Ethical Shades of Gray: International Frequency of Scientific Misconduct and Questionable Research Practices in Health Professions Education

Anthony R. Artino Jr, PhD, Erik W. Driessen, PhD, and Lauren A. Maggio, PhD

Abstract

Purpose
To maintain scientific integrity and engender public confidence, research must be conducted responsibly. Whereas deliberate scientific misconduct such as data fabrication is clearly unethical, other behaviors—often referred to as questionable research practices (QRPs)—exploit the ethical shades of gray that color acceptable practice. This study aimed to measure the frequency of self-reported misconduct and QRPs in a diverse, international sample of health professions education (HPE) researchers.

Method
In 2017, the authors conducted an anonymous, cross-sectional survey study. The web-based survey contained 43 items that asked respondents to rate how often they had engaged in a variety of irresponsible research behaviors. The items were adapted from previously published surveys.

Results
In total, 590 HPE researchers took the survey. The mean age was 46 years (SD = 11.6), and the majority of participants were from the United States (26.4%), Europe (23.2%), and Canada (15.3%). The three most frequently reported irresponsible research behaviors were adding authors who did not qualify for authorship (60.6%), citing articles that were not read (49.5%), and selectively citing papers to please editors or reviewers (49.4%). Additionally, respondents reported misrepresenting a participant's words (6.7%), plagiarizing (5.5%), inappropriately modifying results (5.3%), deleting data without disclosure (3.4%), and fabricating data (2.4%). Overall, 533 (90.3%) respondents reported at least one irresponsible behavior.

Conclusions
Notwithstanding the methodological limitations of survey research, these findings indicate that a substantial proportion of HPE researchers report a range of misconduct and QRPs. Consequently, reforms may be needed to improve the conduct of HPE research.

Downloaded from https://journals.lww.com/academicmedicine by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 03/22/2019

The responsible conduct of research is the foundation of sound scientific practice.[1,2] The need to conduct research in a responsible manner is self-evident—if science is to inform our understanding of how the world works, it must be done in an honest, accurate, and unbiased way.[3]

Irresponsible research behaviors are typically divided into two broad categories: deliberate scientific misconduct (i.e., fabrication, falsification, and plagiarism) and questionable research practices (QRPs).[1] Whereas behaviors like data fabrication are clearly unethical, QRPs exploit the ethical shades of gray that color acceptable research practice and "offer considerable latitude for rationalization and self-deception."[4] Consequently, QRPs are more prevalent and, many have argued, more damaging to science and its public reputation than obvious fraud.[4–8] Taken together, scientific misconduct and QRPs can waste resources; provide an unfair advantage to some researchers over others; damage the scientific record; and provide a poor example for other researchers, especially trainees.[7]

Health professions education (HPE) is not immune to the damaging effects of irresponsible research practices. In HPE, QRPs have been defined as poor data management, inappropriate research procedures, insufficient respect for study participants, improper research design, carelessness in observation and analysis, unsuitable authorship or publishing practices, and derelictions in reviewing and editing.[7] The need to guard against both deliberate scientific misconduct and QRPs is frequently described in the author instructions for scientific and health professions journals, and such guidelines are often patterned on the recommendations of the Committee on Publication Ethics.[9]

Commentaries by HPE journal editors have highlighted instances of QRPs in article submissions, including self-plagiarism, "salami slicing" (i.e., inappropriately dividing a single study into multiple papers), and unethical authorship practices.[10–12] Additionally, a review of four HPE journals found that 13% of original research articles published in 2013 did not address approval by an ethics review board or stated that it was unnecessary, without further discussion.[13] Moreover, a 2017 study of HPE leaders highlighted multiple problematic authorship practices, including honorary authorship and the exclusion of authors who deserved authorship.[14] Notably, only about half of the senior researchers surveyed were

Please see the end of this article for information about the authors.

Correspondence should be addressed to Anthony R. Artino Jr, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814; telephone: (301) 295-3693; e-mail: anthony.artino@usuhs.edu; Twitter: @mededdoc.

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.

Acad Med. 2019;94:76–84.
First published online August 14, 2018
doi: 10.1097/ACM.0000000000002412

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A584.

able to correctly identify the authorship standards used by most medical journals (i.e., the International Committee of Medical Journal Editors [ICMJE] authorship criteria[15]).

Notwithstanding these examples, scientific misconduct and QRPs have received limited attention in the HPE literature. In a recent article,[7] we attempted to raise the community's awareness of these issues and highlight the need to examine their pervasiveness among HPE researchers. Accordingly, we conducted this study to measure the frequency of self-reported misconduct and QRPs in a diverse, international sample of HPE researchers. In doing so, we hope to continue the conversation about responsible research in HPE, with the ultimate goal of promoting high-quality, ethical research.

Method

Context

To measure the frequency of serious research misconduct and other QRPs, many different approaches have been employed. These include counts of confirmed cases of researcher fraud and paper retractions, as well research audits by government funders.[8] Such methods are limited because they are calculated on the basis of misconduct that has been discovered, and detecting such misconduct is difficult.[16] Moreover, distinguishing intentional misconduct from honest mistakes is challenging. Therefore, such approaches significantly underestimate the real frequency of misconduct and QRPs because only researchers know if they have willfully acted in an unethical or questionable manner.

