



# Developing Surveys on Questionable Research Practices: Four Challenging Design Problems

Christian Berggren<sup>1</sup> · Bengt Gerdin<sup>2</sup> · Solmaz Filiz Karabag<sup>1,3</sup>

Accepted: 23 August 2024 / Published online: 2 September 2024  
© The Author(s) 2024

## Abstract

The exposure of scientific scandals and the increase of dubious research practices have generated a stream of studies on Questionable Research Practices (QRPs), such as failure to acknowledge co-authors, selective presentation of findings, or removal of data not supporting desired outcomes. In contrast to high-profile fraud cases, QRPs can be investigated using quantitative, survey-based methods. However, several design issues remain to be solved. This paper starts with a review of four problems in the QRP research: the problem of precision and prevalence, the problem of social desirability bias, the problem of incomplete coverage, and the problem of controversiality, sensitivity and missing responses. Various ways to handle these problems are discussed based on a case study of the design of a large, cross-field QRP survey in the social and medical sciences in Sweden. The paper describes the key steps in the design process, including technical and cognitive testing and repeated test versions to arrive at reliable survey items on the prevalence of QRPs and hypothesized associated factors in the organizational and normative environments. Partial solutions to the four problems are assessed, unresolved issues are discussed, and tradeoffs that resist simple solutions are articulated. The paper ends with a call for systematic comparisons of survey designs and item quality to build a much-needed cumulative knowledge trajectory in the field of integrity studies.

**Keywords** Questionable Research Practices · QRPs · Normative Environment · Organizational Climate · Survey Development · Design Problems · Problem of Incomplete Coverage · Survey Design Process · Baseline Survey · Pre-testing · Technical Assessment · Cognitive Interviews · Social Desirability · Sensitivity · Organizational Controversiality · Challenge of Nonresponse · Qualitative Research · Quantitative Research

## Introduction

The public revelations of research fraud and non-replicable findings (Berggren & Karabag, 2019; Levelt et al., 2012; Nosek et al., 2022) have created a lively interest in studying research integrity. Most studies in this field tend to focus on questionable research practices, QRPs, rather than blatant fraud, which is less common and hard to study with rigorous methods (Butler et al., 2017). Despite the significant contributions of this research about the incidence of QRPs in various countries and contexts, several issues still need to be addressed regarding the challenges of designing precise and valid survey instruments and achieving satisfactory response rates in this sensitive area. While studies in management (Hinkin, 1998; Lietz, 2010), behavioral sciences, psychology (Breakwell et al., 2020), sociology (Brenner, 2020), and education (Hill et al., 2022) have provided guidelines to design surveys, they rarely discuss how to develop, test, and use surveys targeting sensitive and controversial issues such as organizational or individual corruption (Lin & Yu, 2020), fraud (Lawlor et al., 2021), and misconduct. The aim of this study is to contribute to a systematic discussion of challenges facing survey designers in these areas and, by way of a detailed case study, highlight alternative ways to increase participation and reliability of surveys focusing on questionable research practices, scientific norms, and organizational climate.

The following section starts with a literature-based review of four important problems:

- the lack of conceptual consensus and precise measurements,
- the problem of social desirability bias.
- the difficulty of covering both quantitative and qualitative research fields.
- the problem of controversiality and sensitivity.

Section 3 presents an in-depth case study of developing and implementing a survey on QRPs in the social and medical sciences in Sweden 2018–2021, designed to target these problems. Its first results were presented in this journal (Karabag et al., 2024). The section also describes the development process and the survey content and highlights the general design challenges. Section 4 returns to the four problems by discussing partial solutions, difficult tradeoffs, and remaining issues.

## Four Design Problems in the Study of Questionable Research Practices

Extant QRP studies have generated an impressive body of knowledge regarding the occurrence and complexities of questionable practices, their increasing trend in several academic fields, and the difficulty of mitigating them with conventional interventions such as ethics courses and espousal of integrity policies (Gopalakrishna et al., 2022; Karabag et al., 2024; Necker, 2014). However, investigations on the prevalence of QRPs have so far lacked systematic problem analysis. Below, four main problems are discussed.

### The Problem of Conceptual Clarity and Measurement Precision

Studies of QRP prevalence in the literature exhibit high levels of questionable behaviors but also considerable variation in their estimates. This is illustrated in the examples below:

“42% had collected more data after inspecting whether results were statistically significant... and 51% had reported an unexpected finding as though it had been hypothesized from the start (HARKing)” (Fraser et al., 2018, p. 1),  
 “51,3% of respondents engaging frequently in at least one QRP” (Gopalakrishna et al., 2022, p. 1),  
 “...one third of the researchers stated that for the express purpose of supporting hypotheses with statistical significance they engaged in post hoc exclusion of data” (Banks et al., 2016, p. 10).

On a general level, QRPs constitute deviations from the responsible conduct of research, that are not severe enough to be defined as fraud and fabrication (Steneck, 2006). Within these borders, there is no conceptual consensus regarding specific forms of QRPs (Bruton et al., 2020; Xie et al., 2021). This has resulted in a considerable variation in prevalence estimates (Agnoli et al., 2017; Artino et al. Jr, 2019; Fiedler & Schwarz, 2016). Many studies emphasize the role of intentionality, implying a purpose to support a specific assertion with biased evidence (Banks et al., 2016). This tends to be backed by reports of malpractices in quantitative research, such as p-hacking or HARKing, where unexpected findings or results from an exploratory analysis are reported as having been predicted from the start (Andrade, 2021). Other QRP studies, however, build on another, often implicit conceptual definition and include practices that could instead be defined as sloppy or under-resourced research, e.g. insufficient attention to equipment, deficient supervision of junior co-workers, inadequate note-keeping of the research process, or use of inappropriate research designs (Gopalakrishna et al., 2022). Alternatively, those studies include behaviors such as “Fashion-determined choice of research topic”, “Instrumental and marketable approach”, and “Overselling methods, data or results” (Ravn & Sørensen, 2021, p. 30; Vermeulen & Hartmann, 2015) which may be opportunistic or survivalist but not necessarily involve intentions to mislead.

To shed light on the prevalence of QRPs in different environments, the first step is to conceptualize and delimit the practices to be considered. The next step is to operationalize the conceptual approach into useful indicators and, if needed, to reformulate and reword the indicators into unambiguous, easily understood items (Hinkin, 1995, 1998). The importance of careful item design has been demonstrated by Fiedler and Schwarz (2016). They show how the perceived QRP prevalence changes by adding specifications to well-known QRP items. Such specifications include: “*failing to report all dependent measures that are relevant for a finding*”, “*selectively reporting studies related to a specific finding that ‘worked’*” (Fiedler & Schwarz, 2016, p. 46, *italics in original*), or “*collecting more data after seeing whether results were significant in order to render non-significant results significant*” (Fiedler & Schwarz, 2016, p. 49, *italics in original*). These specifications demonstrate the importance of precision in item design, the need for item tests before applications in a large-scale survey, and as the case study in Sect. 3 indicates, the value of statistically analyzing the selected items post-implementation.

