regarding: (a) the percentage of other business scholars who had engaged in each type of misconduct; and (b) among those business scholars who had done so, the percentage that would admit to having done so. In addition, respondents were asked to provide another three responses relating to each QRP item. First, respondents were asked whether they had adopted these practices themselves (0 = No; 1 = Yes). Second, they were asked how frequently this misconduct was carried out (1 = Never, 2 = Once, 3 = Twice or more, 4 = Often). Finally, respondents were asked to indicate whether such practices could be justified (1 = No; 2 = Possibly; 3 = Yes). After completing these questions, respondents had the option to leave comments explaining why they thought the practice in question was (or was not) defensible. We calculated self-admission rate (SAR1) based on the respondents who actually admitted that they had personally engaged in each practice (John et al., 2012). In this case, the respondents who answered 'Yes' were considered to have adopted this practice, and this figure was used to adjust the estimated prevalence (PE1) and produce prevalence estimates derived from the admission estimates (i.e., self-admission rate/admission estimate). In addition, we also calculated self-admission rate (SAR2) and estimated prevalence (PE2) following Fiedler and Schwarz's method (2016), based on respondents who admitted to having used a practice more than once in all of their studies.

Table 1 Self-admission rates and prevalence estimates in percentages for scholars from the US, Germany, Italy, Brazil, and Indonesia

Questionable research practices	Code	US (John et al., 2012)		Germany (Fiedler & Schwarz, 2016)		Italy (Agnoli et al., 2017)		Brazil (Rabelo et al., 2020)		Indonesia (Present study)			
		SAR1	PE1	SAR2	PE2	SAR1	PE1	SAR1	PE1	SAR1	PE1	SAR2	PE2
1. Failing to report all dependent measures that are relevant to a finding	QRP1	63.4	59.03	34.0	8.0	47.9	58.51	22.4	47.43	57.63	61.41	41.53	15.51
2. Collecting more data after seeing whether results are significant in order to render non-significant results significant	QRP2	55.9	61.01	33.0	6.0	53.2	63.18	22.4	49.22	59.32	63.14	40.94	15.46
3. Failing to report all conditions that are relevant to a finding	QRP3	27.7	35.64	25.0	4.0	16.4	40.80	34.5	53.58	42.80	44.59	28.73	7.34
4. Stopping data collection after achieving the desired result concerning a specific finding	QRP4	15.6	38.98	6.9	1.0	10.4	37.30	10.3	40.27	23.94	44.56	12.38	1.72
5. Rounding off p values (e.g., reporting a p-value of .054 as .05)	QRP5	22.0	40.55	22.0	3.9	22.2	46.79	17.7	45.53	22.67	43.63	11.31	1.49
6. Selectively reporting results regarding a specific finding that 'worked'	QRP6	45.8	59.90	42.0	11.0	40.1	65.82	54.7	60.78	49.15	60.52	24.47	6.56
7. Deciding whether to exclude data after looking at the impact of doing so regarding a specific finding	QRP7	38.2	45.24	40.0	8.0	39.7	55.91	19.8	43.50	43.43	50.35	18.86	4.97
8. Claiming to have predicted an unexpected result	QRP8	27.0	47.73	47.0	10.0	37.4	59.41	9.1	41.06	37.29	50.98	16.29	3.32
9. Claiming that results are unaffected by demographic variables (e.g., gender) despite being unsure (or knowing that they are)	QRP9	3.0	18.72	3.0	0.6	3.1	28.55	4.3	33.64	20.34	36.72	9.32	0.99
10. Falsifying data	QRP10	0.6	9.33	3.0	0.3	2.3	18.65	0.86	17.15	44.70	51.31	27.96	7.29
Mean self-admission rate and prevalence estimate		**29.9%**	**41.61%**	**25.6%**	**5.3%**	**27.3%**	**47.49%**	**19.6%**	**43.22%**	**40.13%**	**50.72%**	**23.18%**	**6.47%**

SAR1 is self-admission rate according to John et al. (2012); SAR2 is self-admission rate according to Fiedler and Schwarz (2016); PE1 is prevalence estimate according to John et al. (2012); PE2 is prevalence estimate according to Fiedler and Schwarz (2016)

Results

We calculated the self-admission rate and estimated prevalence from our original sample and compared our results with previous studies conducted in the US (John et al., 2012), Germany (Fiedler & Schwarz, 2016), Italy (Agnoli et al., 2017) and Brazil (Rabelo et al., 2020), as depicted in Table 1. We used the two approaches offered by both John et al. (2012) and Fiedler and Schwarz (2016) to calculate both measures. The main difference in terms of the results yielded by the two approaches lies in the frequency of QRPs' use. In general, estimated prevalence indicates the repeated involvement of scholars in unacceptable research practices. The results of our analysis using both approaches are shown in Fig. 3a and b.

Following Fiedler and Schwarz's (2016) approach, we found the self-admission rate and estimated prevalence across all QRP items for the Indonesian context to be quite low, with the mean self-admission rate (SAR2) and estimated prevalence (PE2) in our case found to be around 23.18% and 6.47%, respectively. We compared these results with the findings reported by Fiedler and Schwarz (2016), and found no significant difference between the two, with p-value of ANOVA $= 0.577 > 0.05$.

Furthermore, when we applied the original approach used by John et al. (2012) for calculating SAR1 and PE1 in our data, we obtained a mean estimated prevalence of 50.72% (with self-admission rate $= 40.13\%$), which does not differ significantly from previous studies. To corroborate this conclusion, we obtained a p-value of ANOVA for Indonesia vs. US of $0.153 > 0.05$. Furthermore, we also obtained similar results for the cases of Italy and Brazil, with p values of 0.579 and $0.124 > 0.05$, respectively. These differences in values occur because the approach taken by John et al. (2012) does not take into account the frequency of QRP, as depicted in Fig. 3a and b. In addition, the discrepancy between these two approaches has inflated the estimated prevalence of QRPs in previous studies to an unknown degree. In this situation, the estimated prevalence is correct in reference to scholars who are often involved in QRPs, at least more than once. In line with this assertion, Fiedler and Schwarz (2016) argue that the prevalence of QRPs represents the proportion of respondents who repeat this misconduct across all studies they have conducted. Therefore, we stand by our previous conclusions.

Moreover, almost all business scholars in Indonesia (93.22%) report having employed at least one QRP. While this percentage is not significantly different from the findings of studies conducted in Italy (88%), the US (91%) and Brazil (85%), with p-value > 0.05, very few of the Indonesian scholars surveyed admitted ever falsely claiming that results were unaffected by demographic variables (QRP 9) or stopping data collection after achieving the desired results concerning a specific finding (QRP4). Similar results were also found in studies in the US, Italy and Brazil.

We then inspected the QRP items individually to find out which actions are most commonly used by business scholars in our context. We found that the self-admission rates for QRP1 (Failing to report all dependent measures that are relevant to a finding), QRP2 (Collecting more data after seeing whether results were significant in order to render nonsignificant results significant), QRP3 (Failing to report all conditions that are relevant to a finding), QRP6 (Selectively reporting results regarding a specific finding that 'worked') and QRP10 (Falsifying data) show that these are the most common QRPs among business scholars in Indonesia.

