Journal of Clinical Epidemiology 154 (2023) 23–32

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Indicators of questionable research practices were identified in 163,129 randomized controlled trials

Johanna A. Damen[a,*], Pauline Heus[a], Herm J. Lamberink[b,c], Joeri K. Tijdink[d,e], Lex Bouter[f,g], Paul Glasziou[h], David Moher[i,j], Willem M. Otte[b,k], Christiaan H. Vinkers[l,m,n,o], Lotty Hooft[a]

[a]*Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*

[b]*Department of Child Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht and Utrecht University, Utrecht, The Netherlands*

[c]*Department of Neurology, Haaglanden Medical Center, Den Haag, The Netherlands*

[d]*Department of Ethics, Law and Humanities, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

[e]*Department of Philosophy, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

[f]*Department of Epidemiology and Data Science, Amsterdam UMC, Amsterdam, The Netherlands*

[g]*Department of Philosophy, Vrije Universiteit, Amsterdam, The Netherlands*

[h]*Institute for Evidence-Based Healthcare, Bond University, Gold Coast, Australia*

[i]*Centre for Journalology, Clinical Epidemiology Program, The Ottawa Hospital Research Institute, Ottawa, Canada*

[j]*School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada*

[k]*Biomedical MR Imaging and Spectroscopy group, Center for Image Sciences, University Medical Center Utrecht and Utrecht University, Utrecht, The Netherlands*

[l]*Department of Psychiatry and Anatomy & Neurosciences, Amsterdam University Medical Center Location Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands*

[m]*Amsterdam Public Health, Mental Health Program and Amsterdam Neuroscience, Mood, Anxiety, Psychosis, Sleep & Stress Program, Amsterdam, The Netherlands*

[n]*Amsterdam Public Health (Mental Health Program) Research Institute, 1081 HV Amsterdam, The Netherlands*

[o]*GGZ inGeest Mental Health Care, 1081 HJ Amsterdam, The Netherlands*

Accepted 29 November 2022; Published online 2 December 2022

Abstract

Objectives: To explore indicators of the following questionable research practices (QRPs) in randomized controlled trials (RCTs): (1) risk of bias in four domains (random sequence generation, allocation concealment, blinding of participants and personnel, and blinding of outcome assessment); (2) modifications in primary outcomes that were registered in trial registration records (proxy for selective reporting bias); (3) ratio of the achieved to planned sample sizes; and (4) statistical discrepancy.

Study Design and Setting: Full texts of all human RCTs published in PubMed in 1996–2017 were automatically identified and information was collected automatically. Potential indicators of QRPs included author-specific, publication-specific, and journal-specific characteristics. Beta, logistic, and linear regression models were used to identify associations between these potential indicators and QRPs.

Author contributions: Johanna A Damen: Conceptualization, Methodology, Formal analysis, Data curation, and Writing–Original Draft. Pauline Heus: Conceptualization, Methodology, and Writing–Review and Editing. Herm J Lamberink: Conceptualization, Methodology, Data curation, and Writing–Review and Editing. Joeri K Tijdink: Conceptualization, Methodology, Writing–Review and Editing, and Funding acquisition. Lex Bouter: Conceptualization, Methodology, and Writing–Review and Editing. Paul Glasziou: Conceptualization and Writing–Review and Editing. David Moher: Conceptualization and Writing–Review and Editing. Willem M Otte: Conceptualization, Methodology, Data curation, and Writing–Review and Editing. Christiaan H Vinkers: Conceptualization, Methodology, Writing–Review and Editing, Supervision, and Funding acquisition. Lotty Hooft: Conceptualization, Methodology, Writing–Review and Editing, and Supervision.

Data availability statement: The risk of bias characterization was done with a large-batch customized Python scripts (version 3; https://github.com/wmotte/robotreviewer_prob). The data management and analyses used R (version 3.6.1). All data are available at https://github.com/wmotte/RCTQuality).

Funding: This work was supported by The Netherlands Organization for Health Research and Development (ZonMw) grant "Fostering Responsible Research Practices" (445001002) (C.V., J.T., and W.O.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declarations of interest: none.

* Corresponding author. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, P.O. Box 85500, Str. 6.131, 3508 GA Utrecht, The Netherlands. Tel.: +31 6 25777481.

E-mail address: J.A.A.Damen@umcutrecht.nl (J.A. Damen).

https://doi.org/10.1016/j.jclinepi.2022.11.020

Results: We included 163,129 RCT publications. The median probability of bias assessed using Robot Reviewer software ranged between 43% and 63% for the four risk of bias domains. A more recent publication year, trial registration, mentioning of CONsolidated Standards Of Reporting Trials-checklist, and a higher journal impact factor were consistently associated with a lower risk of QRPs.

Conclusion: This comprehensive analysis provides an insight into indicators of QRPs. Researchers should be aware that certain characteristics of the author team and publication are associated with a higher risk of QRPs.

Keywords: Responsible research; Bias; Meta-research; RCT; Questionable research; Selective reporting

1. Introduction

Systematic reviews synthesize the results of randomized controlled trials (RCTs) and constitute the backbone of evidence-based medicine. Healthcare professionals rely on these reviews and guidelines to determine which treatments to use in clinical practice. Knowledge gained from RCTs is increasing but methods to minimize bias are not always used, leading to methodological flaws, statistical problems, and interpretation bias (spin), often making the methods and results difficult to reproduce [1,2].