To address these challenges, survey methods have been used to directly ask scientists about their research behaviors.[4–6,17,18] Like the measurement of any socially undesirable behavior, assessing irresponsible research practices via self-report likely underestimates the true prevalence or frequency of the behaviors. Nonetheless, when employed appropriately, survey methods can generate reasonable estimates that provide a general sense of the problem's scope.[19,20]

We administered an anonymous, cross-sectional survey to determine the frequency of research misconduct and other QRPs in a sample of HPE researchers. The Ethical Review Board Committee of the Netherlands Association for Medical Education approved this study (dossier no. 937).

Survey development

We developed the survey by adapting several existing tools. The final survey featured a total of 66 items divided into three sections (see Supplemental Digital Appendix 1, available at http://links.lww.com/ACADMED/A584). The first section included 43 items derived from two previously published surveys assessing research misconduct and QRPs in biomedicine.[5,6] The items asked respondents to identify how often they had engaged in the particular research practice. The practices spanned the research continuum, from data collection and storage to study reporting, collaboration, and authorship. The items employed a six-point, Likert-type, frequency-response scale: "never," "once," "occasionally," "sometimes," "frequently," and "almost always." Each item also included the response option "not applicable to my work."

We slightly modified the original misconduct and QRP items to improve their clarity and relevance to the HPE. For example, the original item "inadequately handled or stored data or (bio)materials" was revised to "inappropriately stored sensitive research data (e.g., data that contains personally identifiable information)." Following these modifications, 19 experienced HPE researchers reviewed the adapted survey items and provided detailed qualitative feedback.[21] Ten of the expert reviewers were women, and all reviewers held doctoral degrees (13 PhDs, 4 MD/PhDs, and 2 MDs). On average, the reviewers had published 98.9 journal articles (SD = 66.1) indexed in PubMed. On the basis of the date of their first HPE publication, they had been publishing in the field for an average of 20.7 years (SD = 9.3). Expert reviewers reported their work location as the United States (n = 9), Europe (n = 4), Canada (n = 3), South America (n = 1), Africa (n = 1), and Australia (n = 1), and all but 2 were identified as full professors.

The expert feedback included comments on item relevance, clarity, missing facets, and suggestions for overall survey improvement. On the basis of the expert feedback, we revised the survey again; revisions included wording modifications and the development of several new HPE-specific items. For example, on the basis of several recommendations, we created the following items related to qualitative research methods, among several others: "misrepresented a participant's words or writings" and "claimed you used a particular qualitative research approach appropriately (e.g., grounded theory) when you knowingly did not."

The second section of the survey included 9 publication pressure items[17] (these data are not reported here), and the final section included 13 demographic items.

Sampling and survey distribution procedures

To create our sample, we used two separate approaches. First, we created a "curated sample" by searching Web of Science for articles in HPE journals published in 2015 and 2016 (see Supplemental Digital Appendix 2, available at http://links.lww.com/ACADMED/A584, for a list of journals). Journals were selected if they focused on general HPE topics and were indexed in Web of Science. To increase international representation, we searched three additional regional databases—Scielo (a database focused on South America), African Journals Online, and Asia Journals Online—for articles in other HPE journals. We did not limit our search to English-language titles but did restrict retrieval to research articles as defined by Web of Science (e.g., original research, reviews). From these articles, we extracted all author e-mail addresses listed in the author information, removing duplicate authors. Authors were consider to be "HPE researchers" if they had published in one of the HPE journals searched. Altogether, this process generated a sample of 1,840 unique HPE researchers. All names and e-mails were entered into Qualtrics, an online survey tool (Qualtrics, Provo, Utah), and the survey was distributed in four waves of e-mail invitations: wave 1 (sent November 13, 2017), wave 2 (sent November 20, 2017), wave 3 (sent November 27, 2017), and wave 4 (sent December 11, 2017).

Next, we collected a "social media sample" by posting anonymous links to the survey on our Twitter and Facebook accounts (posted on December 11, 2017). All survey responses obtained from the social media links were tracked separately from those sent to the curated sample. To help prevent duplicate submissions, respondents in the social media sample were given the option to select "I have already completed this survey" on the informed consent page.

Statistical analyses

Prior to analysis, we screened the data for accuracy and missing values. Next, we calculated the response rate in the curated sample using response rate definition number 6, as delineated by the American Association for Public Opinion Research.[22] Then, to assess potential nonresponse bias in the curated sample, we used wave analysis to calculate a nonresponse bias statistic.[23] In wave analysis, late respondents are considered proxies for nonresponders, and their responses are compared with responses from the first wave. In addition, to determine whether or not it was appropriate to combine the curated and social media samples for analysis, we conducted a multivariate analysis of variance (MANOVA) to compare respondents on several demographic characteristics: age, experience doing HPE research, percentage of work time dedicated to HPE research, and number of peer-reviewed publications. Finally, we calculated descriptive statistics for the total sample, with particular emphasis on the frequency of self-reported misconduct and QRPs. All data analyses were conducted using IBM SPSS statistical software, version 24 (IBM Corporation, New York, New York) and Microsoft Excel 2013 (Microsoft Corporation, Redmond, Washington).