We may note that these selected actions have both similarities and differences from those reported in previous studies in the US, Italy, Germany and Brazil. Specifically, in line with previous studies, our findings regarding the practices QRP1, QRP2 and QRP6 as those most predominantly used by business scholars in Indonesia, is similar to the results from other countries (i.e., the US, Italy, Germany and Brazil). On the other hand, our findings regarding practices QRP3 and QRP10 differ slightly from previous studies. We argue that these two practices are related to each other. In certain situations, when a person fails to report all conditions relevant to the findings, he/she will also be more likely to engage in falsifying data. We note some of the justifications given to support these two practices: difficulties in the data collection process (generally related to several response biases) leading to errors in data analysis; lack of financial support for conducting surveys which results in many scholars taking shortcuts; numerous colleagues committing these acts without being caught; lack of regulations governing scientific fraud and the pressure to increase the number of publications, among others. Although most scholars admit that this violates research ethics, both practices are quite common. A recent study by Macháček and Srholec (2021) reported that scholars in Indonesia are quite involved with predatory journals; which allows research malpractice to thrive. In line with this, previous research has also reported the use of these two practices, albeit in relatively small percentages.

A study conducted by Banks et al. (2016) supports our findings, reporting that more than 50% of management researchers from their sample engaged in QRPs, such as selectively reporting hypotheses on the basis of statistical significance and presenting a post hoc hypothesis as if it had been developed a priori. In addition, Banks et al. (2016) report that management scholars are less likely to choose actions such as falsified data and "rounded off" p values. These results partially support our findings. Finally, we conclude that the self-admission rate percentage for each item in our case was generally slightly higher than in other studies (see Table 1).
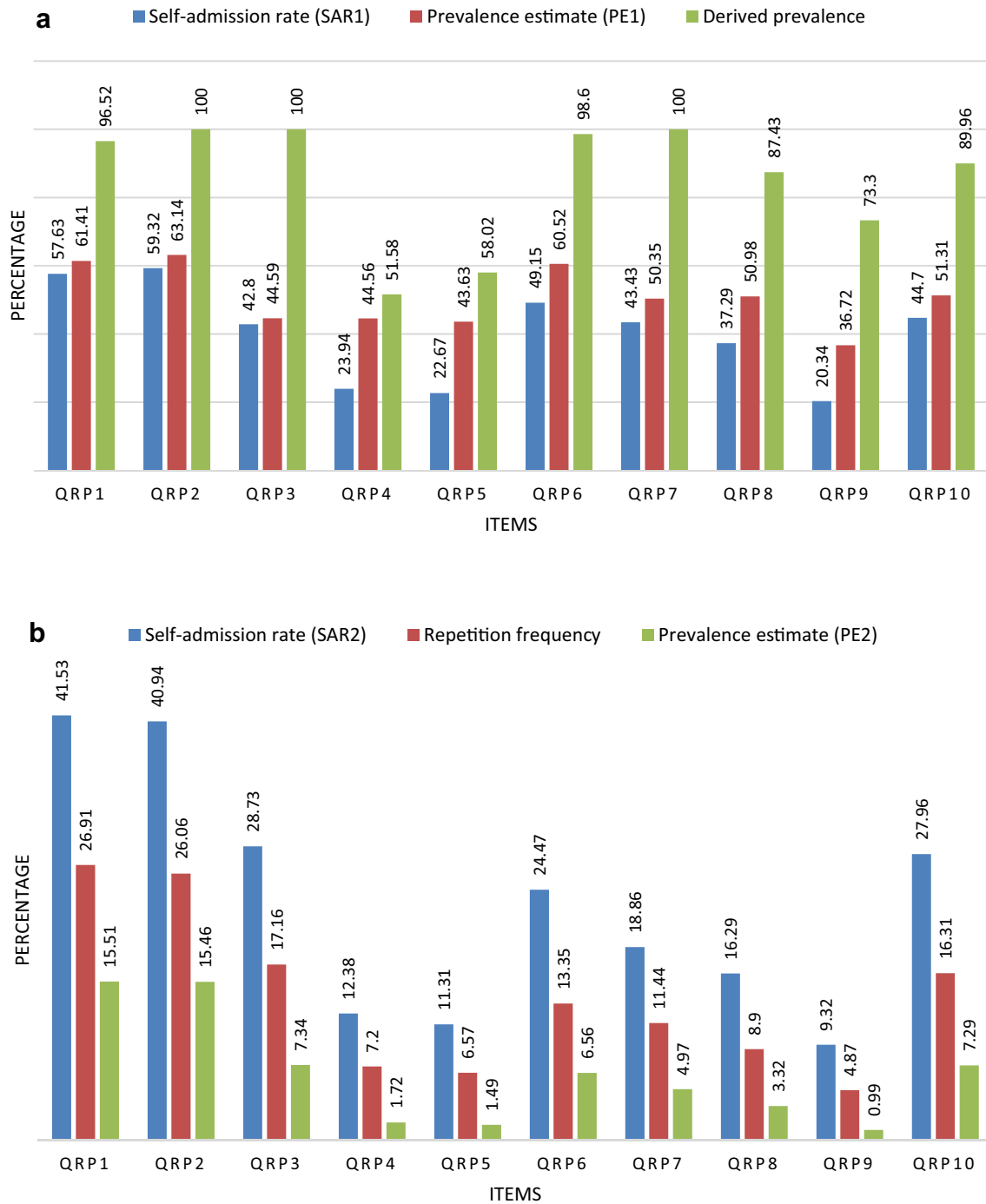
a

b

Fig. 3 **a** Assessing questionable research practices using John et al.'s (2012) Approach. **b** Assessing questionable research practices using Fiedler and Schwarz's (2016) approach

Study #2

Based on the findings of Study 1, which showed a high level of QRP use among business scholars in Indonesia based on our calculations of self-admission rate and estimated prevalence, we conducted Study 2 to follow-up on these findings. In Study 2, the objective was to conduct an empirical investigation to examine the factors that encourage (i.e., justifications and perceived behavioral control) and discourage (i.e., formal sanctions and prevention efforts) QRPs. Study 2

extends the findings from Study 1 to discover the key factors behind this unethical behavior. To achieve the objective of Study 2, all hypotheses were examined to test the proposed model of QRPs using a second data sample collected from scholars who admitted to involvement in QRPs during Study 1.

Sample and Data Collection

For Study 2, we used a sub-section of the same target respondents as Study 1; i.e., those scholars who admitted to being involved in the use of QRPs at least once. For further details on the survey and the sampling frame, see Appendix B in the supplementary material, available online.

We received 156 responses after the deadline for returning the questionnaire was reached. Although this number is smaller than the respondents of Study 1, we must take into account that academics and universities are busier at the end of the calendar year, with university activities and end-of-semester exams. However, such a sample size is still considered acceptable in many social science disciplines (Dillman et al., 2014). We conducted tests to detect several potential biases in the data collection process, including non-response bias and common method bias. We used a t-test to compare the two response waves. Our results found no significant differences ($p > 0.05$) between the two waves using either Levene's test or equality of means testing for the variables tested (see Table 2). We, therefore, conclude that non-response bias is not a threat to our data analysis in this study. In addition, we assessed common method bias, which often occurs when using the self-reporting method. We used Harman's single factor test to detect this bias. We found that five significant factors were extracted, with the largest percentage of variance due to one factor being 36.83% (less than 50%). Therefore, we conclude that common method bias does not occur in our measurements. Once again, while we did not find any evidence of these biases in our results, we do acknowledge that these biases may still exist.

