Faculteit Bedrijfseconomische Wetenschappen

master in de toegepaste economische wetenschappen

Masterthesis

Reporting biases in management research: A systematic map

Thibo Debats
Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische wetenschappen, afstudeerrichting beleidsmanagement

PROMOTOR :

Prof. dr. Stephan BRUNS

2020
2021

Faculteit Bedrijfseconomische Wetenschappen

master in de toegepaste economische wetenschappen

Masterthesis

Reporting biases in management research: A systematic map

Thibo Debats
Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische wetenschappen, afstudeerrichting beleidsmanagement

PROMOTOR :
Prof. dr. Stephan BRUNS

This master thesis was written during the COVID-19 crisis in 2020-2021. This global health crisis might have had an impact on the (writing) process, the research activities and the research results that are at the basis of this thesis.

Preface

Dear reader

This thesis was made to finish my academic career as a student at the university of Hasselt, and it concludes my master's degree in business economics: policy management and was the most important and biggest challenge of this year. The topic of reporting bias was not my first choice but turned out much more interesting than I could have hoped. I was introduced to a side of research that had valid criticisms towards modern science and that showed me that science is not as infallible as I thought it was. The writing of this thesis has thought me to be a lot more critical and skeptical when hearing the word scientific.

Luckily, I was well prepared to write this thesis thanks to the last four years of education that I received at the Uhasselt. Nevertheless, I could have never completed this task without the experience and insight of my promotor Prof. Dr. Stephan Bruns. It is for this reason that I hereby thank him for his patience and guidance throughout the writing of this thesis. I would also like to thank the university of Hasselt, seeing as the university made it possible for me to conduct this research, and in the process learn a lot about scientific literature and the field of research.

I wish you a lot of joy when reading this thesis.

Thibo Debats

Diepenbeek, August 2021

Samenvatting

Niet alleen academici maken gebruik van wetenschappelijke literatuur, ook investeerders en bedrijfsleiders raadplegen de literatuur om de juiste beslissingen te maken. Hun beslissingen hebben daarna een effect op de economische groei van een land (Joel & Brian, 2004; Romer, 1990; Ryan, 2012). Helaas is er tegenwoordig een geloofwaardigheidscrisis over deze wetenschappelijke literatuur aan de gang. Een oorzaak van deze crisis is dat resultaten vaak niet worden gerapporteerd als ze niet het "juiste" aantonen volgens de onderzoeker, ook wel reporting bias genoemd (McGauran et al., 2010). Onderzoekers in managementonderzoek besteden de laatste jaren steeds meer aandacht aan deze bias en het belang hiervan wordt steeds meer erkend (Bettis, 2012; Bruns et al., 2019; Rost & Ehrmann, 2015). Om bij te dragen aan de huidige literatuur tracht deze thesis te achterhalen wat er al is gepubliceerd over reporting bias in managementonderzoek. Daarom is de volgende onderzoeksvraag ontworpen aan de hand van het PECO-model (CEE, 2013): *"Welke literatuur is er gepubliceerd omtrent het effect van reporting bias op managementonderzoek?"* Hiervoor is een systematische map gemaakt volgens de richtlijnen van ROSES (ROSES, 2017), deze methode is gekozen omdat systematische mappen zijn ontworpen om aan de hand van een protocol op een objectieve manier een overzicht te geven van literatuur (Haddaway, Macura, Whaley, & Pullin, 2018; Pae, 2015). Eerst werd er een zoekstring ontworpen aan de hand van de belangrijkste termen uit de literatuur. Vervolgens werden de inclusiecriteria vastgelegd die bepalen of een artikel relevant is. Ten slotte, werd ook de dataverwerking op voorhand vastgelegd, zoals vereist is volgens het protocol.