Results

Of the 1,840 e-mail invitations sent to HPE researchers in the curated sample, 199 were returned as undeliverable, leaving 1,641 potential respondents. Of these, 463 (28.2%) researchers completed at least a portion of the survey (see Supplemental Digital Appendix 1, available at http://links.lww.com/ACADMED/A584, for regional response rates).[22] Results from the wave analysis demonstrated a nonresponse bias statistic of 0.36. On a six-point, frequency-response scale, this represents a 6% difference, which is unlikely to have a meaningful effect on practical interpretation of the results.[23]

The social media sample yielded an additional 127 responses. Results from the MANOVA comparing respondents in the curated sample with those in the social media sample demonstrated statistically significant differences between the two groups $F(5, 524) = 6.67$, $P < .001$. In particular, post hoc analyses indicated that social media respondents were slightly younger (mean = 40.7 years) and more inexperienced in HPE research (mean = 7.5 years) than those in the curated sample (mean = 47.4 years and mean = 11.0 years, respectively). That said, the two groups did not differ in terms of percentage of work time spent doing HPE research activities or mean number of peer-reviewed publications. Therefore, because our goal was to explore the frequency of misconduct and QRPs among a diverse, international sample of HPE researchers, we pooled the two samples and analyzed the data together.

Of the 590 respondents in the pooled sample, the mean age was 46 years (SD = 11.6), and there were 305 (51.7%) women, 246 (41.7%) men, and 39 (6.6%) individuals who did not report gender. As indicated in Table 1, the sample consisted of HPE researchers from across the World Health Organization's six world regions. The majority reported their location as the United States (n = 156; 26.4%), Europe (n = 137; 23.2%), and Canada (n = 90; 15.3%). Respondents' education, area of study, work context and role, academic rank, and primary research activities are presented in Table 1. In addition, respondents reported the following: years involved in HPE in any capacity (mean = 14.9 years, SD = 9.7), years involved in HPE research (mean = 11.3 years, SD = 8.5), percentage of work time spent conducting HPE research (mean = 27.3%, SD = 23.7%), and total number of peer-reviewed publications (mean = 40.1, SD = 55.0).

Table 2 summarizes the frequency of self-reported misconduct and QRPs, and Figure 1 provides a visual representation of these results for the top 20 most frequently reported practices. As indicated, the most frequently reported behaviors were QRPs related to authorship and study reporting practices, as well as issues around data storage, collection, and interpretation. Additionally, 39 (6.7%) respondents reported misrepresenting a participant's words, 31 (5.5%) reported using sections of text from another author's copyrighted material without permission or proper citation, 30 (5.3%) reported inappropriately modifying study results because of pressure from a research advisor or collaborator, 20 (3.4%) reported deleting data before performing analysis without disclosure, and 14 (2.4%) reported fabricating data. Overall, 533 (90.3%) respondents reported at least one irresponsible behavior.

Discussion

This study examined the frequency of self-reported misconduct and QRPs among HPE researchers, practices that may be detrimental to scientific inquiry.[1,2] To our knowledge, this is the first study to explore irresponsible research practices across all phases of HPE research. Taken together, our findings indicate that a substantial proportion of HPE researchers admit to having engaged in a range of irresponsible behaviors, with QRPs reported more frequently than deliberate scientific misconduct. These findings are consistent with the extant literature on research misconduct in other fields[4–6,8,17] and are important because such behaviors can waste resources, provide an unfair advantage to some researchers over others, and ultimately impede scientific progress. Therefore, we believe this study raises important concerns about the conduct of HPE research, suggesting that our community may need to take a hard look at its ethical norms and research culture.

In the current survey study, we asked respondents about a range of problematic behaviors. As Fanelli[8] noted in his review of research misconduct, QRPs (e.g., honorary authorship or excluding study limitations) are qualitatively different from fabrication and falsification because they do not *directly* distort the quality of the science, per se. However, the damage done to the scientific enterprise by these "less severe" and more ambiguous practices may be proportionally greater than deliberate

Table 1
Demographic Characteristics of an International Sample of 590 Health Professions Education Researchers, From a Study of the Frequency of Self-Reported Misconduct and Questionable Research Practices, 2017

Demographic characteristic	No. (%)
Gender	
Female	305 (51.7)
Male	246 (41.7)
Not reported	39 (6.6)
Region	
United States	156 (26.4)
Europe (not including United Kingdom)	96 (16.3)
Canada	90 (15.3)
Australia and New Zealand	42 (7.1)
United Kingdom	41 (6.9)
Africa	37 (6.3)
Asia	37 (6.3)
South/Latin America and Caribbean	18 (3.0)
Middle East	16 (2.7)
Other	12 (2.0)
Not reported	45 (7.6)
Education[a]	
Bachelor's degree	245 (44.4)
Master's degree	306 (55.5)
MD/DO degree	276 (50.0)
PhD/EdD degree	323 (58.5)
Other professional degree	30 (5.4)
Area of study for highest degree selected	
Social science	293 (49.7)
Clinical science	167 (28.3)
Other	42 (7.1)
Basic science	39 (6.6)
Humanities	11 (1.9)
Not reported	38 (6.4)
Work context[a]	
Undergraduate medical education	353 (65.7)
Graduate medical education	308 (57.4)
Continuing medical education	142 (26.4)
Work role	
Researcher	174 (29.5)
Clinician	136 (23.1)
Administrator/program director	89 (15.1)
Teacher	86 (14.6)
Other	65 (11.0)
Not reported	40 (6.8)
Academic rank	
Professor	140 (23.7)
Associate professor	109 (18.5)
Assistant professor	99 (16.8)
Lecturer/instructor	64 (10.9)
Graduate student	42 (7.1)
Postdoctoral fellow	23 (3.9)
Resident	16 (2.7)
Other	52 (8.8)
Not reported	45 (7.6)
Primary research activity	
Mixed-method	280 (47.5)
Quantitative	149 (25.3)
Qualitative	119 (20.2)
Not reported	42 (7.1)

(*Table continues*)

Table 1
(Continued)

Abbreviations: MD, doctor of medicine; DO, doctor of osteopathic medicine.
[a]Percentages total more than 100% because respondents could select more than one category.