A summary of the profile of respondents to this study is shown in Table 3. Most of our respondents are male (55.10%), while the rest are female (44.87%); the majority

Table 3 Characteristics of the sample

Demographic variable	Frequency	Percentage (%)
Gender		
Male	86	55.1
Female	70	44.87
Age (years)		
30–40	26	16.67
41–50	63	40.38
51–60	51	32.69
Over 60 years old	16	10.26
Research Experience		
1–8 years	19	12.18
9–17 years	76	48.72
18–25 years	46	29.49
Over 25 years	15	9.61
Number of Publications (papers)		
1–5	25	16.03
6–10	34	21.79
11–20	43	27.56
Over 20	54	34.62
Academic qualifications (education level)		
Master's degree	31	19.87
Doctorate degree	94	60.26
Post doc	17	10.90
Professor	14	8.97

are aged between 41 and 50 years and possess a doctorate level of academic qualification.

Measurement Items and Scales

We adopted the items used to measure the variables in our model from previous studies (Pratt et al., 2019; Rajah-Kanagasabai & Roberts, 2015; Stone et al., 2009). We selected these items on the grounds that they have proven validity and reliability. A complete list of these items is depicted in Tables 4 and 5 (see Table 1B in Appendix B in the supplementary material for the actual wording of questionnaire used). First, justification (JST) for engaging in QRPs was

Table 2 Assessment of non-response bias

Construct	Sig. Levene's test	Sig. t-test for equality of means
Justifications (JST)	0.971	0.282
Formal Sanctions (FRS)	0.148	0.722
Prevention Efforts (PRE)	0.806	0.393
Perceived Behavioral Control (PBC)	0.662	0.654
Questionable Research Practices (QRP)	0.555	0.431

Table 4 Measurement model assessment of justifications, formal sanctions, prevention efforts, and perceived behavioral control

Indicator/Item	Code	Mean	SD	FL	AVE	α	ρ_c
(A) Justifications (JST)					0.640	0.909	0.925
Because of meeting the expectations of reviewers	JST1	5.468	1.232	0.833			
Because of fear of manuscript rejection	JST2	5.385	1.201	0.850			
Because of negative results	JST3	4.891	1.361	0.744			
Because other researchers do it	JST4	5.045	1.151	0.760			
Because of university pressure for publication	JST5	5.449	1.200	0.869			
Because of monetary reward	JST6	5.244	1.173	0.811			
Because of increasing popularity	JST7	5.314	1.102	0.721			
(B) Formal Sanctions (FRS)					0.838	0.952	0.963
Establishing harsher penalties for researchers who use questionable research practices	FRS1	3.321	1.676	0.889			
Increasing protections for whistleblowers, such as lab staff, who expose research misconduct	FRS2	3.147	1.475	0.887			
The establishment of due process requirements to guide research misconduct investigations	FRS3	3.500	1.658	0.948			
Criminalizing serious forms of questionable research practices	FRS4	3.045	1.486	0.913			
Professional associations establishing formal sanctions for research misconduct	FRS5	3.410	1.621	0.939			
(C) Prevention Efforts (PRE)					0.578	0.797	0.843
Requiring researchers to attend workshops on ethical research practices	PRE1	5.564	1.099	0.869			
Providing grant application writing workshops to junior researchers	PRE2	5.692	1.207	0.833			
Providing mentoring programs to junior researchers	PRE3	5.327	1.122	0.713			
Having journals regularly publish ethnical guidelines	PRE4	5.308	1.408	0.594			
(D) Perceived Behavioral Control (PBC)					0.784	0.908	0.935
It is easy to engage in questionable research practices and not get caught	PBC1	4.885	1.498	0.862			
It is easy to engage in *p* hacking of results	PBC2	4.923	1.551	0.928			
My colleagues often employ questionable research practices and have not been caught	PBC3	4.628	1.614	0.903			
Other researchers employing questionable research practices are often not caught	PBC4	5.160	1.425	0.845			

FL = factor loading, *SD* = standard deviation, *AVE* = Average variance extracted; α = Cronbach's Alpha; ρ_c = Composite reliability

Table 5 Measurement model assessment of questionable research practices

Indicator/Item	Code	Mean	SD	FL	AVE	α	ρ_c
(E) Questionable Research Practices (QRPs)					0.583	0.919	0.933
Failing to report all dependent measures that are relevant to a finding because of negative results	QRP1	5.218	1.242	0.833			
Collecting more data after seeing whether results are significant in order to render non-significant results significant	QRP2	4.628	1.199	0.794			
Failing to report all conditions that are relevant to a finding	QRP3	4.795	1.244	0.787			
Stopping data collection after achieving the desired result concerning a specific finding	QRP4	4.929	1.261	0.709			
Rounding off p values (e.g., reporting a p-value of .054 as .05)	QRP5	5.205	1.295	0.873			
Selectively reporting results regarding a specific finding that 'worked'	QRP6	5.263	1.292	0.826			
Deciding whether to exclude data after looking at the impact of doing so regarding a specific finding	QRP7	5.590	1.154	0.645			
Claiming to have predicted an unexpected result	QRP8	5.615	1.168	0.606			
Claiming that results are unaffected by demographic variables (e.g., gender) despite being unsure (or knowing that they do)	QRP9	4.865	1.415	0.751			
Falsifying data	QRP10	4.782	1.195	0.770			

FL = factor loading, *SD* = standard deviation, *AVE* = Average variance extracted; α = Cronbach's Alpha; ρ_c = Composite reliability

measured using 7 items adapted from Stone et al. (2009) and which have been used by Rajah-Kanagasabai and Roberts (2015). We used a 7-point Likert scale ranging from "1 = very unlikely" to "7 = very likely" to measure this variable. For instance, respondents were asked "how likely are you to engage in QRPs, in order to meet the expectations of reviewers?" and so on. Second, we measured the variables measuring formal sanctions (FRS) and prevention efforts (PRE) using 5 and 4 items, respectively, adopted from Pratt et al. (2019). We used a 7-point Likert scale ranging from "1 = no effect at all" to "7 = very major effect" to measure both variables. For example, respondents were asked to indicate "how much of an effect does the approach described in each item have in reducing QRPs?" etc. Third, the level of perceived behavioral control (PBC) regarding engagement with QRPs was measured using 4 items adapted from Stone et al. (2009). This time we used a 7-point Likert scale ranging from "1 = strongly disagree" to "7 = strongly agree". Respondents were asked questions regarding whether it was easy or difficult to engage in QRPs. Finally, we measured QRPs using items modified by Fiedler and Schwarz (2016) on a 7-point Likert scale ranging from "1 = very problematic" to "7 = very reasonable". Respondents were asked for their assessment regarding each of the procedures or research practices mentioned (see Table 1B in Appendix B in the supplementary material).