Eenmaal het protocol klaar was, werd de zoekstring gebruikt in een database genaamd Web of Science Core Collection, dit gaf 95 artikels die relevant konden zijn. Tijdens een screening op de titels en de samenvattingen van de artikels werden 76 artikels niet relevant gevonden volgens de inclusiecriteria. Vervolgens werden twee artikels verwijderd omdat ze niet beschikbaar waren en nog eens twee tijdens de full tekst screening omdat ze niet relevant waren. Dus 15 relevante artikels bleven over om te analyseren en om meta-data uit te extraheren. Deze analyse toonde aan dat onderzoek naar reporting bias in managementonderzoek een vrij nieuw fenomeen is ten opzichte van andere velden zoals geneeskunde, psychologie en economie. Het eerste relevante artikel was gepubliceerd in 2011, terwijl je in deze andere velden artikels vindt uit 1990, 1959 en 1999 (Ashenfelter, Harmon, & Oosterbeek, 1999; Dickersin, 1990; Sterling, 1959). Onderzoekers zijn wel steeds meer aan het publiceren over het onderwerp maar toch is het onderzoek stilgevallen in 2020 en 2021, een mogelijke verklaring hiervoor is de coronacrisis die is begonnen in maart 2020. Sinds 2017 bevatten de gepubliceerde studies ook meer empirische data om de theorieën te testen, waar in het begin vooral aan theorieontwikkeling werd gedaan. Maar de verhouding tussen ontwikkeling en testen is nog altijd niet goed, zo is het onmogelijk dat alle theorieën worden getest met de huidige verhouding. Dit probleem komt niet alleen voor in onderzoek naar reporting bias maar in hele veld van management (Edwards & Berry, 2010; Kacmar & Whitfield, 2000). Het veld van reporting bias in managementonderzoek heeft ook een relatief lage hoeveelheid aan literatuur, in dit veld leverde de zoekstring maar 95 artikels op. Maar in psychologie was dit 348, in geneeskunde was dit 419 en voor economie was dit 215. Managementonderzoek loopt dus achter op deze velden, maar hun

literatuur kan wel gebruikt worden om nieuwe ideeën op te doen of te beoordelen, zoals het reproduceerbaarheidsproject in de psychologie(Baker, 2015).

Een analyse naar de onderwerpen van de artikels toonde aan dat het meeste onderzoek kijkt naar drie bekende vormen van reporting bias, dit zijn HARKing, P-hacking en publication bias. HARKing werd 14 keer vermeld in 15 artikels, P-hacking 11 keer in 15 artikels en publication bias werd vermeld in bijna de helft met 7 van de 15 artikels. Daarnaast waren er twee minder bekende vormen van reporting bias gevonden namelijk reporting errors en externe druk, het laatste is een nieuwe potentiële vorm van reporting bias. Zoals verwacht werd de vorm reporting errors maar aangehaald in één artikel, dit artikel uit 2019 gaf dan ook aan dat er nog geen eerder onderzoek was uitgevoerd naar reporting errors (Bruns et al., 2019). Externe druk kwam twee keer voor, dit is ook niet verrassend want deze potentiële vorm van reporting bias is voor zover ik weet voor de eerste keer gedefinieerd in deze thesis. Het is gebaseerd op ideeën uit de twee artikels en volgt dezelfde gedachtegang als publication bias, maar werd in de literatuur nog niet bestempeld als een reporting bias. Meer ervaren onderzoekers moeten de beslissing maken of het nuttig is om dit op te nemen als een nieuwe vorm in de toekomst. Verder praatte de literatuur zowel over auteurs van artikels als over redacteurs en reviewers bij wetenschappelijke tijdschriften, omdat beide groepen reporting bias veroorzaken (Rupp, 2011). Beide groepen kregen bijna evenveel aandacht. Ten slotte, bekeek deze thesis ook of de artikels reporting bias op een directe of indirecte manier onderzochten. Direct onderzoek haalt echt nieuwe theorieën of empirische data aan rond reporting bias zelf, terwijl indirect onderzoek het probleem aanhaalt als een deel van een groter probleem. Indirect onderzoek kan wel nog altijd zinvol zijn om inzicht te krijgen in reporting bias. Beide vormen van onderzoek kwamen bijna even veel voor.