misconduct, simply because such practices occur more frequently. For example, 20.1% of respondents reported one or more instances of "salami slicing." Although some may think of this as a minor offense, the practice fills the literature with more articles than is seemingly necessary.[1,11] So, not only does this activity unfairly reward authors and waste resources (e.g., editorial time and journal space), it also can inflate the significance of a given finding, which in turn can distort the outcomes of meta-analyses and other types of systematic reviews.[11,24]

The interpretation of our findings is limited by the nature of the survey methodology employed and, in particular, by the threat of nonresponse bias (especially considering the sensitive nature of the topic under study).[5,6] Therefore, it is reasonable to ask: What do these results really demonstrate about the actual frequency of misconduct and QRPs among HPE researchers? We argue here, as others have previously,[5,8,17] that self-reports of irresponsible practices likely underestimate the real frequency of such behaviors. Researchers who have acted unethically are undoubtedly hesitant to reveal such activities in a survey, despite all assurances of anonymity. What is more, the opposite—researchers admitting to unethical or otherwise questionable practices that they did not do—seems unlikely.[8] Therefore, we speculate that scientific misconduct and QRPs may be even more widespread in our community than our estimates imply. Nevertheless, rather than establish an absolute prevalence of misconduct and QRPs in HPE, we believe these data are better suited for helping the community understand the nature of the most common practices and beginning to find feasible solutions to improve our research enterprise.

QRPs related to authorship were some of the most frequently reported behaviors in this study, particularly the practice of giving or accepting unwarranted authorship ("honorary authorship"). Honorary authorship is unethical because individuals who have not sufficiently contributed to the work unfairly receive credit as an author and misrepresent their contributions in the scientific literature.[25] Our findings corroborate the results of a recent survey of established HPE researchers[14] and of several studies conducted in the field of medicine more broadly, suggesting we are not alone in these practices.[25–27] Such questionable authorship practices have led journals to require that authors sign an author contribution agreement to verify their explicit authorship roles.[28] The effectiveness of such requirements is unknown and could be a fruitful area for additional research.

Of note in this study, the frequency of authors giving honorary authorship and those accepting honorary authorship were not equivalent: 60.6% admitted to adding undeserving authors, whereas only 22.7% admitted to accepting honorary authorship. This mismatch suggests that HPE researchers may not fully understand authorship criteria. It may also be a concrete example of the "Muhammad Ali effect"—the idea that individuals often see themselves as more likely to perform good acts and less likely to perform bad acts than others.[29] Regardless of the mechanism, this finding indicates the need for increased author communication and better shared understanding of author roles and responsibilities, such as those set forth by the ICMJE.[15]

Table 2

Frequency of Self-Reported Misconduct and Questionable Research Practices Among an International Sample of 590 Health Professions Education Researchers, From a Study of Misconduct and Questionable Research Practices, 2017[a]

Category[b]	N	No. (%)						
		Never	Once	Occasionally	Sometimes	Frequently	Almost always	Not applicable
Data collection and storage								
1: Conducted a human-subjects research study without ethics approval (i.e., without IRB approval)	589	465 (78.9)	50 (8.5)	38 (6. 5)	13 (2.2)	6 (1.0)	4 (0.7)	13 (2.2)
2: Circumvented one or more aspects of human-subjects ethics rules (i.e., IRB rules)	585	463 (79.1)	47 (8.0)	49 (8.4)	10 (1.7)	4 (0.7)	0 (0.0)	12 (2.1)
3: Collected course or curriculum data under the guise of "program evaluation" without human-subjects ethics (IRB) approval with the ultimate intent of using the data for research purposes	588	420 (71.4)	45 (7.7)	65 (11.1)	24 (4.1)	9 (1.5)	2 (0.3)	23 (3.9)
4: Inappropriately stored sensitive research data (e.g., data that contains personally identifiable information)	585	353 (60.3)	56 (9.6)	119 (20.3)	36 (6.2)	15 (2.6)	2 (0.3)	4 (0.7)
5: Inappropriately e-mailed sensitive research data (e.g., data that contain personally identifiable information)	585	436 (74.5)	36 (6.2)	84 (14.4)	20 (3.4)	3 (0.5)	0 (0.0)	6 (1.0)
6: Stopped collecting data earlier than planned because the results already reached statistical significance, without formal stopping rules	586	497 (84.8)	22 (3.8)	13 (2.2)	6 (1.0)	3 (0.5)	1 (0.2)	44 (7.5)
7: Fabricated data	586	570 (97.3)	10 (1.7)	2 (0.3)	1 (0.2)	0 (0.0)	1 (0.2)	2 (0.3)
8: Pressured a student or other subordinate to be a study participant in your research	583	536 (91.9)	14 (2.4)	20 (3.4)	4 (0.7)	1 (0.2)	0 (0.0)	8 (1.4)
9: Used students or residents as research subjects without informing the overseeing dean, program director, or other pertinent official	584	468 (80.1)	23 (3.9)	46 (7.9)	17 (2.9)	8 (1.4)	2 (0.3)	20 (3.4)
Data analysis								
10: Deleted data before performing data analysis without disclosure	582	553 (95.0)	9 (1.5)	11 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)	9 (1.5)
11: Ignored a colleague's use of flawed data	581	447 (76.9)	36 (6.2)	46 (7.9)	16 (2.8)	5 (0.9)	1 (0.2)	30 (5.2)
12: Ignored a colleague's questionable interpretation of data	581	404 (69.5)	44 (7.6)	88 (15.1)	17 (2.9)	2 (0.3)	2 (0.3)	24 (4.1)
13: Reported a downwardly rounded *P* value (e.g., reporting that a *P* value of .054 is less than .05)	581	513 (88.3)	11 (1.9)	8 (1.4)	0 (0.0)	0 (0.0)	0 (0.0)	49 (8.4)
14: Misrepresented a participant's words or writings	580	526 (90.7)	16 (2.8)	21 (3.6)	2 (0.3)	0 (0.0)	0 (0.0)	15 (2.6)
15: Decided whether to exclude nonoutlier data after looking at the impact of doing so on the results	584	470 (80.5)	32 (5.5)	32 (5.5)	7 (1.2)	1 (0.2)	0 (0.0)	42 (7.2)
16: In a qualitative study, failed to report disconfirming examples or cases that weaken your conclusions	583	422 (72.4)	32 (5.5)	32 (5.5)	6 (1.0)	1 (0.2)	0 (0.0)	90 (15.4)
17: Collected more data after seeing that the results were almost statistically significant	583	434 (74.4)	33 (5.7)	40 (6.9)	4 (0.7)	3 (0.5)	1 (0.2)	68 (11.7)
18: To confirm a hypothesis, selectively deleted or changed data after performing data analysis	583	523 (89.7)	16 (2.7)	6 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)	38 (6.5)