Data Analysis

We selected the partial least squares-structural equation modeling (PLS-SEM) method, using the SmartPLS 3 software (Ringle et al., 2015), to test the proposed model of QRPs. PLS-SEM is a useful method when the model under consideration is still in the early stages of development. Given that the model we are proposing is still primitive, with a relative scarcity in terms of existing literature and prior knowledge, PLS-SEM is seen as a suitable approach (Lohmöller, 1989; Noonan & Wold, 1986). We calculated the minimum sample requirements for our model under this system and found that our data exceed the minimum requirement

of 146 cases (where the minimum absolute significant path coefficient = 1.97, significant level = 0.05 and required power level = 0.80). In short, we herein follow the best practice reporting guidelines that have been documented regarding this method (Benitez et al., 2020; Latan, 2018).

Results

Tables 4 and 5 present descriptive statistics (means and standard deviations) regarding each indicator variable, while above the diagonal of Table 6 presents the correlations between variables. Following the guidelines proposed by Bedeian (2014), we can confirm that the mean and standard deviation values of the variables in the model do not exceed the maximum and the correlation sign is not reversed. We did not find any correlation greater than 0.70 among all possible pairs of relationships between variables, which gives an initial indication that our data are free from collinearity. To support this conclusion, we calculated the Variance Inflation Factor (VIF) for each predictor and found VIF values < 3, which meets the rule of thumb threshold (see Table 7). In addition, we can confirm that the descriptive statistics and correlations between the variables under study were based on complete data, without missing values.

Measurement Model Assessment

We assessed our measurement model in order to enhance the rigor of our statistical analysis and reporting standards. First, we assessed the convergent validity of the items used to measure constructs in the model, as depicted in Tables 4 and 5. We obtained loading factor values > 0.708 for majority of items, and average variance extracted (AVE) values > 0.5 for all variables (Benitez et al., 2020; Hair et al., 2017), which meets the specified boundaries. Three items have loading factor values between 0.5 and 0.6 which are still considered acceptable, as long as the AVE value meets the required cut-off value to strengthen content validity. Subsequently, we assessed discriminant validity using the hetero-trait-mono-trait ratio (HTMT). As can be seen in below the

Table 6 Assessment of discriminant validity using the hetero-trait mono-trait ratio and correlations

Construct	1	2	3	4	5
FRS	**(0.900)**	0.022	0.082	−0.015	−0.242*
JST	0.046 [0.192;0.062]	**(0.900)**	0.060	−0.201	0.482*
PBC	0.760 [0.806;0.707]	0.054 [0.196;0.066]	**(0.900)**	−0.123	0.302*
PRE	0.129 [0.245;0.083]	0.709 [0.818;0.598]	0.149 [0.281;0.097]	**(0.900)**	−0.310*
QRPs	0.796 [0.859;0.750]	0.118 [0.262;0.112]	0.773 [0.847;0.692]	0.149 [0.276;0.133]	**(0.900)**

Brackets show the lower and upper bounds of the 95% BCa confidence intervals; the values above the diagonal show the correlation between variables; the values below the diagonal show the HTMT index

*Statistically significant at the $p < 0.05$ (two-tailed test)

Table 7 Structural model assessment

Construct	R^2	Adj. R^2	f^2	Q^2	VIF	SRMR	*Harman's single factor*
Justifications (JST)	–	–	0.277	–	2.758	–	36.83% variance explained
Formal Sanctions (FRS)	–	–	0.100	–	1.691	–	
Prevention Efforts (PRE)	–	–	0.057	–	2.891	–	
Perceived Behavioral Control (PBC)	–	–	0.081	–	1.737	–	
Questionable Research Practices (QRPs)	0.633	0.623	–	0.355	–	0.076	

diagonal of Table 6, we did not find HTMT values greater than 0.90 using either the lower or upper bounds of the 95% confidence intervals (Hair et al., 2017). Hence, discriminant validity is confirmed. Finally, we assessed the reliability of internal consistency using Cronbach's alpha (α) and composite reliability (ρ_c). We found that both measures produced values between 0.797 and 0.952, which indicates that the reliability of the measurements is confirmed. Figure 4 shows a visual depiction of the results of this measurement model assessment.

Structural Model Assessment

After confirming the quality of the measurement model, we next assessed the structural model and related parameters. Figure 4 shows that 63.3% of the variance (R^2) among QRPs can be explained by predictors in the model (in this case JST, FRS, PRE and PBC), with this value exceeding the moderate category baseline of 50% indicated by Hair et al. (2017). A very high R^2 value (e.g., > 90%) would indicate over-fit and collinearity issues. Next, we assessed the effect size (f^2) to determine the magnitude of the contribution of each predictor. Values of f^2 ranging from 0.02–0.15 to 0.15–0.35 fall into the medium and large categories, respectively. In our
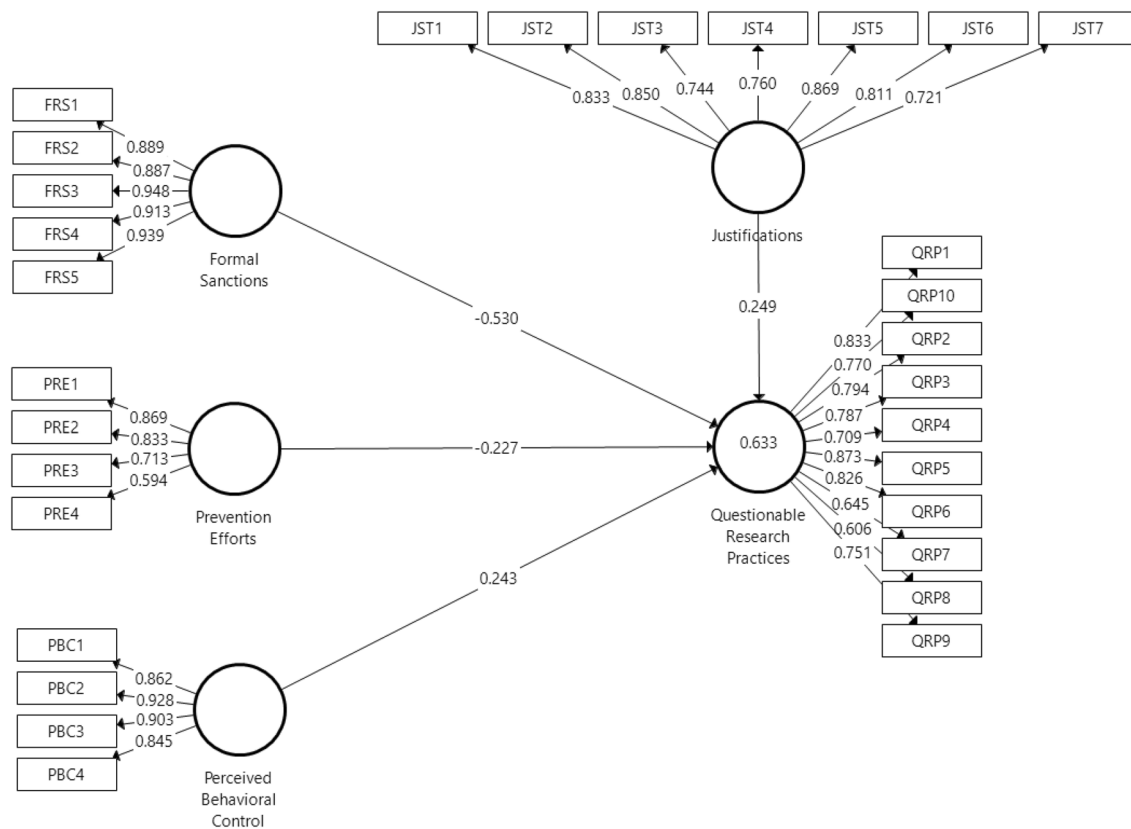