## The Problem of Social Desirability

Case studies of publicly exposed scientific misconduct have the advantage of explicitness and possible triangulation of sources (Berggren & Karabag, 2019; Huistra & Paul, 2022).

Opinions may be contradictory, but researchers/investigators may often approach a variety of stakeholders and compare oral statements with documents and other sources (Berggren & Karabag, 2019). By contrast, quantitative studies of QRPs need to rely on non-public sources in the form of statements and appraisals of survey respondents for the dependent variables and for potentially associated factors such as publication pressure, job insecurity, or competitive climate.

Many QRP surveys use items that target the respondents' personal attitudes and preferences regarding the dependent variables, indicating QRP prevalence, as well as the explanatory variables. This has the advantage that the respondents presumably know their own preferences and practices. A significant disadvantage, however, concerns social desirability, which in this context means the tendency of respondents to portray themselves, sometimes inadvertently, in more positive ways than justified by their behavior. The extent of this problem was indicated in a meta-study by Fanelli (2009), which demonstrated major differences between answers to sensitive survey questions that targeted the respondents' own behavior and questions that focused on the behavior of their colleagues. In the case study below, the pros and cons of the latter indirect approaches are analyzed.

### **The Problem of Covering Both Quantitative and Qualitative Research**

Studies of QRP prevalence are dominated by quantitative research approaches, where there exists a common understanding of the meaning of facts, proper procedures and scientific evidence. Several research fields, also in the social and medical sciences, include qualitative approaches — case studies, interpretive inquiries, or discourse analysis — where assessments of 'truth' and 'evidence' may be different or more complex to evaluate.

This does not mean that all qualitative endeavors are equal or that deceit—such as presenting fabricated interview quotes or referring to non-existent protocols—is accepted. However, while there are defined criteria for reporting qualitative research, such as the Consolidated Criteria for Reporting Qualitative Research (COREQ) (Tong et al., 2007) or the Standards for Reporting Qualitative Research (SRQR checklist) (O'Brien et al., 2014), the field of qualitative research encompasses a wide range of different approaches. This includes comparative case studies that offer detailed evidence to support their claims—such as the differences between British and Japanese factories (Dore, 1973/2011)—as well as discourse analyses and interpretive studies, where the concept of 'evidence' is more fluid and hard to apply. The generative richness of the analysis is a key component of their quality (Flick, 2013). This intra-field variation makes it hard to pin down and agree upon general QRP items to capture such behaviors in qualitative research. Some researchers have tried to interpret and report qualitative research by means of quantified methods (Ravn & Sørensen, 2021), but so far, these attempts constitute a marginal phenomenon. Consequently, the challenges of measuring the prevalence of QRPs (or similar issues) in the variegated field of qualitative research remain largely unexplored.

### **The Problem of Institutional Controversiality and Personal Sensitivity**

Science and academia depend on public trust for funding and executing research. This makes investigations of questionable behaviors a controversial issue for universities and may lead to institutional refusal/non-response. This resistance was experienced by the designers of a

large-scale survey of norms and practices in the Dutch academia when several universities decided not to take part, referring to the potential danger of negative publicity (de Vrieze, 2021). A Flemish survey on academic careers encountered similar participation problems (Aubert Bonn & Pinxten, 2019). Another study on universities' willingness to solicit whistleblowers for participation revealed that university officers, managers, and lawyers tend to feel obligated to protect their institution's reputation (Byrn et al., 2016). Such institutional actors may resist participation to avoid the exposure of potentially negative information about their institutions and management practices, which might damage the university's brand (Byrn et al., 2016; Downes, 2017).

QRP surveys involve sensitive and potentially intrusive questions also from a respondent's personal perspective that can lead to a reluctance to participate and non-response behavior (Roberts & John, 2014; Tourangeau & Yan, 2007). Studies show that willingness to participate declines for surveys covering sensitive issues such as misconduct, crime, and corruption, compared to less sensitive ones like leisure activities (cf. Tourangeau et al., 2010). The method of survey administration—whether face-to-face, over the phone, via the web, or paper-based—can influence the perceived sensitivity and response rate (Siewert & Udani, 2016; Szolnoki & Hoffmann, 2013). In the case study below, the survey did not require any institutional support. Instead, the designers focused on minimizing the individual sensitivity problem by avoiding questions about the respondents' personal practices. To manage this, they concentrated on their colleagues' behaviors (see Sect. 4.2). Even if a respondent agrees to participate, they may not answer the QRP items due to insufficient knowledge about her colleagues' practices or a lack of motivation to answer critical questions about their colleagues' practices (Beatty & Herrmann, 2002; Yan & Curtin, 2010). Additionally, a significant time gap between observing specific QRPs in the respondent's research environment and receiving the survey may make it difficult to recall and accurately respond to the questions. Such issues may also result in non-response problems.

## **Addressing the Problems: Case Study of a Cross-Field QRP Survey – Design Process, Survey Content, Design Challenges**

This section presents a case study of the way these four problems were addressed in a cross-field survey intended to capture QRP prevalence and associated factors across the social and medical sciences in Sweden. The account is based on the authors' intensive involvement in the design and analysis of the survey, including the technical and cognitive testing, and post-implementation analysis of item quality, missing responses, and open respondent comments. The theoretical background and the substantive results of the study are presented in a separate paper (Karabag et al., 2024). Method and language experts at Statistics Sweden, a government agency responsible for public statistics in Sweden, supported the testing procedures, the stratified respondent sampling and administered the survey roll-out.

### **The Survey Design Process – Repeated Testing and Prototyping**

The design process included four steps of testing, revising, and prototyping, which allowed the researchers to iteratively improve the survey and plan the roll-out.

### Step 1: Development of the Baseline Survey

This step involved searching the literature and creating a list of alternative constructs concerning the key concepts in the planned survey. Based on the study's aim, the first and third authors compared these constructs and examined how they had been itemized in the literature. After two rounds of discussions, they agreed on construct formulations and relevant ways to measure them, rephrased items if deemed necessary, and designed new items in areas where the extant literature did not provide any guidance. In this way, *Survey Version 1* was compiled.

### Step 2: Pre-Testing by Means of a Large Convenience Sample

In the second step, this survey version was reviewed by two experts in organizational behavior at Linköping University. This review led to minor adjustments and the creation of *Survey Version 2*, which was used for a major pretest. The aim was both to check the quality of individual items and to garner enough responses for a factor analysis that could be used to build a preliminary theoretical model. This dual aim required a larger sample than suggested in the literature on pretesting (Perneger et al., 2015). At the same time, it was essential to minimize the contamination of the planned target population in Sweden. To accomplish this, the authors used their access to a community of organization scholars to administer *Survey Version 2* to 200 European management researchers.