(*Table continues*)

Table 2
(Continued)

Category[b]	N	No. (%)						
		Never	Once	Occasionally	Sometimes	Frequently	Almost always	Not applicable
19: Reported an unexpected finding as having been hypothesized from the start	583	433 (74.3)	47 (8.1)	55 (9.4)	9 (1.5)	4 (0.7)	2 (0.3)	33 (5.7)
20: Concealed results that contradicted your previous findings or convictions	583	538 (92.3)	23 (3.9)	11 (1.9)	1 (0.2)	0 (0.0)	0 (0.0)	10 (1.7)
21: Claimed you used a particular qualitative research approach appropriately (e.g., grounded theory) when you knowingly did not	566	432 (76.3)	39 (6.9)	22 (3.9)	3 (0.5)	0 (0.0)	0 (0.0)	70 (12.4)
22: Claimed you used a particular qualitative research technique appropriately (e.g., saturation, triangulation) when you knowingly did not	566	433 (76.5)	42 (7.4)	18 (3.2)	3 (0.5)	0 (0.0)	0 (0.0)	70 (12.4)
Study reporting								
23: Spread study results over more papers than is appropriate ("salami slicing")	566	442 (78.1)	59 (10.4)	52 (9.2)	2 (0.4)	1 (0.2)	0 (0.0)	10 (1.8)
24: Deliberately failed to mention important limitations of a study in the published paper	566	517 (91.3)	19 (3.4)	26 (4.6)	3 (0.5)	0 (0.0)	0 (0.0)	1 (0.2)
25: Deliberately failed to mention an organization that funded your research in the published paper	565	549 (97.2)	5 (0.9)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	10 (1.8)
26: Inappropriately modified the results of a study due to pressure from a research advisor or other collaborator	564	531 (94.1)	21 (3.7)	8 (1.4)	1 (0.2)	0 (0.0)	0 (0.0)	3 (0.5)
27: Inappropriately modified the results of a study due to pressure from a funding agency	566	532 (94.0)	5 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	29 (5.1)
28: Failed to disclose relevant financial or intellectual conflicts of interest	567	548 (96.6)	6 (1.1)	4 (0.7)	1 (0.2)	0 (0.0)	0 (0.0)	8 (1.4)
29: Used someone else's ideas without their permission or proper citation	567	536 (94.5)	15 (2.6)	14 (2.5)	1 (0.2)	0 (0.0)	0 (0.0)	1 (0.2)
30: Used sections of text from another author's copyrighted material without permission or proper citation	566	534 (94.3)	13 (2.3)	16 (2.8)	2 (0.4)	0 (0.0)	0 (0.0)	1 (0.2)
31: Used sections of text from your own publications without proper citation ("self-plagiarism")	564	449 (79.6)	35 (6.2)	65 (11.5)	12 (2.1)	0 (0.0)	0 (0.0)	3 (0.5)
32: Selectively cited certain papers just to please editors or reviewers	566	284 (50.2)	57 (10.1)	170 (30.0)	39 (6.9)	7 (1.2)	7 (1.2)	2 (0.4)
33: Cited articles and or materials that you have not read	565	284 (50.3)	39 (6.9)	181 (32.0)	48 (8.5)	12 (2.1)	0 (0.0)	1 (0.2)
34: Selectively cited your own work just to improve your citation metrics	562	391 (69.6)	28 (5.0)	101 (18.0)	24 (4.3)	15 (2.7)	2 (0.4)	1 (0.2)
35: Reused previously published data without disclosure ("duplicate publication")	564	545 (96.6)	11 (2.0)	4 (0.7)	1 (0.2)	0 (0.0)	0 (0.0)	3 (0.5)
36: Used confidential information obtained as a reviewer or editor for your own research or publications	564	543 (96.3)	11 (2.0)	6 (1.1)	0 (0.0)	0 (0.0)	0 (0.0)	4 (0.7)

(*Table continues*)

Table 2
(Continued)