Concerns about the quality of research are certainly not new. Questionable research practices (QRPs) were mentioned in the 1958 Code of Professional Ethics and Practices of Public Opinion Researchers [3]. Banks et al. defined QRPs as design, analytic, or reporting practices that have been questioned because of the potential for the practice to be employed to present biased evidence in favor of an assertion [4]. Examples of QRPs include selective reporting, p-hacking, HARKing (i.e., hypothesizing after results are known) etc. [4−6].

To promote responsible research practices, codes of conduct have been published, including the European Code of Conduct for Research Integrity [7] and a report on Fostering Integrity of Research by the US National Academies of Science [8,9]. Evidence exists for some indicators of QRP. For example, associations have been reported between journal impact factor and risk of bias [10], author experience and effect sizes [11], and study quality and the continent of origin of authors [12,13].

Previous studies focused on one specific QRP and explored a limited set of indicators in small datasets. Furthermore, time trends in quality indicators of RCTs have been described before in large datasets, including the dataset used in the present article [14,15]. To obtain more insight into possible factors associated with QRPs, a large study including more QRPs and a broader set of indicators is necessary. We therefore aimed to validate existing and identify new indicators of QRPs in RCTs. We investigated QRPs concerning risk of bias, modifications in primary outcomes, the ratio of achieved sample size to planned sample size, and statistical discrepancy. The rationale for these QRPs is that they all relate to quality of the study and quality of reporting, which is seen as an essential element of responsible research [7,16]. We focused on demographic and bibliometric indicators, including characteristics of the author team, trial/publication and journal, available during different phases of a project: during trial registration, when a study is submitted for publication, and after a study is published.

2. Methods

A protocol for this study has been made publicly available on the Open Science Framework on December 19, 2018, before start of data collections [17]. Deviations from the protocol are described in Appendix 1.

2.1. Identification of RCTs

We searched PubMed using the Entrez API (https://www.ncbi.nlm.nih.gov/home/develop/api/) via R Statistical Software [18] on November 17, 2017 to identify studies with publication type RCT and automatically excluded non-randomized, animal, pilot, and feasibility studies (Appendix 1). In addition, articles were excluded when the language was other than English. Articles published before 1996 were excluded because in that year the CONsolidated Standards Of Reporting Trials (CONSORT) statement was published aiming to enhance the completeness of reporting of RCTs [19].

We developed web scrapers to automatically download the PDF of each identified RCT via the website of the respective publisher. Downloaded PDFs were transformed to text data in Extensible Markup Format (XML), using GROBID software [20].

2.2. Data collection of QRPs

We assessed the following four QRPs (Box 1):

1. Risk of bias, the probability of bias as determined using Robot Reviewer [21,27] for the domains random sequence generation, allocation concealment, blinding of participants and personnel, and blinding of outcome assessment [21−23].
2. Modifications in primary outcome measures based on comparing first and final versions of the public trial registration records from ClinicalTrials.gov [24].

What is new?

Key findings

- In a sample of 163,129 randomized controlled trial publications we found that a more recent publication year, trial registration, mentioning of CONsolidated Standards Of Reporting Trials-checklist, and a higher journal impact factor were consistently associated with a lower risk of questionable research practices.

What this adds to what was known?

- We validated previously identified associations between indicators and questionable research practices and explored new indicators.

What is the implication?

- Our results might inform future strategies to identify those randomized controlled trials at high risk of questionable research practices.

What should change now?

- Editors, peer reviewers, and readers should be aware that certain characteristics of the author team, the journal, and the publication might be associated with questionable research practices.

3. The ratio of achieved sample size compared to what was planned.
4. Statistical discrepancy, for which we compared the reported *P* value and actual *P* value of the intervention effect estimate calculated from other reported information such as the confidence interval.

Trial registry numbers were collected by searching abstracts and full texts using regular expressions (i.e., sequences of characters that specify a search pattern) and we subsequently obtained public trial registration records from ClinicalTrials.gov.

2.3. Data collection of indicators

Potential indicators of QRPs were selected based on previous evidence, discussions with experts, availability, and feasibility. They are listed in Box 2 and included characteristics of the (1) author team (e.g., gender, number of authors), (2) publication (e.g., reporting of trial registration), and (3) journal (e.g., impact factor). Data were automatically extracted from information indexed in PubMed (e.g., authors, affiliations, etc.) and from the full-text article as XML. Using the PubMed ID, RCTs were linked to Scopus and additional information on characteristics of author teams (e.g., Hirsch-index) was obtained [21,22,27].