Fig. 4 Evaluation of the measurement and structural models

results, the highest f^2 values were found in the relationship between JST and QRPs (0.277) and between FRS and QRPs (0.100). We then considered predictive relevance (Q^2) as an alternative measure of R^2 to ensure the predictive power of the model. A Q^2 value larger than zero is meaningful and hints at a model that is worth testing. We assessed Q^2 through the blindfolding procedure and obtained a value of 0.355 > 0 for QRPs, confirming the predictive power of our model. Finally, we assessed the standardized root mean square residual (SRMR) index, defined as the standardized difference between the observed and predicted correlations. We obtained an SRMR value of 0.076 < 0.08, indicating that the model has a sufficient degree of fit (Kline, 2016).

Testing of Hypotheses

We tested our hypotheses using 95% confidence intervals (CI) and a significance level of 5% (one-tailed test) through the bootstrapping procedure. We used 10,000 subsamples to ensure stability of estimates. We selected confidence interval methods, namely bias-corrected and accelerated (BCa) bootstrapping. Overall, the results obtained support the proposed hypotheses regarding the relationships between the predictors and outcomes. In relation to justification for engaging in QRPs, we found a positive and significant effect of the link JST→QRPs ($\beta = 0.249$, $p < 0.01$ at 95% CI), which confirms Hypothesis 1 (H1). In addition, we found that formal sanctions (FRS) and prevention efforts (PRE) have a negative and significant effect on QRPs. In Table 8, we obtain significance values for the relationships FRS→QRPs ($\beta = -0.530$, $p < 0.01$ at 95% CI) and PRE→QRPs ($\beta = -0.227$, $p < 0.01$ at 95% CI), which show support for Hypothesis 2 (H2) and Hypothesis 3 (H3). Finally, regarding the effect of perceived behavioral control (PBC) on misconduct, our results reflect a positive and significant relationship of PBC→QRPs ($\beta = 0.243$, $p < 0.01$ at 95% CI), which corroborates Hypothesis 4 (H4). From these results we can conclude that JST ($\beta = 0.249$) has a greater influence in encouraging the involvement of business scholars in Indonesia to conduct QRPs compared to PBC ($\beta = 0.243$). We support this conclusion with the resulting f^2 value for JST = 0.277 which falls into the large category, indicating the dominance of this variable in encouraging the use of QRPs. In this case, JST can be considered as one of the main factors that encourage

the involvement of business scholars in Indonesia in the use of QRPs. On the other hand, FRS ($\beta = -0.530$) is a factor that is shown to have a more significant influence in reducing QRPs compared to PRE ($\beta = -0.227$). We support this conclusion with the resulting f^2 value for FRS = 0.100, which falls within the moderate category, indicating the dominance of this variable in reducing the use of QRPs. In this regard, FRS can be considered as an effective means of combating the use of QRPs. However, all the factors tested display an acceptable level of significance, which indicates that they are able to predict QRPs empirically.

Robustness Checks

We performed two robustness tests as a complementary check to strengthen our results. We assessed endogeneity bias using the Heckman test. We did not find a significant difference with our main results. The Heckman test results showed no signs of change after being controlled for by a third variable in each regression equation. Therefore, we conclude that endogeneity bias is not a threat to the validity of our results. Finally, we examined non-linear effects to ensure that the assumptions of linearity in the model are fulfilled. We used Ramsey's regression specification error test (RESET) and quadratic functions to detect this bias. We obtained insignificant results ($p > 0.05$) for all possible non-linear relationships in the model. The same results were also identified in the RESET test ($p > 0.05$). In addition, the f^2 values obtained fall into the small category (i.e., < 0.021). From these results we conclude that our model has been correctly specified (Wooldridge, 2020) and supports the assumption of linearity in the relationships between variables.

Discussion of Study 1 and Study 2

Sixty years ago, Sterling (1959) made a fundamental contribution to the research community by emphasizing the danger of selective reporting practices regarding research findings, which could lead to incorrect conclusions. He noted that most researchers at that time had been involved in some kind of research misconduct. Almost forty years later, this dysfunctional behavior remained largely overlooked and was

Table 8 Testing of hypotheses

Structural path	Coef (β)	SD	*p*-value	95% BCa CI	Conclusion
JST→QRPs	0.249	0.103	0.008**	(0.402, 0.137)**	H1 supported
FRS→QRPs	−0.530	0.081	0.000**	(−0.401, −0.671)**	H2 supported
PRE→QRPs	−0.227	0.094	0.008**	(−0.137, −0.359)**	H3 supported
PBC→QRPs	0.243	0.090	0.003**	(0.386, 0.089)*	H4 supported

**, *Statistically significant at the $p < 0.01$ and $p < 0.05$ (one-tailed test), respectively; *Coef(β)* = beta coefficient, *SD* = standard deviation; *BCa CI* = bias-corrected and accelerated confidence intervals

not considered a critical issue in the development of scientific practice. Greater attention to such misconduct began to develop when an article by Kerr (1998) and a study by John et al. (2012) uncovered these practices. Although scientific fraud and unacceptable research practices will always exist, understanding the prevalence of both is an important step toward crafting strategies to combat them (Anderson et al., 2019; Chambers, 2014).

This study seeks to fill this gap by examining the self-admission rate and estimated prevalence of QRPs among business scholars. In addition, we have also empirically examined the factors that encourage and discourage this misconduct. Given the lack of research addressing QRPs and the effectiveness of strategies to reduce such misconduct in business research (Leung, 2011; Shaw, 2017), the two studies (Study 1 and Study 2) contained herein can be considered one of the first to provide relevant empirical evidence on this topic.

In general, our results indicate similarities between the use of QRP items among different countries (e.g., the US, Germany, Italy and Brazil), but some differences can also be highlighted. In terms of similarity, most of our sample (93.22%) acknowledged all QRP items presented to them, indicating that this list of items was recognizable to business scholars from Indonesia, as well as those from other countries. On the other hand, in terms of differences, business scholars in Indonesia tend to use different QRPs in their studies compared to other countries. Finally, we found certain strategies to reduce QRPs to be very effective, such as formal sanctions and prevention efforts.