Vervolgens werden er kennistekorten en potentiële kennisclusters geïdentificeerd aan de hand van deze resultaten. Er is een kennistekort over reporting errors en deze thesis raadt ook aan om meer direct onderzoek uit te voeren. Daarnaast zijn er net kennisclusters in de onderwerpen van HARKing en P-hacking. Op deze onderwerpen kan men dus een systematische review uitvoeren om de aanwezige literatuur samen te zetten. Een betere optie is een levende systematische review, hierbij wordt regelmatig een update uitgevoerd, zo kan de kennis op één plaats en up-to-date worden bijgehouden terwijl de kennis groeit (Elliott et al., 2017). Dit is ook een optie voor het onderzoek naar de rol van auteurs, redacteurs en reviewers. Ten slotte, kunnen we ook nog concluderen dat er een positieve trend is in het aantal onderzoeken dat wordt uitgevoerd over reporting bias in managementonderzoek.

Er is dus nog maar beperkt onderzoek gebeurd naar reporting bias in managementonderzoek, dit betekent dat er geen zekerheid is over hoe erg het veld getroffen is. Wat doet twijfelen aan de kwaliteit en betrouwbaarheid van managementonderzoek. Zo is er een kans dat onderzoekers theorieën zijn aan het bouwen op foutief onderzoek dat toch als bewezen wordt aanzien. Meer onderzoek is dan ook nodig om schandalen zoals in de psychologie te vermijden (Shea, 2012). Verder heeft het gebrek aan zekerheid ook invloed op de geloofwaardigheid van evidence-based management en ook op opleidingen die zijn gebaseerd op wetenschappelijke literatuur. Dit onderzoek heeft net zoals alle onderzoeken gebreken. Zo was deze thesis beperkt in de tijd waardoor niet alle mogelijke bronnen van relevante literatuur zijn doorzocht. Verder was er ook maar één

reviewer voor alle artikels, dit is ongebruikelijk bij een systematische map. Normaal worden systematische mappen en reviews dan ook door meerdere onderzoekers uitgevoerd, maar in dit geval was dat geen optie. Ten slotte, zou de zoekstring ook vollediger zijn geweest als de term reporting error ook was opgenomen, maar dit had waarschijnlijk geen groot effect op het resultaat.

Table of contents

List of figures

List of tables

1. Introduction

The existence of high-quality and trustworthy scientific research is valuable to many parties in society, not only to researchers or the sciences themselves. Investors and firms often look for opportunities within the scientific world that can indicate what to invest in (Joel & Brian, 2004; Ryan, 2012). These investments and business decisions even influence the economic growth of a country (Romer, 1990). It is this importance of the scientific literature that makes the current credibility crisis very worrying. Multiple fields such as psychology (Collaboration, 2015), economics (Christensen & Miguel, 2018) and medicine (Stupple, Singerman, & Celi, 2019) are all questioning the validity of their body of literature.

Ioannidis (2005) even states that most study designs and settings make it more likely for research claims to be false than true, one reason for this is the bias that is present in a lot of studies. The article defines bias as the combination of several data, design, analysis and presentation choices that create certain research findings that should not be produced (Ioannidis, 2005). One such bias is reporting bias, a phenomenon in which the reporting of the results is dependent on the nature or direction of the results (McGauran et al., 2010). One of the first forms of reporting bias to be discovered was publication bias, the first evidence for it was found in psychology in 1959 (Dickersin, 1997). Specifically, publication bias is a form of reporting bias that occurs at the study-level, either because authors decide not to publish studies without statistically significant results or when reviewers and editors reject papers without statistically significant results (Rosenthal, Kleid, & Cohen, 1979). Closely related to this bias is P-hacking, the difference is that P-hacking occurs at the analysis-level, so in this case researchers conduct multiple analyzes but only report the statistically significant results (Simonsohn, Nelson, & Simmons, 2014). Another well-known form is HARKing, which happens when a researcher makes a hypothesis after the results are already known without reporting this process (Bettis, 2012; Kerr, 1998). Hypothesizing after the results is not always a bad practice, such behavior can be beneficial for science if the usage is mentioned in the discussion of an article (Hollenbeck & Wright, 2017). There are many more biases that can be seen as reporting biases such as citation bias, location bias and language bias (McGauran et al., 2010), but the last one this thesis will discuss are reporting errors. Reporting errors occur when the reported statistical information in an article is inconsistent. For example, when asterisks that denote significance levels differ from the significance levels based on the calculated t-value (Bruns et al., 2019).