Box 1 Methods for collecting information on questionable research practices

Risk of bias
Risk of Bias domains was extracted via open source software provided by Robot Reviewer [21]. Robot Reviewer is developed to score bias for four domains of the Cochrane Risk of Bias tool: random sequence generation, allocation concealment, blinding of participants and personnel, and blinding of outcome assessment [22]. Robot Reviewer assesses the probability that a study has bias rather than dichotomizing it into high or low risk of bias. Level of agreement between Robot Reviewer and human raters was similar for most domains (average human–human agreement 79% [range 71% to 85%], human–Robot Reviewer agreement 65% [range 39% to 91%]) [21,23].
Modifications in primary outcomes
Changes made to the primary outcome after the trial had started, as reported in the trial registration on Cinicaltrials.gov. Changes were first automatically extracted by comparing the first and final version of the primary outcome as registered in the study protocols on clinicaltrials.gov. Additions and deletions of complete outcome measures were extracted. The algorithm was too sensitive for changes in the content: if any textual change was present, the primary outcome was flagged as changed. These flagged studies were subsequently manually checked to distinguish between significant and insignificant (e.g., typo's) changes [24].
Ratio of achieved sample size compared to what was planned
We calculated the ratio of actual sample size and planned sample size based on the power calculation provided in the public trial registration records from ClinicalTrials.gov. This information could be extracted directly from the trial registration record. A manual check was performed for all publications and protocols where the ratio of the number of enrolled and estimated participants was > 100, that is, 100 times more enrolled than was estimated.
Statistical discrepancy
Comparison of reported *P* value and actual *P* value of the intervention effect estimate calculated from other reported information. Based on the reported relative risk, odds ratio or hazard ratio in combination with its 95% confidence intervals, the *P* value was recomputed. This value was compared with the reported *P* value using a script by Georgescu and Wren [25]. For *t*-tests, Chi-square values, F-values z-statistics, and correlations, the R-package StatCheck [26] was used to check the correct reporting of the *P* value. Inconsistent *P* values (defined as a difference ≥ 0.01) were marked. Every inconsistency where the adjusted *P* value crosses the level of 0.05 compared to the original *P* value was labeled as statistical discrepancy.

Box 2 Collected demographic and bibliometric indicators (more details can be found in Appendix 1)

Author team

- Gender of first and last author [11,13,28,29] (https://genderize.io/)
- Proportion of female authors in the author team
- Total number of authors [11,30]
- Continent of first and last author [12,13,31]
- Number of countries to which the author team is affiliated
- Hirsch-index of first and last author in the year before publication [4,11,30]
- Academic age of first and last author (i.e., number of years between the trial publication and first publication by this author) [11,13,32]
- Uninterrupted presence of first and last author (i.e., the number of years the author has published at least one article in sequentially without interruption) [13]
- Number of collaborations of the first and last author (i.e., total number of co-authorships until year of publication)
- Number of institutions represented in the author team [12]
- Ranking of institution of first and last author in the Academic Ranking of World Universities (www.shanghairanking.com)

Trial/publication

- Trial registration
- Financial support (industrial, other, and none) [11,12,31,33]
- Year of publication
- Conflict of interest
- Mentioning of the CONSORT Statement
- Positive and negative word frequencies in abstract [34]
- Number of words and number of names mentioned in acknowledgments

Journal

- Medical field [12,33]
- Journal impact factor in the year before publication [33,35,36]
- Impact factor change compared to previous year
- Number of publications of the journal per year
- Journal publisher
- Continent of journal

Detailed descriptions of definitions and methods for outcomes and indicators are described in Appendix 1.

2.4. Statistical analyses

Detailed analyses are described in Appendix 1. In short, associations between indicators and outcomes were assessed using univariable and multivariable regression models. Three multivariable regression models were fitted per outcome: (1) a full model including all indicators (Box 2); (2) a reduced model including indicators available upon journal submission of an article but before publication; and (3) a reduced model including indicators available upon trial design and registration but before the trial is completed [37]. Indicators in the model were selected based on *a priori* group discussions on the relevance of the indicators. We used beta regression models (R package 'betareg' [38]) for probability of bias, logistic regression (R package 'rms' [39]) for modifications in primary outcomes and statistical discrepancy, and linear regression (R package base [18]) for the log-transformed ratio of achieved to planned sample size. For all multivariable models, one indicator from an indicator pair was excluded based on discussions between authors if there was multicollinearity (i.e., Spearman correlation > 0.8). Indicators with more than 40% missing values were excluded from analyses. For the other indicators and QRPs, missing values were imputed 20 times using Multiple Imputation by Chained Equations [40]. Data transformations were applied if required and a Bonferroni correction for multiple testing was applied. Goodness of fit was assessed in terms of explained variance (i.e., R^2).

3. Results

3.1. Study flow

The search identified 445,159 records, of which 138,422 were excluded automatically because they were likely not

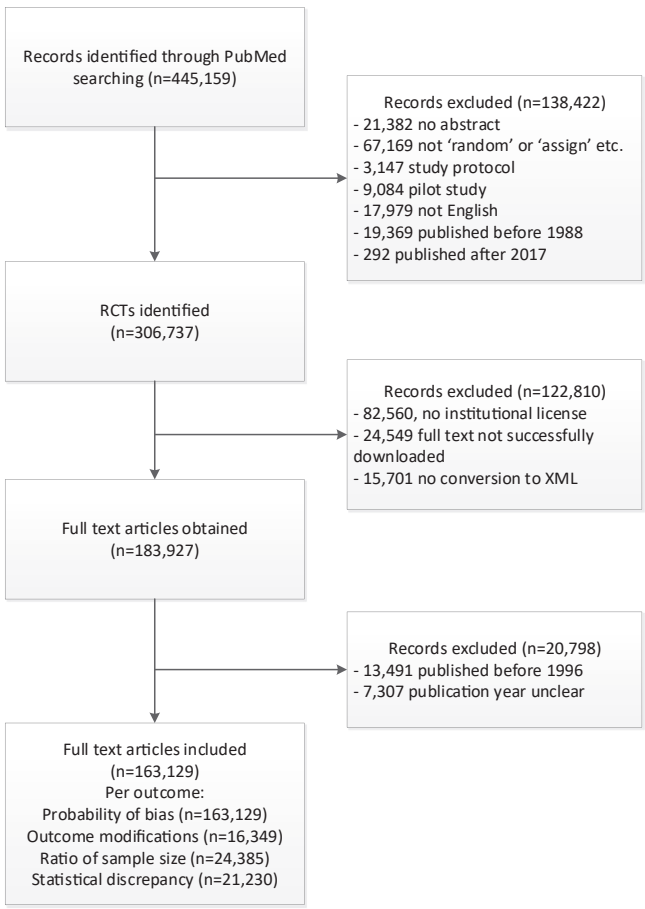

Fig. 1. Flow chart.