This mass pre-testing yielded 163 responses. The data were used to form preliminary factor structures and test a structural equation model. Feedback from a few of the respondents highlighted conceptual issues and duplicated questions. *Survey Version 3* was developed and prepared for detailed pretesting based on this feedback.

### Step 3: Focused Pre-Testing and Technical Assessment

This step focused on the pre-testing and technical assessment. The participants in this step's pretesting were ten researchers (six in the social sciences and four in the medical sciences) at five Swedish universities: Linköping, Uppsala, Gothenburg, Gävle, and Stockholm School of Economics. Five of those researchers mainly used qualitative research methods, two used both qualitative and quantitative methods, and three used quantitative methods. In addition, Statistics Sweden conducted a technical assessment of the survey items, focusing on wording, sequence, and response options.<sup>1</sup> Based on feedback from the ten pretest participants and the Statistics Sweden assessment, *Survey Version 4* was developed, translated into Swedish, and reviewed by two researchers with expertise in research ethics and scientific misconduct.

It should be highlighted that Swedish academia is predominantly bilingual. While most researchers have Swedish as their mother tongue, many are more proficient in English, and a minority have limited or no knowledge of Swedish. During the design process, the two language versions were compared item by item and slightly adjusted by skilled bilingual

<sup>1</sup> Wallenborg Likidis (2019). *Academic norms and scientific attitudes: Metrology Review of a survey for doctoral students, researchers and academic teachers (In Swedish: Akademiska normer och vetenskapliga förhållningssätt. Mätteknisk granskning av en enkät till doktorander, forskare och akademiska lärare)*. Prod. nr. 8,942,146, Statistics Sweden, Örebro.

researchers. This task was relatively straightforward since most items and concepts were derived from previously published literature in English. Notably, the Swedish versions of key terms and concepts have long been utilized within Swedish academia (see for example Berggren, 2016; Hasselberg, 2012). To secure translation quality, the language was controlled by a language expert at Statistics Sweden.

#### **Step 4: Cognitive Interviews by Survey and Measurement Experts**

Next, cognitive interviews (Willis, 2004) were organized with eight researchers from the social and medical sciences and conducted by an expert from Statistics Sweden (Wallenborg Likidis, 2019). The participants included four women and four men, ranging in age from 30 to 60. They were two doctoral students, two lecturers, and four professors, representing five different universities and colleges. Additionally, two participants had a non-Nordic background. To ensure confidentiality, no connections are provided between these characteristics and the individual participants.

An effort was made to achieve a distribution of gender, age, subject, employment, and institution. Four social science researchers primarily used qualitative research methods, while the remaining four employed qualitative and quantitative methods. Additionally, four respondents completed the Swedish version of the survey, and four completed the English version.

The respondents completed the survey in the presence of a methods expert from Statistics Sweden, who observed their entire response process. The expert noted spontaneous reactions and recorded instances where respondents hesitated or struggled to understand an item. After the survey, the expert conducted a structured interview with all eight participants, addressing details in each section of the survey, including the missive for recruiting respondents. Some respondents provided oral feedback while reading the cover letter and answering the questions, while others offered feedback during the subsequent interview.

During the cognitive interview process, the methods expert continuously communicated suggestions for improvements to the design team. A detailed test protocol confirmed that most items were sufficiently strong, although a few required minor modifications. The research team then finalized *Survey Version 5*, which included both English and Swedish versions (for the complete survey, see Supplementary Material S1).

Although the test successfully captured a diverse range of participants, it would have been desirable to conduct additional tests of the English survey with more non-Nordic participants; as it stands, only one such test was conducted. Despite the participants' different approaches to completing the survey, the estimated time to complete it was approximately 15–20 min. No significant time difference was observed between completing the survey in Swedish and English.

#### **Design Challenges – the Dearth of an Item-Specific Public Quality Discussion**

The design decision to employ survey items from the relevant literature as much as possible was motivated by a desire to increase comparability with previous studies of questionable research practices. However, this approach came with several challenges. Survey-based studies of QRPs rely on the respondents' subjective assessments, with no possibility to compare the answers with other sources. Thus, an open discussion of survey problems would

be highly valuable. However, although published studies usually present the items used in the surveys, there is seldom any analysis of the problems and tradeoffs involved when using a particular type of item or response format and meager information about item validity. Few studies, for example, contain any analysis that clarifies which items that measured the targeted variables with sufficient precision and which items that failed to do so.

Another challenge when using existing survey studies is the lack of information regarding the respondents' free-text comments about the survey's content and quality. This could be because the survey did not contain any open questions or because the authors of the report could not statistically analyze the answers. As seen below, however, open respondent feedback on a questionnaire involving sensitive or controversial aspects may provide important feedback regarding problems that did not surface during the pretest process, which by necessity targets much smaller samples.

## Survey Content

The survey started with questions about the respondent's current employment and research environment. It ended with background questions on the respondents' positions and the extent of their research activity, plus space for open comments about the survey. The core content of the survey consisted of sections on the organizational climate (15 items), scientific norms (13 items), good and questionable research practices (16 items), perceptions of fairness in the academic system (4 items), motivation for conducting research (8 items), ethics training and policies (5 items); and questions on the quality of the research environment and the respondent's perceived job security.

## Sample and Response Rate

All researchers, teachers, and Ph.D. students employed at Swedish universities are registered by Statistics Sweden. To ensure balanced representation and perspectives from both large universities and smaller university colleges, the institutions were divided into three strata based on the number of researchers, teachers, and Ph.D. students: more than 1,000 individuals (7 universities and university colleges), 500–999 individuals (3 institutions), and fewer than 500 individuals (29 institutions). From these strata, Statistics Sweden randomly sampled 35%, 45%, and 50% of the relevant employees, resulting in a sample of 10,047 individuals. After coverage analysis and exclusion of wrongly included, 9,626 individuals remained.

The selected individuals received a personal postal letter with a missive in both English and Swedish informing them about the project and the survey and notifying them that they could respond on paper or online. The online version provided the option to answer in either English or Swedish. The paper version was available only in English to reduce the cost of production and posting. The missive provided the recipients with comprehensive information about the study and what their involvement would entail. It emphasized the voluntary character of participation and their right to withdraw from the survey at any time, adding: *"If you do not want to answer the questions, we kindly ask you to contact us. Then you will not receive any reminders."* Sixty-three individuals used this decline option. In line with standard Statistics Sweden procedures, survey completion implied an agreement to participation and to the publication of anonymized results and indicated participants' understand-

ing of the terms provided (Duncan & Cheng, 2021). An email address was provided for respondents to request study outputs or for any other reason. The survey was open for data collection for two months, during which two reminders were sent to non-responders who had not opted out.