There is evidence that these reporting biases are prevalent in many fields such as psychiatry (Andrade, 2021), economics (Brodeur, Cook, & Heyes, 2020) and strategic management (Harrison, Banks, Pollack, O'Boyle, & Short, 2017). These biases and questionable practices such as P-hacking and HARKing could be discovered if there was a self-correcting mechanism through replication studies. But the prevalence of replication studies is rather low, a study from 1996 determined that in 18 leading business journals between 1970 and 1991 less than 10% were replication or extension studies of empirical work in the disciplines of accounting, economics and finance areas, and 5% or less in marketing and management (Hubbard & Vetter, 1996). Most of the replications even contradicted the results of the original studies. An examination of three major marketing journals between 1974 and 1989 even revealed that no replications were published and that only 2,4% of the papers were replications with extensions (Hubbard & Armstrong, 1994). This later dropped to 1,2%

(Evanschitzky, Baumgarth, Hubbard, & Armstrong, 2007). The prevalence of the reporting biases combined with the replication crises brings up a lot of questions about the trustworthiness of the current scientific literature.

The awareness in management research about these problems and more specifically about reporting bias has risen over the last years. Bettis (2012) is a well-known paper that brings up the issue of HARKing as "the search for asterisks". Furthermore, he claims that a lot of researchers do not know that their research is based on questionable research practices. Rost & Ehrmann (2015) studied the omission or suppression of certain empirical findings. Bruns et al. (2019) claimed to be the first to research reporting errors in innovation research. Even the reliability of evidence-based management has been questioned, an article about this topic even mentions publication bias and HARKing as underlying reasons. It is for these reasons that this research attempts to add something to the current literature regarding reporting bias in management research (Kepes, Bennett, & McDaniel, 2014).

2. Methods

2.1. Research question

The objective of this research is to contribute to the existing literature regarding reporting bias in management research. Hitherto, there has barely been any systematic analyses of reporting bias in management research (Rost & Ehrmann, 2015). So, the specific objective is to give an overview of the published relevant literature that can be used to identify what the current literature is about and what it studied. The PECO-model was used to construct a first version of the research question in a structured way and to ensure that all the relevant parts of a research question are included (CEE, 2013). The different elements in the PECO-model are presented below.

Population: Management research
Exposure: Reporting bias
Comparator: No reporting bias
Outcome: Measure of prevalence of or the consequences of or reasons for reporting bias

These elements lead to the following question: What is the effect and occurrence of reporting bias in management research? This version was changed to reflect that this research attempts to give an overview of the published literature and does not attempt to answer the question. So, this research asks the primary research question: **What literature has been published regarding the effect of reporting bias in management research?** Furthermore, two secondary questions were made to make the primary question more specific and easier to answer. The first one is interested in the topics that the literature is about, the questions is as follows: *What topics of reporting bias are discussed in the published literature regarding reporting bias in management research?* The second question is interested in when the topic was studied and tries to determine how important the topic has been over the years. This is the second research question: *When was the relevant literature about reporting bias in management research published?*

2.2. Research strategy

The optimal research strategy was decided based on the research question and the possbile relevant approaches. The objective was to use the already existing literature to give an overview of what has already been discussed in the field, this narrows the possible methods down to literature reviews or maps. First, the decision was made to take a systematic approach instead of a narrative one, seeing as systematic approaches tend to be more comprehensive and objective. Furthermore, the objective and systematic selection criteria make the inclusion of articles less susceptible to selection bias (Pae, 2015). Then it was decided that a systematic map was more appropriate than a systematic review. A systematic map does not try to answer a specific question but tries to describe what research is out there, this fits with the attempt to give an overview of the literature (James, Randall, & Haddaway, 2016).