Category[b]	N	No. (%) Never	Once	Occasionally	Sometimes	Frequently	Almost always	Not applicable
Collaboration and authorship								
37: Refused to share data with legitimate colleagues	564	535 (94.9)	11 (2.0)	6 (1.1)	0 (0.0)	2 (0.4)	2 (0.4)	8 (1.4)
38: Added one or more authors to a paper who did not qualify for authorship ("honorary authorship")	563	219 (38.9)	91 (16.2)	163 (29.0)	54 (9.6)	28 (5.0)	5 (0.9)	3 (0.5)
39: Accepted authorship for which you did not qualify ("honorary authorship")	564	433 (76.8)	78 (13.8)	43 (7.6)	6 (1.1)	1 (0.2)	0 (0.0)	3 (0.5)
40: Demanded authorship for which you did not qualify ("honorary authorship")	565	550 (97.3)	8 (1.4)	3 (0.5)	1 (0.2)	0 (0.0)	0 (0.0)	3 (0.5)
41: Omitted a contributor who deserved authorship	563	531 (94.3)	23 (4.1)	7 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.4)
42: Submitted (or resubmitted) a manuscript or grant application without consent from one or more of the authors	563	499 (88.6)	27 (4.8)	29 (5.2)	4 (0.7)	2 (0.4)	0 (0.0)	2 (0.4)
43: Submitted the same manuscript to multiple journals at once ("duplicate" or "double submission")	564	558 (98.9)	4 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.4)

Abbreviation: IRB indicates institutional review board.
[a]Responses were in answer to the question "In your work as an HPE researcher, how often have you engaged in any of the following behaviors, even if it has been only on a single occasion? If applicable, please consider your experiences with both quantitative and qualitative research." Percentages are calculated using data from individuals who responded to the item (i.e., nonresponder data are not included in the denominator).
[b]Misconduct and questionable research practices are listed in the order presented on the survey.

A complete discussion of all the behaviors assessed in this study is outside the scope of this article. Although readers may debate the degree to which some of these behaviors are unethical or otherwise problematic, we should note, as described above, that seemingly minor infractions can have far-reaching negative consequences. For example, by employing data manipulation techniques like "*P* hacking" (i.e., massaging data or analyses until nonsignificant results become significant)[30,31] or taking advantage of other types of "researcher degrees of freedom,"[32] scientists can discover "illusory results"[33] that actually represent artifacts of their study design and analytic approach, as opposed to legitimate findings that can be replicated.[34,35] Some have argued that such behaviors are the result of individual researchers responding to a set of incentives, the most important of which are rewards for the *quantity* (not the quality) of publications.[17,36] If one accepts this thesis, then the most effective way to improve research practices, and the quality of the corresponding science, may be to change incentive structures at the institutional level.[37]

Limitations and future directions

This study has several important limitations. First, nonresponse bias is a legitimate concern, especially considering the modest response rate (28.2%) in the curated sample. That said, recent research suggests that response rate may be a flawed indicator of response quality and representativeness.[38,39] Moreover, notwithstanding its inherent limitations, our wave analysis results suggest that nonresponse bias was limited in our sample. Nonetheless, investigators should build on these initial results by examining misconduct and QRPs among a larger, more global sample of HPE researchers.

A second limitation relates to the innate challenge of assessing complex, context-specific research behaviors with a survey that requires respondents to self-assess their own practices.[40,41] Because some of the irresponsible practices on our survey are judgment calls, their evaluation likely requires more detail and nuance than a single survey item can provide.[40] So, while the practice of fabricating data is fairly straightforward (and never justified), the same cannot be said for something like inappropriately storing sensitive research data. The latter practice is open to interpretation: What is considered "inappropriate storage" to one researcher might seem innocuous to another. We attempted to reduce this type of ambiguity in our survey items by employing a rigorous expert review process. But, ultimately, "survey self-reports can never fully rule out ambiguities in meaning, limitations in autobiographical memory, or motivated biases."[40] Thus, future work might apply qualitative research methods to address some of these limitations and further unpack the nature of irresponsible research practices in HPE.

Finally, we administered our survey in English and did not ask respondents to focus on a particular time period. These implementation and design choices, particularly the latter decision to use an open-ended time period, could