Once Statistics Sweden had collected the answers, they were anonymized and used to generate data files delivered to the authors. Statistics Sweden also provided anonymized information about age, gender, and type of employment of each respondent in the dataset delivered to the researchers. Of the targeted individuals, 3,295 responded, amounting to an overall response rate of 34.2%. An analysis of missing value patterns revealed that 290 of the respondents either lacked data for an entire factor or had too many missing values dispersed over several survey sections. After removing these 290 responses, we used SPSS algorithms (IBM-SPSS Statistics 27) to analyze the remaining missing values, which were randomly distributed and constituted less than 5% of the data. These values were replaced using the program's imputation program (Madley-Dowd et al., 2019). The final dataset consisted of 3,005 individuals, evenly distributed between female and male respondents (53,5% vs. 46,5%) and medical and social scientists (51,3% vs. 48,5%). An overview of the sample and the response rate is provided in Table 1, which can also be found in (Karabag et al., 2024). As shown in Table 1, the proportion of male and female respondents, as well as the proportion of respondents from medical and social science, and the age distribution of the respondents compared well with the original selection frame from Statistics Sweden.

## Revisiting the Four Problems. Partial Solutions and Remaining Issues

### Managing the Precision Problem - the Value of Factor Analyses

As noted above, the lack of conceptual consensus and standard ways to measure QRPs has resulted in a huge variation in estimated prevalence. In the case studied here, the purpose was to investigate deviations from research integrity and not low-quality research in general. This conceptual focus implied that selected survey items regarding QRP should build on the core aspect of intention, as suggested by Banks et al. (2016, p. 323): "design, analytic, or reporting practices that have been questioned because of the potential for the practice to be employed with the purpose of presenting biased evidence in favor of an assertion". After scrutinizing the literature, five items were selected as general indicators of QRP, irrespective of the research approach (see Table 2).

An analysis of the survey responses indicated that the general QRP indicators worked well in terms of understandability and precision. Considering the sensitive nature of the items, features that typically yield very high rates of missing data (Fanelli, 2009; Tourangeau & Yan, 2007), our missing rates of 11–21% must be considered modest. In addition, there were a few critical comments on the item formulation in the open response section at the end of the survey (see below).

Regarding the explanatory (independent) variables, the survey was inspired by studies showing the importance of the organizational climate and the normative environment within academia (Anderson et al., 2010). Organizational climate can be measured in several ways; the studied survey focused on items related to a collegial versus a competitive climate. The analysis of the normative environment was inspired by the classical norms of science articu-

lated by Robert Merton in his CUDOS framework: communism (communalism), universalism, disinterestedness, and organized skepticism (Merton, 1942/1973). This framework has been extensively discussed and challenged but remains a key reference (Anderson et al., 2010; Chalmers & Glasziou, 2009; Kim & Kim, 2018; Macfarlane & Cheng, 2008). Moreover, we were inspired by the late work of Merton on the ambivalence and ambiguities of scientists (Merton, 1942/1973), and the counter norms suggested by Mitroff (1974). Thus, the survey involved a composite set of items to capture the contradictory normative environment in academia: classical norms as well as their counter norms.

To reduce the problems of social desirability bias and personal sensitivity, the survey design avoided items about the respondent's personal adherence to explicit ideals, which are common in many surveys (Gopalakrishna et al., 2022). Instead, the studied survey focused on the normative preferences and attitudes within the respondent's environment. This necessitated the identification, selection, and refinement of 3–4 items for each potentially relevant norm/counter-norm. The selection process was used in previous studies of norm subscription in various research communities (Anderson et al., 2007; Braxton, 1993; Bray & von Storch, 2017). For the norm "skepticism", we consulted studies in the accounting literature of the three key elements of professional skepticism: questioning mind, suspension of judgment and search for knowledge (Hurt, 2010).

The first analytical step after receiving the completed survey set from Statistics Sweden was to conduct a set of factor analyses to assess the quality and validity of the survey items related to the normative environment and the organizational climate. These analyses suggested three clearly identifiable factors related to the normative environment: (1) a counter norm factor combining Mitroff's particularism and dogmatism ('Biasedness' in the further analysis), and two Mertonian factors: (2) Skepticism and (3) Openness, a variant of Merton's Communalism (see Table 3). A fourth Merton factor, Disinterestedness, could not be identified in our analysis.

The analytical process for organizational climate involved reducing the number of items from 15 to 11 (see Table 4). Here, the factor analysis suggested two clearly identifiable factors, one related to collegiality and the other related to competition (see Table 4). Overall, the factor analyses suggested that the design efforts had paid off in terms of high item quality, robust factor loadings, and a very limited need to remove any items.

In a parallel step, the open comments were assessed as an indication of how the study was perceived by the respondents (see Table 5). Of the 3005 respondents, 622 provided comprehensible comments, and many of them were extensive. 187 comments were related to the respondents' own employment/role, 120 were related to the respondents' working conditions and research environment, and 98 were related to the academic environment and atmosphere. Problems in knowing details of collegial practices were mentioned in 82 comments.

## Reducing Desirability Bias - the Challenge of Nonresponse

It is well established that studies on topics where the respondent has anything embarrassing or sensitive to report suffer from more missing responses than studies on neutral subjects and that respondents may edit the information they provide on sensitive topics (Tourangeau & Yan, 2007). Such a social desirability bias is applicable for QRP studies which explicitly target the respondents' personal attitudes and behaviors. To reduce this problem, the studied

**Table 1** Selection and characteristics of respondents

	Selected		Responded		Usable respondents	
	Number	Percent	Number	Percent	Number	Percent
Randomized	10047					
Excluded after analysis of coverage errors	421					
Had inquiry mailed	9626		3295	34.2	3005	31.2
The three strata						
Universities and university colleges with > 1,000 individuals	6184		2205	35.7		
Universities and university colleges with 500–999 individuals	963		312	32.4		
Universities and university colleges with less than 500 individuals	2470		778	31.5		
Not identified	9					
Dropouts						
No response	6142					
Declined participation	63					
Not retrieved	124					
Empty form	2					
Total dropouts	6331	65.8				
Sex						
Males	4216	43.8	1441	43.7	1334	44.4
Females	5410	56.2	1854	56.3	1671	55.6
	9626	100.0	3295	100.0	3005	100.0
Age						
-30 yrs.	906	9.4	276	8.4	248	8.3
31–40 yrs.	2749	28.6	933	28.3	866	28.8
41–50 yrs.	2566	26.7	846	25.7	769	25.6
51–60 yrs.	1953	20.3	658	20.0	600	20.0
61+ yrs.	1452	15.1	582	17.7	522	17.4
	9626	100.0	3295	100.0	3005	100.0
Research Field						
Medicine	5154	53.5	1685	51.1	1541	51.3

Table 1 (continued)

	Selected		Responded		Usable respondents	
	Number	Percent	Number	Percent	Number	Percent
Social science	4472	46.5	1610	48.9	1464	48.7
	9626	100.0	3295	100.0	3005	100.0

**Table 2** Five QRP items regarding general research practices

In my research environment, it happens that colleagues.....

---

...copy the work of others without providing appropriate references.

...selectively present findings that confirm their own opinions;

...copy the analyses and findings from their previous papers without providing appropriate references.