describing an RCT (Fig. 1). After excluding references for which we could not obtain the full text (n = 122,810) and that were published before 1996 or with an unclear publication year (n = 20,798), we included 163,129 in the analyses for each of the four probability of bias outcomes. For ratio of achieved to planned sample size and modifications in primary outcomes, we excluded additional references due to the unavailability of registration on ClinicalTrials.gov. For statistical discrepancy, references were excluded because no combination of P value and test statistic could be identified, leaving 21,230 references.

3.2. Components of questionable research practices

The median probabilities of bias ranged between 43% (interquartile range [IQR] 18%–59%) for randomization and 63% (IQR 40%–75%) for blinding of patients and personnel (Table 1). Twenty two percent (95% confidence interval [CI] 21%–23%) of studies had modified their primary outcome and we found a median ratio of achieved to planned sample size of 1 (IQR 0.98–1.04). In 370 of 21,230 publications (1.7% [95% CI 1.6%–1.9%]), we identified statistical discrepancy.

3.3. Demographic and bibliometric indicators

The majority of the included publications had a male first author (61.8% [95% CI 61.6%–62.1%]) and a male last author (73.6% [95% CI 73.4%–73.8%]), with a median of 33% (IQR 17%–50%) female authors (Appendix 2). Author teams included a median of six (IQR 5–9) authors.

The most frequent medical discipline was general medicine (10.0% [95% CI 9.9%–10.2%]), 12.8% (95% CI 12.6%–13.0%) of publications mentioned the word CONSORT, and for 28.8% (95% CI 28.6%–29.1%) we identified a trial registration number. Articles were published in journals with a median impact factor of 2.93 (IQR 1.99–4.41).

The indicators ranking of the institution of the first author, ranking of the institution of the last author, financial support, number of words in the acknowledgment, and number of names in the acknowledgment were excluded from further analyses because of the large amount of missing data.

3.4. Univariable analyses

Results of univariable analyses are presented in Appendix 2. None of the indicators showed a statistically significant consistently positive or negative association for every type of QRP.

3.5. Multivariable models with data available from the trial publication

The indicators continent of first author, academic age of first and last author, academic presence of first author, and number of collaborations of first and last authors were excluded from multivariable models due to high correlations with other indicators in the model.

3.5.1. Risk of bias

In the multivariable models (Appendix 2), the following indicators were found to be associated with a lower probability of bias for at least three of four domains: a higher proportion of female coauthors, publications with the last author from Oceania, a more recent publication year, reporting a trial registration number, mentioning of CONSORT, higher journal impact factor, and publications from a large publisher. Publications with the last author

Table 1. Descriptive statistics of questionable research practices

Questionable research practice	Value[a]	Number of references for which this outcome was available
Probability of bias (as assessed by Robot Reviewer[b])		
Probability of bias in randomization	0.43 (0.18–0.59)	163,129
Probability of bias in allocation concealment	0.59 (0.40–0.71)	163,129
Probability of bias in blinding of patients and personnel	0.63 (0.40–0.75)	163,129
Probability of bias in blinding of outcome assessment	0.55 (0.44–0.64)	163,129
Modifications in primary outcome in public registration	3,615/16,349 (22.1% [95% CI 21.5–22.8])	16,349
Ratio of achieved compared to planned sample size	1 (0.98–1.04)	24,385
Statistical discrepancy	370/21,230 (1.7% [95% CI 1.6–1.9])	21,230

[a] Values are N (% [95% CI]) or median (25th–75th percentile).
[b] Robot Reviewer assesses the probability that a study has bias rather than dichotomizing it into high or low risk of bias. We here present the median probabilities. See methods section for definitions of questionable research practices.

Table 2. Results from multivariable reduced models that include indicators of QRPs available upon submission of an article but before publication (model 2)