Specifically, we can elaborate our findings as follows. In Study 1, we found the mean self-admission rate (SAR1 = 40.13%; SAR2 = 23.18%) and estimated prevalence (PE1 = 50.72%; PE2 = 6.47%) of business scholars' involvement in QRPs to be quite high. When we compared the self-admission rate (SAR1) and the estimated prevalence (PE1) of our sample with previous studies from other countries according to the approach of John et al. (2012), these numbers were slightly above average for each individual practice. This difference may depend on several key factors, such as research culture, supervision and educational regulations that apply in the respective countries of study. In Indonesia, the research culture is progressing gradually and there remains an absence of regulations governing scientific fraud. Under such conditions, unacceptable research practices may occur more than in other countries with more developed research systems. However, this does not mean that countries with such systems are free from such misconduct. In addition, the estimated prevalence among business scholars calculated when we used the approach of Fiedler and Schwarz (2016) was found to be lower, but in line with the findings of Fiedler and Schwarz (2016). We noted estimated prevalence (PE2) values for each practice ranging from 0.99% to

15.51%, which is far below the self-admission rate (SAR2), which ranges from 9.32% to 41.53%. Additionally, we found that Indonesian business scholars reported their involvement in QRP1, QRP2, QRP3, QRP6 and QRP10 more frequently, with the use of QRP10 being the highest compared to previous studies. Conversely, use of QRP4, QRP5 and QRP9 were less prevalent. These findings reinforce the results of Banks et al. (2016) in the realm of quantitative research. Therefore, our findings can be taken as a warning that cannot be ignored by stakeholders (e.g., editors, funders, supervisors, chancellors, and others).

Furthermore, in Study 2, we found empirical evidence regarding factors that can encourage and discourage QRPs. First, we found a positive and significant relationship between justification (JST) and the involvement of business scholars in QRPs. That is, the higher the level of JST given by scholars, the greater the likelihood of them engaging in misconduct. We found pressure for publications from universities, fear of rejection of manuscripts, meeting the expectations of reviewers and gaining monetary reward to be the main reasons encouraging business scholars in Indonesia to become involved in QRPs. Such JSTs are very relevant, considering that the Indonesian research ministry is working to increase the number of publications from universities in Indonesia. Therefore, there is pressure on scholars in Indonesia to publish in international journals which are indexed by SSCI or Scopus. To achieve publication, it is inevitable that certain scholars, in trying to meet the expectations of reviewers, will not report negative results that could cause rejection of the manuscript (Anderson et al., 2019; Chen, 2018; Harvey, 2017; Hollenbeck & Wright, 2017). In addition, financial incentives are provided by several leading universities and institutions in Indonesia based on publication. These factors seem to encourage business scholars in Indonesia to engage with QRPs. Our findings corroborate the results of previous studies that indicate a positive relationship between JST and research misconduct (Necker, 2014; Rajah-Kanagasabai & Roberts, 2015; Stone et al., 2009).

We found a negative and significant relationship between formal sanctions (FRS) and prevention efforts (PRE) and the prevalence of unacceptable research practices. We thus argue that both of these strategies work to reduce QRPs. FRSs include having to return research funding, being blacklisted from journals, the risk of being dismissed from one's job or perhaps even criminal proceedings. We found that such FRSs have a greater effect in reducing QRPs. That is, potential perpetrators might rethink engaging in unacceptable practices given the severity of the penalties. In Indonesia, only half of these FRSs are applied, namely returning research budgets and being blacklisted. However, there are no regulations leading to criminal proceedings for scientific fraud in Indonesia. As pointed out by Pratt et al. (2019), unacceptable research practices can occur when existing

sanctions may be considered too light by the actors. On the other hand, we found PREs such as conducting workshops on ethical research practices or providing assistance to junior researchers to be useful precautions against QRPs. In this regard, such activities will increase the ethical awareness of scholars to avoid unethical behavior. Our findings jointly reinforce the results of previous studies that found a negative effect of PREs and FRSs on QRPs (Honig et al., 2014; Pratt et al., 2019).

Finally, we found a positive and significant effect between PBC and the involvement of business scholars in QRPs. That is, the easier scholars perceive engaging in QRPs to be, the more likely they are to engage in such practices. In this sense, when the likelihood of being caught is small, perpetrators may not worry about engaging in such practices. Conversely, when the likelihood of being caught is high, they may perceive QRPs as difficult a behavior to enact. We found that our sample was motivated by the ease of using QRPs without being caught. In addition, the actions of colleagues with regard to QRPs can also play an important role for business scholars in Indonesia engaging in QRPs. These findings support previous studies which found a positive relationship between PBC and research misconduct (Rajah-Kanagasabai & Roberts, 2015; Stone et al., 2009).

Theoretical and Practical Implications

Based on our findings from both studies, several theoretical and practical implications may be provided. First, regarding Study 1, our results have theoretical implications for a broader understanding of QRPs in the realm of quantitative research. While previous studies have revealed limitations on data and evidence related to QRPs in this field (Leung, 2011; Shaw, 2017), our results provide two important findings regarding the self-admission rate (SAR1 and SAR2) and estimated prevalence (PE1 and PE2) of business scholars' engagement in QRPs. Our findings provide new insights into the frequency of QRPs' use in business research, whereas previous studies, for example from Banks et al. (2016), do not provide this information, and only ask whether this practice has been used "at least once". As stated by Fiedler and Schwarz (2016), the frequency of individuals using QRPs "at least once in all studies" does not give a clear view of the overall prevalence of QRPs. Therefore, we also report information on the use of QRPs "twice or more often", which allows us to better estimate the overall prevalence of QRPs. Our results show that while previous findings indicate involvement in QRPs at least once to be very high (in this case, based on self-admission rate), only a small percentage of business scholars admit to using such practices twice or more often in their studies.

In addition, our findings have practical implications for stakeholders (in this case editors, funders, supervisors, deans and others) by showing that unacceptable research practices in business research exist and cannot be denied. Based on understanding the existence of such misconduct, preventive measures as well as related policies and strategies can be developed to combat these unethical research practices in the future. For example, deans of faculty and editors of business research journals should emphasize the importance of avoiding QRPs to graduate students and researchers, especially the practices of not reporting all dependent measures that are relevant to their findings (QRP1) and collecting more data after seeing whether the results are significant in order to render non-significant results significant (QRP2). Both of these QRPs tend to increase publication bias and potential errors (Francis, 2013).

Second, in Study 2, our findings regarding which factors can encourage or discourage QRPs provide new empirical evidence for the QRP literature. Our findings have theoretical implications and expand the state-of-the-art research related to QRPs, given that there are no previous studies addressing the relationships between the variables tested here. As highlighted by Shaw (2017), empirical research related to QRPs remains scarce. Hence, our findings enrich the academic corpus on QRPs. In addition, our findings also reinforce previous studies related to research misconduct (Rajah-Kanagasabai & Roberts, 2015; Stone et al., 2009). The practical implications of these findings relate to the application of formal sanctions and prevention efforts which are considered effective to reduce the involvement of scholars in QRPs. As highlighted by Sijtsma (2016), it is more important to find ways to control this misconduct than to identify instances of it. Therefore, deans of faculty and editors of journals should employ prevention efforts to minimize such misconduct (e.g., conducting workshops or webinars related to ethical research practices) and introduce formal sanctions for perpetrators (e.g., being blacklisted). In addition, where pressure regarding publication and monetary rewards are provided by universities, both should be rational and balanced in order to prevent scholars from engaging in QRPs. High levels of publication pressure can trigger stress and anxiety (O'Boyle et al., 2017), leading to increased engagement in QRPs, while monetary rewards can increase scientific fraud. Finally, these efforts are only intended to raise awareness among business scholars, with the potential to increase transparency, quality, and the impact of their research findings.