...are listed as co-authors even when they have not contributed to the research or to the writing.

...fail to acknowledge someone who has worked on a paper as a co-author.

---

**Table 3** Factor analysis of norm-related items

Items	Component		
Colleagues in my research environment...	1	2	3
... assess the work of other scientists primarily based on the status of the scientists.	0.83		
... assess a paper's findings based on the reputation of the journal in which it is published.	0.78		
... tend to think that only research that supports their own findings is good science.	0.73		
... hesitate to present findings that contradict their previous publications.	0.71		
... constantly use the same theoretical frameworks and references.	0.67		
... accept or reject new scientific results based on their intuition (gut feeling).	0.52		
... are skeptical about conclusions if they cannot find independent evidence.		0.78	
... are good at noticing inconsistencies in explanations.		0.64	
... consider and cite findings that challenge their own beliefs		0.43	
... value careful assessments by competent peers more than public praise.		0.44	
... find it important to inform external researchers investigating similar problems about their work in progress.			0.87
... readily invite external researchers to review their methods and results.			0.85
... argue that scientists must protect their findings to ensure priority in publishing.			0.55

Principal Components Analysis with Oblimin rotation;

Variance explained: 58%; Kaiser-Meyer-Olkin Measure of Sampling Adequacy=0.86

**Table 4** Factor analysis of organizational climate-related items

Items	Component	
Colleagues in my research environment...	1	2
...help each other to improve their research applications.	0.91	
...help each other solve publication problems, respond to critical reviewers, draft letters to editors.	0.87	
...receive valuable comments from each other.	0.80	
...are open to a critical discussion of their own ideas.	0.72	
...care about each other.	0.59	
...have little contact with each other.	0.56	
...treat their research areas as their own property		0.76
...have a hard time accepting criticism without becoming defensive.		0.69
...give all the attention to only a few colleagues.		0.68
...put their own interests first.		0.68
...compete with each other for research resources.		0.78

Principal Components Analysis with Oblimin rotation

Variance explained=62%; Kaiser-Meyer-Olkin Measure of Sampling Adequacy=0.93

**Table 5** Key themes in open-ended comments

Theme	Number	Percent of total
Details about own employment or role	187	30%
Details on working conditions or environment	120	19%
General or neutral on the study	98	16%
Difficulties responding to items	82	13%
General or neutral on ethics/data handling	43	7%
General on funding/competition	39	6%
Positive comments on the study	30	5%
Comment on qualitative part of the study	16	3%
Negative on the study/irrelevant	7	1%
Total number of responses	622	100%

survey applied a non-self-format focusing on the behaviors and preferences of the respondents' colleagues. Relevant survey items from published studies were rephrased from self-format designs to non-self-questions about practices in the respondent's environment, using the format: "In my research environment, colleagues..." followed by a five-step incremental response format from "(1) never" to "(5) always". In a similar way the survey avoided "should"-statements about ideal normative values: "Scientists and scholars should critically examine...". Instead, the survey used items intended to indicate the revealed preferences in the respondent's normative environment regarding universalism versus particularism or openness versus secrecy.

As indicated by Fanelli (2009), these redesign efforts probably reduced the social desirability bias significantly. At the same time, however, the redesign seemed to increase a problem not discussed by Fanelli (2009): an increased uncertainty problem related to the respondents' difficulties of knowing the practices of their colleagues in questionable areas. This issue was indicated by the open comment at the end of the studied survey, where 13% of the 622 respondents pointed out that they lacked sufficient knowledge about the behavior of their colleagues to answer the QRP questions (see Table 5). One respondent wrote:

*"It's difficult to answer questions about 'colleagues in my research area' because I don't have an insight into their research practices; I can only make informed guesses and generalizations. Therefore, I am forced to answer 'don't know' to a lot of questions".*

Regarding the questions on general QRPs, the rate of missing responses varied between 11% and 21%. As for the questions targeting specific QRP practices in quantitative and qualitative research, the rate of missing responses ranged from 38 to 49%. Unfortunately, the non-response alternative to these questions ("Don't know/not relevant") combined the two issues: the lack of knowledge and the lack of relevance. Thus, we don't know what part of the missing responses related to a non-presence of the specific research approach in the respondent's environment and what part signaled a lack of knowledge about collegial practices in this environment.

## Measuring QRPs in Qualitative Research - the Limited Role of Pretests

Studies of QRP prevalence focus on quantitative research approaches, where there exists a common understanding of the interpretation of scientific evidence, clearly recommended procedures, and established QRP items related to compliance with these procedures. In the heterogeneous field of qualitative research, there are several established standards for reporting the research (O'Brien et al., 2014; Tong et al., 2007), but, as noted above, hardly any commonly accepted survey items that capture behaviors that fulfill the criteria for QRPs. As a result, the studied survey project designed such items from the start during the survey development process. After technical and cognitive tests, four items were selected. See Table 6.

Despite the series of pretests, however, the first two of these items met severe criticism from a few respondents in the survey's open commentary section. Here, qualitative researchers argued that the items were unduly influenced by the truth claims in quantitative studies, whereas their research dealt with interpretation and discourse analysis. Thus, they rejected the items regarding selective usage of respondents and of interview quotes as indicators of questionable practices:

*"The alternative regarding using quotes is a bit misleading. Supporting your results by quotes is a way to strengthen credibility in a qualitative method...."*

*"The question about dubious practices is off target for us, who work with interpretation rather than solid truths. You can present new interpretations, but normally that does not imply that previous 'findings' should be considered incorrect."*

*"The questions regarding qualitative research were somewhat irrelevant. Often this research is not guided by a given hypothesis, and researchers may use a convenient sample without this resulting in lower quality."*

One comment focused on other problems related to qualitative research:

*"Several questions do not quite capture the ethical dilemmas we wrestle with. For example, is the issue of dishonesty and 'inaccuracies' a little misplaced for us who work with interpretation? ...At the same time, we have a lot of ethical discussions, which, for example, deal with power relations between researchers and 'researched', participant observation/informal contacts and informed consent (rather than patients participating in a study)".*

Unfortunately, the survey received these comments and criticism only after the full-scale rollout and not during the pretest rounds. Thus, we had no chance to replace the contested items with other formulations or contemplate a differentiation of the subsection to target

**Table 6** QRP questions regarding qualitative research practices

---

In my research environment, it happens that colleagues...

---

select respondents to support their own hypotheses.

only use interview quotes or other data that confirm the results they want to present.

allege that a paper is based on methods, such as observations or independent coding, without really using these methods.

exaggerate the number of interviews or observations that a paper is based on.

---

specific types of qualitative research with appropriate questions. Instead, we had to limit the post-roll-out survey analysis to the last two items in Table 6, although they captured devious behaviors rather than gray zone practices.