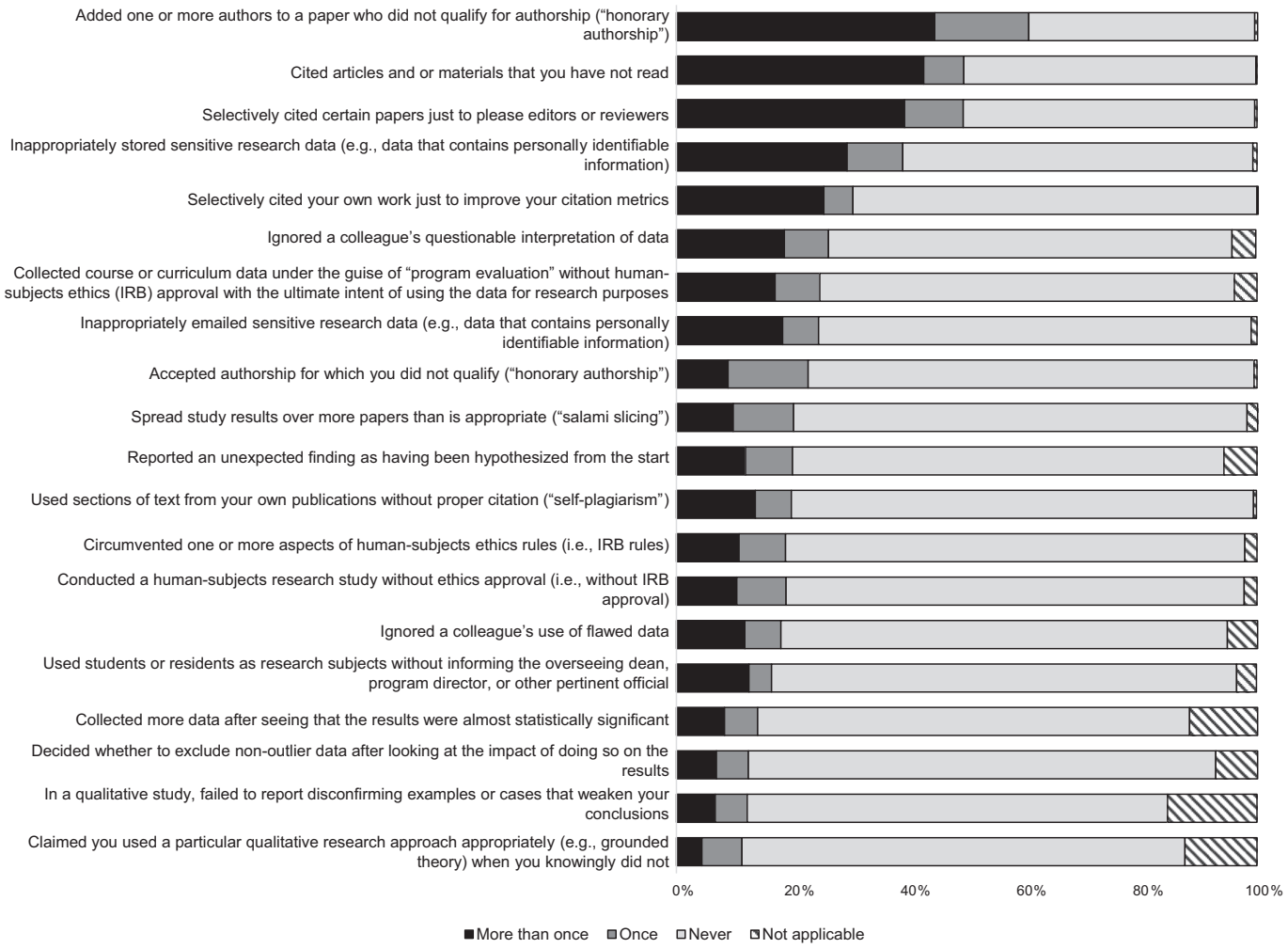

Figure 1 Stacked bar graph showing the frequency of self-reported questionable research practices among an international sample of 590 health professions education researchers for the top 20 most frequently reported practices, from an international study of frequency of self-reported misconduct and questionable research practices, 2017. To improve readability, the response options of "occasionally," "sometimes," "frequently," and "almost always" were collapsed into a single frequency option labeled "more than once."
Abbreviation: IRB indicates institutional review board.

have negatively affected data quality. For example, several respondents noted in their written comments that ethical standards related to human-subjects research had evolved over time. Therefore, respondents reporting a given QRP may have been referring to activities that occurred decades prior, long before more stringent standards of conduct were in place. Future research could address this problem by limiting the time period that respondents are asked to consider.[5]

Recommendations for practice

As the field of HPE continues to mature, it is essential that we explicitly confront our obligation to conduct research in an ethical and responsible manner. Previously, we suggested a number of recommendations to improve practice,[7] and the findings reported here indicate that the time is *now* to implement these and other policy changes. These institution-level approaches embrace the idea that scientific misconduct and QRPs are not simply the result of individual researchers acting badly. Instead, important contextual factors likely influence researcher behavior, including social norms, power disparities, institutional policies, and academic incentives.[7] Examples of initiatives being tried elsewhere include promotion and tenure guidelines that privilege publication quality over quantity,[42] study preregistration plans,[33] and other open-science practices (e.g., open data and materials sharing).[43] All of these approaches require study to determine their efficacy in HPE research and publication.

In summary

Cultivating the responsible conduct of research is essential if we are to maintain scientific integrity and engender public confidence in our research.[1,2] This study raises important concerns about the conduct of HPE research and presents a somewhat pessimistic picture of our community. We should be clear, though—most HPE scientists surveyed did *not* report the vast majority of irresponsible research behaviors. They are presumably acting ethically and doing good science. Nevertheless, we believe that reforms are needed. In addition, we recommend future research to monitor research misconduct and QRPs in HPE, and to evaluate the effectiveness of policies designed to improve the integrity of our research enterprise.

Acknowledgments: The authors wish to sincerely thank all of the health professions education researchers who took the time (and had the courage) to complete the survey.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: The Ethical Review Board Committee of the Netherlands Association for Medical Education approved this study (dossier no. 937).

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences, the U.S. Navy, the Department of Defense, or the U.S. Government.

Previous presentations: An earlier version of this article was posted to bioRxiv as a preprint on January 31, 2018. The preprint is entitled Ethical Shades of Gray: Questionable Research Practices in Health Professions Education. doi: https://doi.org/10.1101/256982.

A.R. Artino Jr is professor of medicine, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland; ORCID: http://orcid.org/0000-0003-2661-7853.

E.W. Driessen is professor of medical education, Faculty of Health, Medicine and Life Science, Maastricht University, Maastricht, the Netherlands; ORCID: https://orcid.org/0000-0001-8115-261X.

L.A. Maggio is associate professor of medicine, Department of Medicine, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland; ORCID: https://orcid.org/0000-0002-2997-6133.