	Probability of bias		
Indicator	**Bias in randomization**	**Bias in allocation concealment**	**Bias in blinding of patients and personnel**
Gender of first author: male	−0.014 (−0.044; 0.016)	−0.011 (−0.038; 0.016)	**−0.053 (−0.081; −0.024)**
Gender of last author: male	−0.023 (−0.055; 0.009)	−0.016 (−0.042; 0.010)	**−0.064 (−0.093; −0.035)**
Proportion of female authors	**−0.144 (−0.200; −0.088)**	**−0.053 (−0.103; −0.003)**	*0.071 (0.019; 0.123)*
Number of authors	**−0.009 (−0.012; −0.005)**	**−0.007 (−0.010; −0.004)**	0.002 (−0.002; 0.005)
Continent of last author: Africa	**−0.142 (−0.237; −0.048)**	**−0.122 (−0.202; −0.041)**	0.072 (−0.018; 0.163)
Continent of last author: Asia	−0.018 (−0.049; 0.013)	*0.101 (0.074; 0.127)*	0.013 (−0.017; 0.042)
Continent of last author: Middle and South America	0.063 (−0.008; 0.135)	0.056 (−0.004; 0.115)	**−0.073 (−0.139; −0.006)**
Continent of last author: North America	*0.110 (0.083; 0.136)*	*0.104 (0.082; 0.126)*	**−0.026 (−0.049; −0.002)**
Continent of last author: Oceania	**−0.217 (−0.275; −0.159)**	**−0.225 (−0.273; −0.177)**	0.008 (−0.041; 0.057)
Number of countries	*0.020 (0.011; 0.029)*	0.003 (−0.005; 0.011)	**−0.044 (−0.052; −0.035)**
H−index of first author	**−0.001 (−0.001; −0.000)**	**−0.001 (−0.001; −0.000)**	**−0.003 (−0.004; −0.003)**
H-index of last author	*0.001 (0.000; 0.001)*	0.000 (−0.000; 0.001)	**−0.001 (−0.001; −0.000)**
Academic age of last author: sqrt	0.002 (−0.007; 0.010)	−0.000 (−0.007; 0.007)	*0.018 (0.010; 0.025)*
Number of institutions: sqrt	**−0.060 (−0.081; −0.039)**	**−0.066 (−0.085; −0.048)**	0.001 (−0.019; 0.021)
Percentage of positive words in abstract	0.036 (−0.019; 0.090)	0.043 (−0.004; 0.089)	*0.063 (0.013; 0.112)*
Percentage of negative words in abstract	0.043 (−0.045; 0.131)	0.008 (−0.067; 0.084)	0.005 (−0.075; 0.086)
Medical discipline[a]	Appendix 2	Appendix 2	Appendix 2
Mentioning of CONSORT	**−0.394 (−0.427; −0.361)**	**−0.318 (−0.346; −0.290)**	0.029 (−0.001; 0.058)
Trial registration	**−0.437 (−0.462; −0.412)**	**−0.417 (−0.438; −0.395)**	**−0.155 (−0.178; −0.133)**

All values are regression coefficients from multivariable models for 1 unit increase in the indicator. For all outcomes, except for the ratio of achieved compared to planned sample size, negative values are good, that is, less questionable and more responsible (e.g., lower risk of bias). Statistically significant values are marked in bold (lower risk of QRP) and italics (higher risk of QRP). For all categorical variables regarding the continent of authors or journal, Europe is taken as the reference category.

Abbreviation: Sqrt, square root.

[a] The indicator 'medical discipline' was included in the model but removed from the table to improve readability.

[b] The indicator having a trial registration could not be included in the models predicting modifications in the outcome and ratio of achieved compared to planned sample size, as these outcomes were only available for trials that have a trial registration.

from North America were associated with a higher probability of bias than publications with the last author from Europe. Compared to the category of general medicine, many medical disciplines were associated with either consistently higher (e.g., hematology) or lower (e.g., anesthesiology) probability of bias.

3.5.2. Modifications in the primary outcome

Publications with the last author from North America or Oceania had a higher risk of modifications in the outcome than publications with the last author from Europe. Also, a higher h-index of the first and last authors and having more institutions involved were associated with a higher risk.

3.5.3. Ratio of achieved compared to sample size

A higher number of countries involved were associated with a higher ratio of achieved to planned sample size (i.e., higher achieved sample size). Having more institutions involved was associated with a lower ratio.

3.5.4. Statistical discrepancy

Publications reporting a trial registration number were associated with a lower risk of statistical discrepancy.

We found conflicting associations or found no associations with consistent directions over multiple QRPs for the research experience of the last author (i.e., active research years), use of positive or negative words in abstracts, or changes in journal impact factors.

3.6. Multivariable models restricted to data available upon submission to a journal (i.e., before trial publication)

Models that contain indicators available upon submission of an article to a journal but before publication (model 2) showed similar trends to the models with postpublication indicators (Table 2 and Appendix 2). Again, a higher proportion of female authors was associated with a lower probability of bias (except for the domain blinding

Probability of bias	Modifications in outcome	Ratio of achieved compared to target sample size	Statistical discrepancy
Bias in blinding of outcome assessment			
−0.008 (−0.025; 0.009)	0.041 (−0.171; 0.254)	0.009 (−0.028; 0.045)	0.001 (−0.589; 0.591)
−0.020 (−0.037; −0.002)	−0.069 (−0.301; 0.162)	−0.002 (−0.041; 0.037)	0.156 (−0.439; 0.751)
−0.047 (−0.080; −0.014)	−0.105 (−0.602; 0.391)	−0.017 (−0.090; 0.055)	0.847 (−0.495; 2.189)
−0.003 (−0.005; −0.001)	0.011 (−0.012; 0.033)	0.001 (−0.002; 0.004)	−0.003 (−0.074; 0.069)
−0.003 (−0.060; 0.053)	−0.269 (−0.938; 0.400)	−0.003 (−0.087; 0.080)	−0.402 (−2.880; 2.076)
0.011 (−0.007; 0.029)	−0.106 (−0.320; 0.107)	0.015 (−0.015; 0.045)	−0.061 (−0.912; 0.789)
−0.068 (−0.109; −0.026)	0.089 (−0.350; 0.529)	0.013 (−0.053; 0.080)	−0.605 (−3.021; 1.812)
0.003 (−0.012; 0.018)	*0.284 (0.141; 0.427)*	−0.013 (−0.038; 0.012)	−0.071 (−0.589; 0.448)
−0.123 (−0.154; −0.092)	*0.371 (0.092; 0.651)*	−0.032 (−0.082; 0.019)	0.054 (−0.871; 0.979)
0.001 (−0.005; 0.006)	0.030 (−0.018; 0.079)	**0.012 (0.005; 0.019)**	0.064 (−0.111; 0.240)
−0.000 (−0.001; −0.000)	*0.005 (0.001; 0.009)*	0.001 (−0.000; 0.001)	−0.002 (−0.016; 0.012)
−0.001 (−0.001; −0.001)	*0.004 (0.001; 0.007)*	−0.000 (−0.001; 0.000)	−0.001 (−0.013; 0.012)
0.009 (0.004; 0.014)	−0.039 (−0.097; 0.018)	0.000 (−0.009; 0.010)	−0.017 (−0.206; 0.173)
−0.018 (−0.030; −0.005)	*0.171 (0.015; 0.328)*	−0.019 (−0.041; 0.002)	−0.196 (−0.644; 0.251)
0.017 (−0.015; 0.048)	−0.125 (−0.533; 0.284)	−0.015 (−0.075; 0.045)	0.355 (−0.400; 1.109)
0.047 (−0.004; 0.098)	−0.209 (−0.759; 0.341)	−0.023 (−0.121; 0.076)	0.595 (−0.479; 1.668)
Appendix 2	Appendix 2	Appendix 2	Appendix 2
−0.114 (−0.133; −0.096)	0.036 (−0.145; 0.217)	−0.019 (−0.045; 0.007)	−0.237 (−0.893; 0.419)
−0.193 (−0.207; −0.178)	Not applicable[b]	Not applicable[b]	**−0.685 (−1.260; −0.110)**