Recommendations for Improvement

Albert Einstein once said that if we knew what we were doing, it wouldn't be called research (Mazzola & Deuling, 2013). Thus, are there any solutions which can effectively tackle the use of QRPs? What should be done to combat this issue in the realm of business research? For this purpose, we provide the following several suggestions for universities, reviewers and editors of journals to consider.

First, we believe that the issue of QRPs is related to the prevalent ethics and norms for conducting good research. The problem arises from a decision taken within individual scholars themselves to engage in misconduct, and ethics and norms are thereby violated. In this regard, developing ethical standards for publication and emphasizing their importance to all faculty members and researchers is a good solution to minimize this issue (Bosco et al., 2016; Braun & Roussos, 2012; O'Boyle et al., 2017). In addition, we also recommend that each university and related institution develop a whistle-blowing reporting system. As far as we know, whistleblowing in relation to scientific fraud is uncommon. A study by Latan et al. (2021) indicates that whistleblowing systems can work effectively in revealing wrongdoing, including scientific fraud. When a researcher is involved in QRPs, only he/she and maybe a third person (e.g., close friends, colleagues, co-authors, lab staff or even students) know about it. Therefore, by encouraging the role of whistleblowers, misconduct such as QRPs can be uncovered and minimized. In such a climate, potential perpetrators will think twice because they are more likely to be caught.

Second, we recommend reviewers to ask authors to follow general reporting standards in reporting their findings (Benitez et al., 2020; Latan, 2018). In this case, authors are asked to report all the metrics needed for transparency purposes (Aguinis et al., 2018). We believe that with more transparent reporting of results, misconduct will be more easily detected. As indicated by Bergh et al. (2017) and Simonsohn et al. (2014), several statistical tests, such as sensitivity tests, can be used to detect QRPs. In addition, reviewers should encourage authors to transparently express their use of post hoc hypotheses (Hollenbeck & Wright, 2017). As Hollenbeck and Wright (2017) argue, transparent HARKing (THARKing) can be an appropriate solution, besides being useful for developing new knowledge.

Finally, we recommend that editors of journals take the following actions to minimize QRPs, as a final preventive effort. First, in the guidelines provided to authors, editors should require authors to make their research data and materials publicly available (Sijtsma, 2016; Wigboldus & Dotsch, 2016). Thus, anyone will be able to replicate the results of the study findings. We argue that this method may already have been applied by most journals. Second, conducting a

multi-stage review (Kepes et al., 2014), often referred to as a pre-registered report, is recommended (Bosco et al., 2016; Chambers, 2014; Nosek & Lakens, 2014). In this way, the peer-review process will be divided into two stages. First, scholars submit their research proposal, including research objectives, hypotheses, research models and study design before conducting their scientific investigation. Next, assessments will be carried out separately before and after the data are collected and analyzed in order to produce empirical findings. A recent study from Bloomfield et al. (2018) concludes that such procedures are effective in reducing publication bias. In fact, we found that several journals have implemented this procedure.

Although pre-registered reports offer several benefits, there is no perfect system. As Yamada (2018) argues, pre-registered reports have several limitations and this procedure is easy to crack. First, despite a journal using a pre-registration system for the submission process, QRPs can still be used. For example, scholars may still report positive findings as a result of data falsification. Second, scholars may still be able to circumvent the logic of this procedure without anyone knowing. That is, the analysis can be conducted first, before the research proposal is registered. This is called pre-registration after the results are known (PARKing). However, we argue that there is no panacea that can eliminate the issue of QRPs completely; we can only work to minimize it.

Limitations and Future Research Directions

Our studies are not free from limitations. First, our study only considers the QRP items proposed by John et al. (2012) and modified by Fiedler and Schwarz (2016). This list may not be exhaustive in capturing all unacceptable research practices. In addition, our study only focuses on misconduct in the realm of quantitative research, without investigating the stream of qualitative approaches. Second, our findings are limited to the Indonesian context and business research, which may differ from other countries and fields. Therefore, our results may not be generalizable to different contexts to show the most prevalent unacceptable research practices. Finally, our study only considers factors such as JST, FRS, PRE, and PBC in influencing QRPs. These factors may not be exhaustive, given that there are still several other variables that might affect QRPs, especially in encouraging or discouraging misconduct.

We suggest several directions for future research to enrich the corpus of research into QRPs. First, future research might consider using different QRP items. For example, the QRP items used by Motyl et al. (2017) may be useful for further investigation. In addition, we suggest that future researchers develop new items for measuring QRPs. Second, based on our literature review, there are no previous studies

addressing unethical research practices in the realm of qualitative research. Although the stream of qualitative research differs from quantitative research, we believe that unacceptable research practices may also exist and require further investigation in this stream. Furthermore, we suggest that future studies could conduct a cross-cultural comparison of QRPs to determine the level of difference in such practices around the world. Finally, empirical investigations related to QRPs should be expanded by examining other factors such as subjective norms, descriptive norms, strains, or ethical awareness that may lead to new avenues for future studies.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

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), e0172792.

Aguinis, H., Ramani, R. S., & Alabduliader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals, 12*(1), 1–28.

Ajzen, I. (2005). *Attitudes, personality and behaviour* (2nd ed.). Open University Press.

Anderson, B. S., Wennberg, K., & McMullen, J. S. (2019). Editorial: Enhancing quantitative theory-testing entrepreneurship research. *Journal of Business Venturing, 34*, 105928.

Banks, G. C., O'Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., et al. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management, 42*(1), 5–20.

Bedeian, A. G. (2014). "More than meets the eye": A guide to interpreting the descriptive statistics and correlation matrices reported in management research. *Academy of Management Learning & Education, 13*(2), 121–135.

Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education, 9*(4), 715–725.

Behfar, K., & Okhuysen, G. A. (2018). Perspective—discovery within validation logic: Deliberately surfacing, complementing, and substituting abductive reasoning in hypothetico-deductive inquiry. *Organization Science, 29*(2), 323–340.

Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management, 57*(2), 103168.

Bergh, D. D., Sharp, B. M., & Li, M. (2017). Tests for identifying "red flags" in empirical findings: Demonstration and recommendations for authors, reviewers, and editors. *Academy of Management Learning & Education, 16*(1), 110–124.

Bloomfield, R. J., Rennekamp, K. M., & Steenhoven, B. A. (2018). No system is perfect: Understanding how registration-based editorial processes affect reproducibility and investment in research quality. *Journal of Accounting Research, 56*(2), 313–362.

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology, 65*(6), 709–750.

Braun, M., & Roussos, A. J. (2012). Psychotherapy researchers: Reported misbehaviors and opinions. *Journal of Empirical Research on Human Research Ethics, 7*(5), 25–29.

Butler, N., Delaney, H., & Spoelstra, S. (2017). The gray zone: Questionable research practices in the business school. *Academy of Management Learning & Education, 16*(1), 94–109.

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex, 49*, 609–610.

Chambers, C. D. (2014). Ten reasons why journals must review manuscripts before results are know. *Addiction, 110*, 9–13.

Chen, G. (2018). Editorial: Supporting and enhancing scientific rigor. *Journal of Applied Psychology, 103*(4), 359–361.

Cox, A., Craig, R., & Tourish, D. (2018). Retraction statements and research malpractice in economics. *Research Policy, 47*(5), 924–935.