References

1 Steneck NH. Fostering integrity in research: Definitions, current knowledge, and future directions. Sci Eng Ethics. 2006;12:53–74.
2 Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435:737–738.
3 Steneck NH. Introduction to the responsible conduct of research. https://ori.hhs.gov/sites/default/files/rcrintro.pdf. Accessed July 30, 2018.
4 John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012;23:524–532.
5 Tijdink JK, Bouter LM, Veldkamp CLS, van de Ven PM, Wicherts JM, Smulders YM. Personality traits are associated with research misbehavior in Dutch scientists: A cross-sectional study. Dorta-González P, ed. PLoS One. 2016;11(9):e0163251.
6 Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: Results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1:17.
7 Maggio LA, Artino AR, Picho K, Driessen EW. Are you sure you want to do that? Fostering the responsible conduct of medical education research. Acad Med. 2018;93:544–549.
8 Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. Tregenza T, ed. PLoS One. 2009;4:e5738.
9 Committee on Publication Ethics. Promoting integrity in research and its publication. https://publicationethics.org. Accessed July 30, 2018.
10 Brice J, Bligh J, Bordage G, et al. Publishing ethics in medical education journals. Acad Med. 2009;84(10 suppl):S132–S134.
11 Eva KW. How would you like your salami? A guide to slicing. Med Educ. 2017;51:456–457.
12 ten Cate O. Why the ethics of medical education research differs from that of medical research. Med Educ. 2009;43:608–610.
13 Hally E, Walsh K. Research ethics and medical education. Med Teach. 2016;38:105–106.
14 Uijtdehaage S, Mavis B, Durning SJ. Whose paper is it anyway? Authorship criteria according to established scholars in health professions education. Acad Med. 2018;93:1171–1175.
15 International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. http://www.icmje.org/icmje-recommendations.pdf. Published 2017. Accessed July 30, 2018.
16 White C, ed. The COPE Report 2000: Annual Report of the Committee on Publication Ethics. London, UK: BMJ Books; 2000.
17 Tijdink JK, Verbeke R, Smulders YM. Publication pressure and scientific misconduct in medical scientists. J Empir Res Hum Res Ethics. 2014;9:64–71.
18 Anderson MS, Horn AS, Risbey KR, Ronning EA, De Vries R, Martinson BC. What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists. Acad Med. 2007;82:853–860.
19 Schaeffer NC, Dykema J. Questions for surveys: Current trends and future directions. Public Opin Q. 2011;75:909–961.
20 Tourangeau R, Yan T. Sensitive questions in surveys. Psychol Bull. 2007;133:859–883.
21 Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE guide no. 87. Med Teach. 2014;36:463–474.
22 American Association for Public Opinion Research (AAPOR). Standard definitions: Final dispositions of case codes and outcome rates for surveys. 9th ed. http://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx. Revised 2016. Accessed January 1, 2018.
23 Phillips AW, Reddy S, Durning SJ. Improving response rates and evaluating nonresponse bias in surveys: AMEE guide no. 102. Med Teach. 2016;38:217–228.
24 Supak Smolcić V. Salami publication: Definitions and examples. Biochem Med (Zagreb). 2013;23:237–241.
25 Wislar JS, Flanagin A, Fontanarosa PB, Deangelis CD. Honorary and ghost authorship in high impact biomedical journals: A cross sectional survey. BMJ. 2011;343:d6128.
26 Kornhaber RA, McLean LM, Baber RJ. Ongoing ethical issues concerning authorship in biomedical journals: An integrative review. Int J Nanomed. 2015;10:4837–4846.
27 Vera-Badillo FE, Napoleone M, Krzyzanowska MK, et al. Honorary and ghost authorship in reports of randomised clinical trials in oncology. Eur J Cancer. 2016;66:1–8.
28 Lundberg GD, Flanagin A. New requirements for authors: Signed statements of authorship responsibility and financial disclosure. JAMA. 1989;262:2003–2004.
29 Allison ST, Messick DM, Goethals GR. On being better but not smarter than others: The Muhammad Ali effect. Soc Cogn. 1989;7:275–295.
30 Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLoS Biol. 2015;13:e1002106.
31 Nuzzo R. Scientific method: Statistical errors. Nature. 2014;506:150–152.
32 Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011;22:1359–1366.
33 Gehlbach H, Robinson CD. Mitigating illusory results through preregistration in education. J Res Educ Eff. October 2017:1–20.
34 Picho K, Maggio LA, Artino AR Jr. Science: The slow march of accumulating evidence. Perspect Med Educ. 2016;5:350–353.
35 Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124.
36 Horton R. Offline: What is medicine's 5 sigma? Lancet. 2015;385:1380.
37 Smaldino PE, McElreath R. The natural selection of bad science. R Soc Open Sci. 2016;3:160384.
38 Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. JAMA. 2012;307:1805–1806.
39 Halbesleben JR, Whitman MV. Evaluating survey quality in health services research: A decision framework for assessing nonresponse bias. Health Serv Res. 2013;48:913–930.
40 Fiedler K, Schwarz N. Questionable research practices revisited. Soc Psychol Personal Sci. 2016;7:45–52.
41 Eva KW, Regehr G. Self-assessment in the health professions: A reformulation and research agenda. Acad Med. 2005;80(10 suppl):S46–S54.
42 Nazim Ali S, Young HC, Ali NM. Determining the quality of publications and research for tenure or promotion decisions. Libr Rev. 1996;45:39–53.
43 Kidwell MC, Lazarević LB, Baranski E, et al. Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. Macleod MR, ed. PLOS Biol. 2016;14:e1002456.