of participants and personnel where the reverse was found). The h-index of first and last authors was associated with a higher risk of primary outcome modifications in the public registration. Studies that mentioned CONSORT and reported a trial registration number were consistently associated with a lower probability of bias.

Differences with the full models were that a higher number of authors were associated with a lower probability of bias. The h-index of the first author was associated with a lower probability of bias in all four domains in the reduced models but not in the full models.

3.7. Multivariable models with data available upon trial registration

The models that contained indicators available upon trial registration (but before trial completion) only included the indicators gender of last author, the continent of last author, h-index of last author, academic age of last author, and medical discipline. For the four domains, almost all of these indicators were associated with probability of bias (Appendix 2).

3.8. Explained variance

In terms of explained variance, the reduced models had lower values than the full models (Appendix 2). The highest R^2 values were seen for full models predicting bias in allocation concealment and bias in randomization (0.138 and 0.122, respectively). The lowest R^2 was found for the reduced model, using data available during trial design and registration, predicting ratio of achieved to planned sample size (R^2 0.002).

4. Discussion

We investigated the association between trial characteristics and QRPs and found associations with QRPs for many of the studied indicators (e.g., gender, publication year, h-index, mentioning of CONSORT). The most robust indicators that were consistently associated with a lower risk of several QRPs included (1) a higher journal impact factor, (2) a journal from a large publisher (such as Elsevier or Springer), (3) having a trial registration, and (4) mentioning of the CONSORT reporting guideline. We could not identify any association between the percentage of positive or negative words in an abstract and the risk of QRP.

4.1. Comparison to previous literature

Several researchers mapped the frequency of QRPs [4,14,15,41,42]. In our study, we observed that *P* values did not correspond to the given test statistics in 1.7% of the articles. This is similar to a previous publication in

which inconsistencies were observed in 1.6% of studies [43]. It is, however, lower than a study that found statistical discrepancy in 38% of articles published in 2001 in Nature and 25% in the BMJ [44]. A possible explanation for these differences is that *P* values were manually collected and checked in that study while we made use of an automated script which might have missed large parts of the *P* value test-statistic combinations while a manual check is not restricted to this specific type of format and therefore could identify more of these combinations.

We found that gender (higher proportion of female authors) was associated with lower probability of bias. Previous research has shown that female authors tend to report more conservative effect sizes but also that female first authors are more likely to overestimate effects [11,13]. A recent survey conducted among Dutch academics has shown that a lower academic rank and a female gender were associated with a lower responsible research practice score (i.e., less responsible) [45]. Many studies have focused on the association between impact factor and QRPs. In agreement with our findings, higher impact factors were found to be associated with a lower probability of bias [10,36,46] but also with better reporting [47]. Surprisingly, we found that a higher h-index was associated with a higher risk of primary outcome modifications. We hypothesize this might be related to the fact that h-index is partly driven by the number of publications and therefore, more experienced researchers often have a higher h-index. In the past years, there has been a change in research culture, with more attention to responsible conduct or research [48].

4.2. Recommendations for future research

Although it is not possible to draw conclusions about causal relations based on our study, our results might inform future strategies to identify those RCTs at a high risk of QRPs. In this explorative study, we showed there are associations between indicators and the presence of QRPs. However, the low explained variance of our regression models suggests these cannot be used for individual risk predictions. Furthermore, this suggests there is still a lot of variation between studies that could not be explained by the indicators we studied. A future step could be to study more indicators from other QRP domains to inform a prediction model that can be applied to flagging trial protocols, manuscripts, or articles with a high risk of QRPs which need to be scrutinized more closely. It should also be noted that such a prediction model should not be used on its own but always combined with further (manual) examination of the existence of QRPs.

Two indicators that consistently showed associations with a lower probability of bias across all four studied QRPs are reporting a trial registration number and mentioning CONSORT in a manuscript, which both relate to strategies aimed at enhancing usability of study results. This confirms the importance of requiring trial registration and complying with reporting guidelines.