The ROSES guidelines were chosen to conduct the systematic map (ROSES, 2017). First, ROSES is more tailored towards observational data than PRISMA. Secondly, ROSES has stricter reporting guidelines as well, this seemed interesting as the topic of this research is reporting bias (Haddaway et al., 2018). However, the guidelines were not followed precisely because of the nature of this research, a master thesis limited to one year. Time restrictions made it impossible to publish a protocol and to wait on feedback from researchers in the field. The steps of the protocol were still followed to attain the structure that they provide. It is for this reason that deviations from these original plans will be discussed and explained, this will assure some of the transparency that is normally gained from first publishing a protocol.

2.3. **Searches**

One database was searched for studies using English search terms, this is the Web of Science Core Collection (Clarivate, n.d.). Some important search terms were selected from the relevant literature to test search strings. These terms were the names of reporting biases or frequently used words within the literature of reporting bias. Table 1 shows the terms that were tested for the number of articles, the used test string and number of results are also reported in this thesis (see Annex A).

Table 1: Search terms tested for the number of results.

Tested search terms	
"publication bias*"	(non-results OR "non results")
"reporting bias*"	("data fabrication*" OR data-fabrication*)
(data-snooping OR "data snooping")	"statistical methodology"
(p-hacking OR "p hacking")	"research malpractice*"
"selective reporting*"	"questionable research practice*"
"credibility crisis*"	"theory testing"
harking	Falsification

The time restrictions made it impossible to be completely comprehensive, so it was important to include the most relevant literature while still limiting the volume. Table 2 shows the terms that were selected to form the search string used in this research. It still contained most important terms despite the restriction.

Table 2: Search terms used in the final search string.

Used search terms
"publication bias*"
"reporting bias*"
"p-hacking" OR "p hacking"
"data-snooping" OR "data snooping"
"selective reporting*"
"questionable research practice*"

"harking"

Systematic reviews and analyses were also excluded from the search, by utilizing the Boolean "NOT" with the terms in Table 3.

Table 3: Terms used to exclude systematic reviews.

Exclusion search terms
"meta-analy*"
"meta analy*"
"systematic review"

The search was narrowed further down to only include articles that are relevant to the management field, because this is part of the research question. This was achieved using the field tag WC and the category management. All the important literature for this topic is published in one language, so the search was also restricted to English. Lastly, it was decided to only include articles. This results in the following complete search string (see Figure 1).

Figure 1: Complete search string.

> ((TS= (("publication bias*" OR "reporting bias*" OR "p-hacking" OR "p hacking" OR "data-snooping" OR "data snooping" OR "selective reporting*" OR "questionable research practice*" OR "harking") NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC = management)) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)

The results of this search string on Web of Science were downloaded as an excel file. This search was also the only source of literature used to find relevant articles, no other databases or websites were used. An estimation of the comprehensiveness of the results was made with the help of benchmark articles. These benchmark articles were included, this indicates that the search was comprehensive enough, the check was performed by Prof. Bruns. Lastly no search updates were required because this research did not last longer than two years.

2.4. **Screening**

All articles that resulted from the search were then screened for relevance using predetermined inclusion criteria, an article is deemed irrelevant if it breaks one of these criteria. The first exclusions happened while reading the title and abstract of each article, during this the inclusion criterion for topic were applied more loosely, to ensure that no articles were excluded needlessly. Afterwards the full text availability of each article was checked by collecting pdf files or links to online sources, articles that were unobtainable were removed. Lastly a relevance screening was conducted while reading the full text, in this phase all the articles that did not meet the inclusion criteria for topic were removed. There were several inclusion criteria used to assess the relevance of each article. All of them are discussed below together with an explanation why this criterion was chosen.

Inclusion criterion for topic: The topic was deemed relevant if the article mentioned any form of reporting bias. This could be HARKing, P-hacking, publication bias, reporting errors or questionable research practices. This means that other forms of bias were excluded. The reason for this criterion is that this research is about reporting bias not about bias in general.