Why then was this criticism of QRP items related to qualitative research not exposed in the pretest phase? This is a relevant question, also for future survey designers. An intuitive answer could be that the research team only involved quantitative researchers. However, as highlighted above, the pretest participants varied in their research methods: some exclusively used qualitative methods, others employed mixed methods, and some utilized quantitative methods. This diversity suggests that the selection of test participants was appropriate. Moreover, all three members of the research team had experience of both quantitative and qualitative studies. However, as discussed above, the field of qualitative research involves several different types of research, with different goals and methods – from detailed case studies grounded in original empirical fieldwork to participant observations of complex organizational phenomena to discursive re-interpretations of previous studies. Of the 3,005 respondents who answered the survey in a satisfactory way, only 16 respondents, or 0,5%, had any critical comments about the QRP items related to qualitative research. A failure to capture the objections from such a small proportion in a pretest phase is hardly surprising. The general problem could be compared with the challenge of detecting negative side-effects in drug development. Although the pharmaceutical firms conduct large-scale tests of candidate drugs before government approval, doctors nevertheless detect new side-effects when the medicine is rolled out to significantly more people than the test populations – and report these less frequent problems in the additional drug information (Galeano et al., 2020; McNeil et al., 2010).

In the social sciences, the purpose of pre-testing is to identify problems related to ambiguities and bias in item formulation and survey format and initiate a search for relevant solutions. A pre-test on a small, selected subsample cannot guarantee that all respondent problems during the full-scale data collection will be detected. The pretest aims to reduce errors to acceptable levels and ensure that the respondents will understand the language and terminology chosen. Pretesting in survey development is also essential to help the researchers to assess the overall flow and structure of the survey, and to make necessary adjustments to enhance respondent engagement and data quality (Ikart, 2019; Presser & Blair, 1994).

In our view, more pretests would hardly solve the epistemological challenge of formulating generally acceptable QRP items for qualitative research. The open comments studied here suggest that there is no one-size-fits-all solution. If this is right, the problem should rather be reformulated to a question of identifying different strands of qualitative research with diverse views of integrity and evidence which need to be measured with different measures. To address this challenge in a comprehensive way, however, goes far beyond the current study.

### **Controversiality and Collegial sensitivity - the Challenge of Predicting Nonresponse**

Studies of research integrity, questionable research practices, and misconduct in science tend to be organizationally controversial and personally sensitive. If university leaders are asked to support such studies, there is a considerable risk that the answer will be negative. In the case studied here, the survey roll-out was not dependent on any active organizational participation since Statistics Sweden possessed all relevant respondent information

in-house. This, we assumed, would take the controversiality problem off the agenda. Our belief was supported by the non-existent complaints regarding a potential negativity bias from the pretest participants. Instead, the problem surfaced when the survey was rolled out, and all the respondents contemplated the survey. The open comment section at the end of the survey provided insights into this reception.

Many respondents provided positive feedback, reflected in 30 different comments such as:

*"Thank you for doing this survey. I really hope it will lead to changes because it is needed".*

*"This is an important survey. However, there are conflicting norms, such as those you cite in the survey, /concerning/ for example, data protection. How are researchers supposed to be open when we cannot share data for re-analysis?"*

*"I am glad that the problems with egoism and non-collegiality are addressed in this manner".*

Several of them asked for more critical questions regarding power, self-interest, and leadership:

*"What I lack in the survey were items regarding academic leadership. Otherwise, I am happy that someone is doing research on these issues".*

*"A good survey but needs to be complemented with questions regarding researchers who put their commercial interests above research and exploit academic grants for commercial purposes".*

A small minority criticized the survey for being overly negative towards academia:

*"A major part of the survey feels very negative and /conveys/ the impression that you have a strong pre-understanding of academia as a horrible environments".*

*"Some of the questions are uncomfortable and downright suggestive. Why such a negative attitude towards research?"*

*"The questions have a tendency to make us /the respondents/ informers. An unpleasant feeling when you are supposed to lay information against your university".*

*"Many questions are hard to answer, and I feel that they measure my degree of suspicion against my closest colleagues and their motivation ... Several questions I did not want to answer since they contain a negative interpretation of behaviors which I don't consider as automatically negative".*

A few of these respondents stated that they abstained from answering some of the 'negative questions', since they did not want to report on or slander their colleagues. The general impact is hard to assess. Only 20% of the respondents offered open survey comments, and only seven argued that questions were "negative". The small number explains why the issue of negativity did not show up during the testing process. However, a perceived sense of negativity may have affected the willingness to answer among more respondents than those who provided free test comments.

## Conclusion - The Needs for a Cumulative Knowledge Trajectory in Integrity Studies

In the broad field of research integrity studies, investigations of QRPs in different contexts and countries play an important role. The comparability of the results, however, depends on the conceptual focus of the survey design and the quality of the survey items. This paper starts with a discussion of four common problems in QRP research: the problems of precision, social desirability, incomplete coverage, and organizational controversiality and sensitivity. This is followed by a case study of how these problems were addressed in a detailed survey design process. An assessment of the solutions employed in the studied survey design reveals progress as well as unresolved issues.

Overall, the paper shows that the problem and challenges of precision could be effectively managed through explicit conceptual definitions and careful item design.

The problem of social desirability bias was probably reduced by means of a non-self-response format referring to preferences and behaviors among colleagues instead of personal behaviors. However, an investigation of open respondent comments indicated that the reduced risk of social bias came at the expense of higher uncertainty due to the respondents' lack of insight in the concrete practices of their colleagues.

The problem of incomplete coverage of QRPs in qualitative research, the authors initially linked to "the lack of standard items" to capture QRPs in qualitative studies. Open comments at the end of the survey, however, suggested that the lack of such standards would not be easily managed by the design of new items. Rather, it seems to be an epistemological challenge related to the multifarious nature of the qualitative research field, where the understanding of 'evidence' is unproblematic in some qualitative sub-fields but contested in others. This conjecture and other possible explanations will hopefully be addressed in forthcoming epistemological and empirical studies.

Regarding the problem of controversiality and sensitivity, previous studies show that QRP research is a controversial and sensitive area for academic executives and university brand managers. The case study discussed here indicates that this is a sensitive subject also for rank-and-file researchers who may hesitate to answer, even when the questions do not target the respondents' own practices but the practices and preferences of their colleagues. Future survey designers may need to engage in framing, presenting, and balancing sensitive items to reduce respondent suspicions and minimize the rate of missing responses. Reflections on the case indicate that this is doable but requires thoughtful design, as well as repeated tests, including feedback from a broad selection of prospective participants.

In conclusion, the paper suggests that more resources should be spent on the systematic evaluation of different survey designs and item formulations. In the long term, such investments in method development will yield a higher proportion of robust and comparable studies. This would mitigate the problems discussed here and contribute to the creation of a much-needed cumulative knowledge trajectory in research integrity studies.

An issue not covered here is that surveys, however finely developed, only give quantitative information about patterns, behaviors, and structures. An understanding of underlying thoughts and perspectives requires other procedures. Thus, methods that integrate and triangulate qualitative and quantitative data — known as mixed methods (Karabag & Berggren, 2016; Ordu & Yilmaz, 2024; Smajic et al., 2022)— may give a deeper and more complete picture of the phenomenon of QRP.