Surprisingly, we found that a higher number of countries was associated with a higher ratio of achieved to planned sample size, while a higher number of institutions was associated with a lower ratio. Furthermore, having a last author from Oceania was associated with a lower probability of bias and a higher risk of modifying outcomes. Further research can focus on finding out whether these associations can be confirmed independently or whether they are just chance findings.

4.3. Strengths and limitations

We evaluated QRPs in RCTs covering a large proportion of all published RCTs included in PubMed. Using automated data collection, we were able to obtain a large amount of data.

Our analysis also has several limitations. First, we have not manually screened all included and excluded articles. This allowed us to include a large number of RCTs, but it is possible that we have included articles that do not report an RCT, that we have included multiple publications about the same RCT, and that we have excluded articles that were reporting an RCT. Especially, poorly written articles were more likely to be misclassified. Articles for which no PDF was available had to be excluded, which might have led to a selective set of RCTs included in our analyses. For the QRP related to selective reporting of outcomes, we restricted ourselves to RCTs registered on ClinicalTrials.gov, while many European trials are only registered in European Union Drug Regulating Authorities Clinical Trials Database. This might have led to selective exclusion of European RCTs for this QRP.

Second, the automated data collection might have led to misclassification of indicators and QRPs. We expect that this has diluted the associations. Due to the high amount of missing data, caused by problems with automatic data collection, we had to exclude five of our predefined indicators. Furthermore, only 21,230 articles were available for evaluating statistical discrepancy because in the other articles we were not able to identify a *P* value test-statistic combination in the required format. For the outcome risk of bias, we relied on Robot Reviewer software. Evaluations of this tool indicated moderate to good agreement with human reviewers for the random sequence generation and allocation concealment domains; however, a varying agreement was found for the domains on blinding [14,23,49].

Third, we planned to collect information on the quality of reporting, defined as adherence to the CONSORT reporting guideline as determined with software developed by StatReviewer [50], but this turned out not possible due to time constraints of the software developers.

Finally, although we applied a Bonferroni correction, we still tested hundreds of indicator−outcome associations. Furthermore, the large size of our dataset might have resulted in statistically significant but irrelevant associations (i.e., small effect size).

5. Conclusion

Our analyses show that gender, author continent, publication year, h-index, mentioning of CONSORT, trial registration, medical discipline, and journal impact factor were all associated (in different directions) with the risk of QRPs.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2022.11.020.

References

[1] Bouter LM. Fostering responsible research practices is a shared responsibility of multiple stakeholders. J Clin Epidemiol 2018;96:143−6.

[2] Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. Circ Res 2015;116:116−26.

[3] Riley JW. Proceedings of the thirteenth conference on public opinion research. Public Opin Q 1958;22(2):169−216.

[4] Banks GC, Rogelberg SG, Woznyj HM, Landis RS, Rupp DE. Evidence on questionable research practices: the good, the bad, and the ugly. J Bus Psychol 2016;31:323−38.

[5] Wicherts JM, Veldkamp CL, Augusteijn HE, Bakker M, Van Aert R, Van Assen MA. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. Front Psychol 2016;7:1832.

[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] Hermerén G. The European code of conduct for research integrity; 2017:161. Available at https://allea.org/code-of-conduct/. Accessed December 23, 2022.

[8] National Academies of Sciences, Engineering, and Medicine. Fostering Integrity in Research. Washington, DC: The National Academies Press; 2017. Available at https://doi.org/10.17226/21896.

[9] Final rule for clinical trials registration and results information submission (42 CFR Part 11). Available at https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission. Accessed November 11, 2021.

[10] Zhou J, Li J, Zhang J, Geng B, Chen Y, Zhou X. The relationship between endorsing reporting guidelines or trial registration and the impact factor or total citations in surgical journals. PeerJ 2022;10:e12837.

[11] Fanelli D, Costas R, Ioannidis JP. Meta-assessment of bias in science. Proc Natl Acad Sci U S A 2017;14:3714−9.

[12] Sosa JA, Mehta P, Thomas DC, Berland G, Gross C, McNamara RL, et al. Evaluating the surgery literature: can standardizing peer-review today predict manuscript impact tomorrow? Ann Surg 2009;250(1):152−8.

[13] Maggio LA, Dong T, Driessen EW, Artino A. Factors associated with scientific misconduct and questionable research practices in Health professions education. bioRxiv 2018;2:74−82.

[14] Vinkers CH, Lamberink HJ, Tijdink JK, Heus P, Bouter L, Glasziou P, et al. The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. PLoS Biol 2021;19(4):e3001162.

[15] Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K. Boutron I.,et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. BMJ 2017;357:j2490.

[16] World Health Organisation. Code of conduct for responsible research. 2017. Available at. https://www.who.int/docs/default-source/wpro−-documents/regional-committee/nomination-regional-director/code-of-conduct/ccrr.pdf?sfvrsn=b2cb450_2&ua=1. Accessed November 13, 2021

[17] Damen J, Lamberink HJ, Tijdink JK, Otte WM, Vinkers C, Hooft L, et al. Predicting questionable research practices in randomized clinical trials. Open Science Framework 2018. Available at https://osf.io/27f53/. Accessed November 13, 2021.

[18] R: A language and environment for statistical computing [program]. Vienna, Austria: R Foundation for Statistical Computing; 2016.

[19] Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637−9.

[20] Inferring gender from names on the web. A comparative evaluation of gender detection methods. Proceedings of the 25th International Conference Companion on World Wide Web. Switzerland: International World Wide Web Conferences Steering Committee; 2016.