Crede, M., & Harms, P. (2019). Questionable research practices when using confirmatory factor analysis. *Journal of Managerial Psychology, 34*(1), 18–30.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed mode surveys: The tailored design method* (4th ed.). Wiley.

Fanelli, D. (2010). "'Positive'" results increase down the hierarchy of the sciences. *PLoS One, 5*(4), e10068.

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science, 7*(1), 45–52.

Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology, 57*, 153–169.

Fulton, B. R. (2018). Organizations and survey research: Implementing response enhancing strategies and conducting nonresponse analyses. *Sociological Methods & Research, 47*(2), 240–276.

Grand, J. A., Rogelberg, S. G., Banks, G. C., Landis, R. S., & Tonidandel, S. (2018). From outcome to process focus: Fostering a more robust psychological science through registered reports and results-blind reviewing. *Perspectives on Psychological Science, 13*(4), 448–456.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Sage Publications.

Hall, J., & Martin, B. R. (2019). Towards a taxonomy of research misconduct: The case of business school research. *Research Policy, 48*, 414–427.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *Journal of Finance, 72*(4), 1399–1440.

Hill, A. D., Bolton, J. F., & White, M. A. (2020). A call to find knowledge in our non-findings. *Strategic Organization, 18*(4), 645–654.

Hollenbeck, J. R., & Wright, P. M. (2017). Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management, 43*(1), 5–18.

Honig, B., Lampel, J., Siegel, D., & Drnevich, P. (2014). Ethics in the production and dissemination of management research: Institutional failure or individual fallibility? *Journal of Management Studies, 51*(1), 118–142.

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.

Kepes, S., Bennett, A. A., & McDaniel, M. A. (2014). Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. *Academy of Management Learning & Education, 13*(3), 446–466.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217.

Khan, M. J., & Trønnes, P. C. (2019). p-hacking in experimental audit research. *Behavioral Research in Accounting, 30*(1), 119–131.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Gulford Press.

Krishna, A., & Peter, S. M. (2018). Questionable research practices in student final theses − Prevalence, attitudes, and the role of the supervisor's perceived attitudes. *PLoS One, 13*(8), e0203470.

Latan, H. (2018). PLS path modeling in hospitality and tourism research: The golden age and days of future past. In F. Ali, S. M. Rasoolimanesh, & C. Cobanoglu (Eds.), *Applying partial least squares in tourism and hospitality research* (pp. 53–83). Bingley.

Latan, H., Chiappetta Jabbour, C. J., Lopes de Sousa Jabbour, A. B. (2021). To blow or not to blow the whistle: The role of rationalization in the perceived seriousness of threats and wrongdoing. *Journal of Business Ethics, 169*, 517–535.

Latan, H., Ringle, C. M., & Chiappetta Jabbour, C. J. (2018). Whistleblowing intentions among public accountants in Indonesia: Testing for the moderation effects. *Journal of Business Ethics, 152*, 573–588.

Leung, K. (2011). Presenting post hoc hypotheses as a priori: Ethical and theoretical issues. *Management and Organization Review, 7*(3), 471–479.

Linder, C., & Farahbakhsh, S. (2020). Unfolding the black box of questionable research practices: Where is the line between acceptable and unacceptable practices? *Business Ethics Quarterly, 30*(3), 335–360.

Lohmöller, J.-B. (1989). Basic principles of model building: Specification, estimation and evaluation. In H. Wold (Ed.), *Theoretical empiricism: A general rationale for scientific model-building* (pp. 1–25). Paragon House.

Macháček, V., & Srholec, M. (2021). Predatory publishing in Scopus: Evidence on cross-country diferences. *Scientometrics, 126*, 1897–1921.

Mazzola, J. J., & Deuling, J. K. (2013). Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I-O journal articles. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 6*(3), 279–284.

Miller, J. (2017). Hypothesis testing in the real world. *Educational and Psychological Measurement, 77*(4), 663–672.

Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., et al. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Journal of Personality and Social Psychology, 13*(1), 34–58.

Murphy, K. R., & Aguinis, H. (2019). HARKing: How badly can cherry-picking and question trolling produce bias in published results? *Journal of Business and Psychology, 34*, 1–17.

Necker, S. (2014). Scientific misbehavior in economics. *Research Policy, 43*, 1747–1759.

Noonan, R., & Wold, H. (1986). Partial least squares path analysis. *The International Encyclopedia of Education, 7*, 3769–3775.

Nosek, B. A., & Lakens, D. (2014). Editorial: Registered reports − A method to increase the credibility of published results. *Social Psychology, 45*(3), 137–141.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management, 43*(2), 376–399.

Olsson-Collentine, A., van Assen, M. A. L. M., & Hartgerink, C. H. J. (2019). The prevalence of marginally significant results in psychology over time. *Psychological Science, 30*(4), 576–586.

Pratt, T. C., Reisig, M. D., Holtfreter, K., & Golladay, K. A. (2019). Scholars' preferred solutions for research misconduct: Results from a survey of faculty members at America's top 100 research universities. *Ethics & Behavior, 29*(7), 510–530.

Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. *Psychological Science, 27*(7), 1036–1042.

Rabelo, A. L. A., Farias, J. E. M., Sarmet, M. M., Joaquim, T. C. R., Hoersting, R. C., Victorino, L., et al. (2020). Questionable research practices among Brazilian psychological researchers: Results from a replication study and an international comparison. *International Journal of Psychology, 55*(4), 674–683.

Rajah-Kanagasabai, C. J., & Roberts, L. D. (2015). Predicting self-reported research misconduct and questionable research practices in university students using an augmented Theory of Planned Behavior. *Frontiers in Psychology, 6*, 535.

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. In. Boenningstedt: SmartPLS GmbH.

Rubin, M. (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology, 21*, 308–320.

Savalei, V., & Dunn, E. (2015). Is the call to abandon p-values the red herring of the replicability crisis? *Frontiers in Psychology, 6*, 245.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17*(4), 551–566.

Schwartz, M. S. (2016). Ethical decision-making theory: An integrated approach. *Journal of Business Ethics, 139*(4), 755–776.

Shaw, J. D. (2017). From the editors: Advantages of starting with theory. *Academy of Management Journal, 60*(3), 819–822.

Sijtsma, K. (2016). Playing with data—Or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika, 81*, 1–15.

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.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681.

Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why has the number of scientific retractions increased? *PLoS One, 8*(7), e68397.

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*(285), 30–34.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49*(1), 108–112.

Stone, T. H., Jawahar, I. M., & Kisamore, J. L. (2009). Using the theory of planned behavior and cheating justifications to predict academic misconduct. *Career Development International, 14*(3), 221–241.

Stricker, J., & Günther, A. (2019). Scientific misconduct in psychology: A systematic review of prevalence estimates and new empirical data. *Zeitschrift Für Psychologie, 227*(1), 53–63.

Vogel, D., & Homberg, F. (2021). P-hacking, p-curves, and the PSM–performance relationship: Is there evidential value? *Public Administration Review, 81*(2), 191–20.

Wigboldus, D. H. J., & Dotsch, R. (2016). Encourage playing with data and discourage questionable reporting practices. *Psychometrika, 81*(1), 27–32.

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.

Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in Psychology, 9*, 1831.