**Acknowledgements** We thank Jennica Wallenberg Likidis, Statistics Sweden, for providing expert support in the survey design. We are grateful to colleagues Ingrid Johansson Mignon, Cecilia Enberg, Anna Dreber Almenberg, Andrea Fried, Sara Liin, Mariano Salazar, Lars Bengtsson, Harriet Wallberg, Karl Wennberg, and Thomas Magnusson, who joined the pretest or cognitive tests. We also thank Ksenia Onufrey, Peter Hedström, Jan-Ingvar Jönsson, Richard Öhrvall, Kerstin Sahlin, and David Ludvigsson for constructive comments or suggestions.

**Author contributions** Conceptualization: CB. Survey Design: SFK, CB. Methodology: SFK, BG, CB. Visualization: SFK, BG. Funding acquisition: SFK. Project administration and management: SFK. Writing – original draft: CB. Writing – review & editing: CB, BG, SFK. Approval of the final manuscript: SFK, BG, CB.

**Funding** Open access funding provided by Linköping University. Swedish Forte: Research Council for Health, Working Life and Welfare ([https://www.vr.se/swecris?#/project/2018-00321\\_Forte](https://www.vr.se/swecris?#/project/2018-00321_Forte)) Grant No. 2018-00321.

Open access funding provided by Linköping University.

**Data Availability** The data supporting the findings of this study are available from the corresponding author, upon reasonable request.

## Declarations

**Ethics approval and consent to participate** The Swedish Act concerning the Ethical Review of Research Involving Humans (2003:460) defines the type of studies which requires an ethics approval. In line with the General Data Protection Regulation (EU 2016/67), the act is applicable for studies that collect personal data that reveal racial or ethnic origin, political opinions, trade union membership, religious or philosophical beliefs, or health and sexual orientation. The present study does not involve any of the above, why no formal ethical permit was required. The ethical aspects of the project and its compliance with the guidelines of the Swedish Research Council (2017) were also part of the review process at the project's public funding agency Forte.

**Competing Interests** The authors declare that they have no competing interests.

**Supporting Information** The complete case study survey of social and medical science researchers in Sweden 2020.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS One*, 12(3), e0172792.
- Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics*, 13, 437–461.
- Anderson, M. S., Ronning, E. A., Devries, R., & Martinson, B. C. (2010). Extending the Mertonian norms: Scientists' subscription to norms of Research. *The Journal of Higher Education*, 81(3), 366–393. <https://doi.org/10.1353/jhe.0.0095>
- Andrade, C. (2021). HARKing, cherry-picking, p-hacking, fishing expeditions, and data dredging and mining as questionable research practices. *The Journal of Clinical Psychiatry*, 82(1), 25941.

- Artino Jr, A. R., Driessen, E. W., & Maggio, L. A. (2019). Ethical shades of gray: International frequency of scientific misconduct and questionable research practices in health professions education. *Academic Medicine*, *94*(1), 76–84.
- Aubert Bonn, N., & Pinxten, W. (2019). A decade of empirical research on research integrity: What have we (not) looked at? *Journal of Empirical Research on Human Research Ethics*, *14*(4), 338–352.
- Banks, G. C., O'Boyle Jr, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., & Adkins, C. L. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, *42*(1), 5–20.
- Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. *Survey Nonresponse*, *71*, 86.
- Berggren, C. (2016). *Scientific Publishing: History, practice, and ethics (in Swedish: Vetenskaplig Publicering: Historik, Praktik Och Etik)*. Studentlitteratur AB.
- Berggren, C., & Karabag, S. F. (2019). Scientific misconduct at an elite medical institute: The role of competing institutional logics and fragmented control. *Research Policy*, *48*(2), 428–443. <https://doi.org/10.1016/j.respol.2018.03.020>
- Braxton, J. M. (1993). Deviancy from the norms of science: The effects of anomie and alienation in the academic profession. *Research in Higher Education*, *54*(2), 213–228. <https://www.jstor.org/stable/40196105>
- Bray, D., & von Storch, H. (2017). The normative orientations of climate scientists. *Science and Engineering Ethics*, *23*(5), 1351–1367.
- Breakwell, G. M., Wright, D. B., & Barnett, J. (2020). Research questions, design, strategy and choice of methods. *Research Methods in Psychology*, 1–30.
- Brenner, P. S. (2020). Why survey methodology needs sociology and why sociology needs survey methodology: Introduction to understanding survey methodology: Sociological theory and applications. In *Understanding survey methodology: Sociological theory and applications* (pp. 1–11). [https://doi.org/10.1007/978-3-030-47256-6\\_1](https://doi.org/10.1007/978-3-030-47256-6_1)
- Bruton, S. V., Medlin, M., Brown, M., & Sacco, D. F. (2020). Personal motivations and systemic incentives: Scientists on questionable research practices. *Science and Engineering Ethics*, *26*(3), 1531–1547.
- 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.
- Byrn, M. J., Redman, B. K., & Merz, J. F. (2016). A pilot study of universities' willingness to solicit whistleblowers for participation in a study. *AJOB Empirical Bioethics*, *7*(4), 260–264.
- Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89.
- de Vrieze, J. (2021). Large survey finds questionable research practices are common. *Science*. <https://doi.org/10.1126/science.373.6552.265>
- Dore, R. P. (1973/2011). *British Factory Japanese Factory: The origins of National Diversity in Industrial Relations, with a New Afterword*. University of California Press/Routledge.
- Downes, M. (2017). University scandal, reputation and governance. *International Journal for Educational Integrity*, *13*, 1–20.
- Duncan, L. J., & Cheng, K. F. (2021). Public perception of NHS general practice during the first six months of the COVID-19 pandemic in England. *F1000Research*, *10*.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, *4*(5), e5738.
- Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, *7*(1), 45–52.
- Flick, U. (2013). *The SAGE Handbook of Qualitative Data Analysis*. sage.
- Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS One*, *13*(7), e0200303.
- Galeano, D., Li, S., Gerstein, M., & Paccanaro, A. (2020). Predicting the frequencies of drug side effects. *Nature Communications*, *11*(1), 4575.
- Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in the Netherlands. *PLoS One*, *17*(2), e0263023.
- Hasselberg, Y. (2012). *Science as Work: Norms and Work Organization in Commodified Science (in Swedish: Vetenskap Som arbete: Normer och arbetsorganisation i den kommodifierade vetenskapen)*. Gidlunds förlag.
- Hill, J., Ogle, K., Gottlieb, M., Santen, S. A., & Artino Jr, A. R. (2022). Educator's blueprint: a how-to guide for collecting validity evidence in survey-based research. *AEM Education and Training*, *6*(6), e10835.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*(5), 967–988.

- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104–121.
- Huistra, P., & Paul, H. (2022). Systemic explanations of scientific misconduct: Provoked by spectacular cases of norm violation? *Journal of Academic Ethics*, 20(1), 51–65.
- Hurt, R. K. (2010). Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice & Theory*, 29(1), 149–171.
- Ikart, E. M. (2019). Survey questionnaire survey pretesting method: An evaluation of survey questionnaire via expert reviews technique. *Asian Journal of Social Science Studies*, 4(2), 1.
- Karabag, S. F., & Berggren, C. (2016). Misconduct, marginality and editorial practices in management, business and economics journals. *PLoS One*, 11(7), e0159492. <https://doi.org/10.1371/journal.pone.0159492>
- Karabag, S. F., Berggren, C., Pielaszkievicz, J., & Gerdin, B. (2024). Minimizing questionable research practices—the role of norms, counter norms, and micro-organizational ethics discussion. *Journal of Academic Ethics*, 1–27. <https://doi.org/10.1007/s10805-024-09520-z>
- Kim, S. Y., & Kim, Y. (2018). The ethos of Science and its correlates: An empirical analysis of scientists' endorsement of Mertonian norms. *Science Technology and Society*, 23(1), 1–24. <https://doi.org/10.1177/0971721817744438>
- Lawlor, J., Thomas, C., Guhin, A. T., Kenyon, K., Lerner, M. D., Consortium, U., & Drahota, A. (2021). Suspicious and fraudulent online survey participation: Introducing the REAL framework. *Methodological Innovations*, 14(3), 20597991211050467.
- Levelt, W. J., Drenth, P., & Noort, E. (2012). *Flawed science: The fraudulent research practices of social psychologist Diederik Stapel (in Dutch: Falende wetenschap: De frauduleuze onderzoekspraktijken van social-psycholoog Diederik Stapel)*. Commissioned by the Tilburg University, University of Amsterdam and the University of Groningen. <https://doi.org/http://hdl.handle.net/11858/00-001M-0000-0010-258A-9>
- Lietz, P. (2010). Research into questionnaire design: A summary of the literature. *International Journal of Market Research*, 52(2), 249–272.
- Lin, M. W., & Yu, C. (2020). Can corruption be measured? Comparing global versus local perceptions of corruption in East and Southeast Asia. In *Regional comparisons in comparative policy analysis studies* (pp. 90–107). Routledge.
- Macfarlane, B., & Cheng, M. (2008). Communism, universalism and disinterestedness: Re-examining contemporary support among academics for Merton's scientific norms. *Journal of Academic Ethics*, 6, 67–78.
- Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, 110, 63–73.
- McNeil, J. J., Piccenna, L., Ronaldson, K., & Ioannides-Demos, L. L. (2010). The value of patient-centred registries in phase IV drug surveillance. *Pharmaceutical Medicine*, 24, 281–288.
- Merton, R. K. (1942/1973). The normative structure of science. In *The sociology of science: Theoretical and empirical investigations*. The University of Chicago Press.
- Mitroff, I. I. (1974). Norms and counter-norms in a select group of the Apollo Moon scientists: A case study of the ambivalence of scientists. *American Sociological Review*, 39(4), 579–595. <https://doi.org/10.2307/2094423>
- Necker, S. (2014). Scientific misbehavior in economics. *Research Policy*, 43(10), 1747–1759. <https://doi.org/10.1016/j.respol.2014.05.002>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., & Nuijten, M. B. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719–748.
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, 89(9). [https://journals.lww.com/academicmedicine/fulltext/2014/09000/standards\\_for\\_reporting\\_qualitative\\_research\\_a.21.aspx](https://journals.lww.com/academicmedicine/fulltext/2014/09000/standards_for_reporting_qualitative_research_a.21.aspx)
- Ordu, Y., & Yılmaz, S. (2024). Examining the impact of dramatization simulation on nursing students' ethical attitudes: A mixed-method study. *Journal of Academic Ethics*, 1–13.
- Perneger, T. V., Courvoisier, D. S., Hudelson, P. M., & Gayet-Ageron, A. (2015). Sample size for pre-tests of questionnaires. *Quality of life Research*, 24, 147–151.
- Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? *Sociological Methodology*, 73–104.
- Ravn, T., & Sørensen, M. P. (2021). Exploring the gray area: Similarities and differences in questionable research practices (QRPs) across main areas of research. *Science and Engineering Ethics*, 27(4), 40.
- Roberts, D. L., & John, F. A. S. (2014). Estimating the prevalence of researcher misconduct: a study of UK academics within biological sciences. *PeerJ*, 2, e562.
- Siewert, W., & Udani, A. (2016). Missouri municipal ethics survey: Do ethics measures work at the municipal level? *Public Integrity*, 18(3), 269–289.

- Smajic, E., Avdic, D., Pasic, A., Prcic, A., & Stancic, M. (2022). Mixed methodology of scientific research in healthcare. *Acta Informatica Medica*, 30(1), 57–60. <https://doi.org/10.5455/aim.2022.30.57-60>
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics*, 12, 53–74.
- Szolnoki, G., & Hoffmann, D. (2013). Online, face-to-face and telephone surveys—comparing different sampling methods in wine consumer research. *Wine Economics and Policy*, 2(2), 57–66.
- Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), 349–357. <https://doi.org/10.1093/intqhc/mzm042>
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859.
- Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly*, 74(3), 413–432.
- Vermeulen, I., & Hartmann, T. (2015). Questionable research and publication practices in communication science. *Communication Methods and Measures*, 9(4), 189–192.
- Wallenborg Likidis, J. (2019). *Academic norms and scientific attitudes: Metrology review of a survey for doctoral students, researchers and academic teachers (In Swedish: Akademiska normer och vetenskapliga förhållningssätt. Mätteknisk granskning av en enkät till doktorander, forskare och akademiska lärare)*. Prod.nr. 8942146, Statistics Sweden, Örebro.
- Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.
- Xie, Y., Wang, K., & Kong, Y. (2021). Prevalence of research misconduct and questionable research practices: A systematic review and meta-analysis. *Science and Engineering Ethics*, 27(4), 41.
- Yan, T., & Curtin, R. (2010). The relation between unit nonresponse and item nonresponse: A response continuum perspective. *International Journal of Public Opinion Research*, 22(4), 535–551.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Christian Berggren<sup>1</sup>  · Bengt Gerdin<sup>2</sup>  · Solmaz Filiz Karabag<sup>1,3</sup> 

✉ Solmaz Filiz Karabag  
solmaz.filiz.karabag@liu.se

Christian Berggren  
christian.berggren@liu.se

Bengt Gerdin  
bengt.gerdin@uu.se

<sup>1</sup> Department of Management and Engineering [IEI], Linköping University, Linköping SE-581 83, Sweden

<sup>2</sup> Department of Surgical Sciences, Uppsala University, Uppsala University Hospital, entrance 70, Uppsala SE-751 85, Sweden

<sup>3</sup> Department of Civil and Industrial Engineering, Uppsala University, Box 169, Uppsala SE-751 04, Sweden