[21] Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc 2016;23(1):193−201.

[22] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

[23] Gates A, Vandermeer B, Hartling L. Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool. J Clin Epidemiol 2018;96:54−62.

[24] Lamberink HJ, Vinkers CH, Lancee M, Damen JAA, Bouter LM, Otte WM, et al. Clinical trial registration patterns and changes in primary outcomes of randomized clinical trials from 2002 to 2017. JAMA Intern Med 2022;182(7):779−82.

[25] Georgescu C, Wren JD. Algorithmic identification of discrepancies between published ratios and their reported confidence intervals and P-values. Bioinformatics 2018;34(10):1758−66.

[26] Epskamp S, Nuijten M. statcheck: extract statistics from articles and recompute p values. R Package version 01 0. 2016. Available at https://cran.r-project.org/web/packages/statcheck/index.html. Accessed December 23, 2022.

[27] Marshall IJ, Kuiper J, Banner E, Wallace BC. Automating biomedical evidence synthesis: RobotReviewer. Proc Conf Assoc Comput Linguistics Meet; 2017:7−12.

[28] Campbell LG, Mehtani S, Dozier ME, Rinehart J. Gender-heterogeneous working groups produce higher quality science. PLoS One 2013;8:e79147.

[29] Otte WM, Tijdink JK, Weerheim PL, Lamberink HJ, Vinkers CH. Adequate statistical power in clinical trials is associated with the combination of a male first author and a female last author. Elife 2018;7:e34412.

[30] Smart RJ, Susarla SM, Kaban LB, Dodson TB. Factors associated with converting scientific abstracts to published manuscripts. J Craniofac Surg 2013;24(1):66−70.

[31] van der Steen JT, van den Bogert CA, van Soest-Poortvliet MC, Fazeli Farsani S, Otten RHJ, Ter Riet G, et al. Determinants of selective reporting: a taxonomy based on content analysis of a random selection of the literature. PLoS One 2018;13:e0188247.

[32] Khadem-Rezaiyan M, Dadgarmoghaddam M. Research misconduct: a report from a developing Country. Iranian J Public Health 2017;46(10):1374.

[33] Zwierzyna M, Davies M, Hingorani AD, Hunter J. Clinical trial design and dissemination: comprehensive analysis of clinicaltrials.gov and PubMed data since 2005. BMJ 2018;361:k2130.

[34] Vinkers CH, Tijdink JK, Otte WM. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. BMJ 2015;351:h6467.

[35] Frank RA, McInnes MDF, Levine D, Kressel HY, Jesurum JS, Petrcich W, et al. Are study and journal characteristics reliable indicators of ‘‘truth’’ in imaging research? Radiology 2018;287:215−23.
[36] Gluud LL, Sorensen TI, Gotzsche PC, Gluud C. The journal impact factor as a predictor of trial quality and outcomes: cohort study of hepatobiliary randomized clinical trials. Am J Gastroenterol 2005; 100(11):2431−5.
[37] Baerlocher MO, Newton M, Gautam T, Tomlinson G, Detsky AS. The meaning of author order in medical research. J Investig Med 2007;55:174−80.
[38] Zeileis A, Cribari-Neto F, Gruen B, Kosmidis I, Simas AB, Rocha AV, et al. Package ‘betareg’. R Package; 2016. Available at https://cran.r-project.org/web/packages/betareg/index.html. Accessed December 23, 2022.
[39] Harrell FE. Package ‘rms’. Vanderbilt Univ. 2019. Available at https://cran.r-project.org/web/packages/rms/index.html. Accessed December 23, 2022.
[40] Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1−67.
[41] Xie Y, Wang K, Kong Y. Prevalence of research misconduct and questionable research practices: a systematic review and meta-analysis. Sci Eng Ethics 2021;27(4):41.
[42] John LK, Loewenstein G, Prelec DJPs. Measuring prevalence questionable Res practices incentives truth telling 2012;23(5): 524−32.
[43] Nuijten MB, Hartgerink CH, Van Assen MA, Epskamp S, Wicherts JM. The prevalence of statistical reporting errors in psychology (1985−2013). Behav Res Methods 2016;48(4):1205−26.
[44] García-Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol 2004;4:13.
[45] Gopalakrishna G, Wicherts JM, Vink G, Stoop I, van den Akker OR, Ter Riet G, et al. Prevalence of responsible research practices among academics in The Netherlands. F1000Res 2022;11:471.
[46] Barbui C, Cipriani A, Malvini L, Tansella M. Validity of the impact factor of journals as a measure of randomized controlled trial quality. J Clin Psychiatry 2006;67(1):37−40.
[47] Elcivan M, Kowark A, Coburn M, Hamou HA, Kremer B, Clusmann H, et al. A retrospective analysis of randomized controlled trials on traumatic brain injury: evaluation of CONSORT item adherence. Brain Sci 2021;11(11):1504.
[48] Gerrits EM, Bredenoord AL, van Mil MHW. Educating for responsible research practice in biomedical sciences: towards learning Goals. Sci Educ (Dordr) 2022;31(4):977−96.
[49] Armijo-Olivo S, Craig R, Campbell S. Comparing machine and human reviewers to evaluate the risk of bias in randomized controlled trials. Res Synth Methods 2020;11(3):484−93.
[50] StatReviewer. Available at https://www.ariessys.com/views-and-press/newsletter-archive/june-2016/statreviewer-presents-at-emug/. Accessed November 26, 2019.