Inclusion criterion for reporting bias: There was a criterion placed on what kind of reporting bias was included, the reporting bias had to be in research. Articles about reporting bias from managers or within projects or towards the public were excluded. This criterion was chosen because this research is specifically interested in reporting bias in research.

Inclusion criterion for fields: Only studies about management research or about scientific research in general were included. This was sometimes a difficult criterion to apply because management is made up of different subfields and is highly related to other fields. Fields that were closely related to management like leadership and organizational research were included, while fields like psychology, finance and business were excluded. This was important because other fields like psychology have done a lot more than management research to prevent the problem of bias. This could lead to very different problems regarding reporting bias. Studies about scientific research in general were included because it was deemed general enough to be applicable to management research.

Inclusion criterion for study design: All study designs were included except systematic reviews and meta-analyses. This decision was made because the important studies in these systematic reviews would most likely also occur in the search and to limit the number of relevant articles.

Inclusion criterion for language: The only language that was included was English. This is only mentioned for completeness because all the articles that resulted from the search were in English even if the language filter was not included in the search string.

There was no consistency checking in the application of these criteria, seeing as there was only one reviewer. The reason for each exclusion can be found in the annex to ensure transparency and to enable the reader of this thesis to judge the quality of the consistency (see Annex B and Annex D).

2.5. Critical appraisal

No critical appraisal was performed, this decision was made because a systematic map does not attempt to summarize or judge the literature but to give an overview. It is also difficult to assess the external validity of an article when a question has not been specifically formulated as with systematic reviews (James et al., 2016). A critical appraisal is also optional in the ROSES guidelines (ROSES, 2017).

2.6. Data extraction

The data was extracted in two stages, during the first stage important information about the articles was written down together with a short summary. This information was used to figure out what

differences there were between the articles and what variables were interesting to identify shortcomings in the current literature. Afterwards in the second stage, all articles were read again to code the articles, extract the necessary information, and make more comprehensive summaries. These variables were then put together into an excel file to give an overview of the extracted data. Table 4 shows the most important kinds of meta-data that were recorded. A full explanation about the criteria of judgement will be given in the results sections together with the results themselves. All needed data was present in the data, so there was no need to handle missing data.

Table 4: Important meta-data that was extracted, with extraction methods.

Variable	Criteria of judgement	Place of extraction
Year of publication	/	Web of science database
Discipline	How does the author call it in the article?	Article
Agent of focus	Is the relative focus on authors or journals?	Article
Theory development or testing	Does testing occur in the article?	Article
Direct or indirect research into reporting bias	Is reporting bias the focus of the article or part of a bigger problem?	Article
Topic	Coding into predefined categories.	Article

2.7. Data synthesis

After the data extraction the excel file with all the collected data was used to make frequency tables and graphs of the different variables. These were than utilized to identify big difference in the frequencies to spot where there was a need for more research or an abundance for a systematic review.

It is important to mention here that, the promotor of this thesis, Prof. Bruns coauthored one of the articles included in this dataset (Bruns et al., 2019). But he did not partake in the screenings or the coding of the articles, he did however judge the inclusion of all the articles after the title and abstract screening.

2.8. Deviation from the original protocol

There were two deviations from the original protocol. First it was planned to also include articles from fields related to management such as organizational psychology and finance, but the field of inclusion was narrowed down to limit the number of relevant articles because of time constraints.

Secondly, the original protocol did not intend to use a two-stage approach for the data extraction. But after the initial extraction it became clear that the quality could improve if another reading and extraction were conducted. The initial information was utilized to improve the categories that the meta-data was sorted into. Furthermore, it was decided to include summaries of each article in the

report. This is not common in systematic maps but was included to give a better overview of the data and because the relatively low number of articles allowed it.

3. Results

3.1. Search results

First, it is important to mention that there was only one reviewer who carried out all the coding and screenings. This means that no consistency checking happened throughout any of the results.

Figure 3 shows the process of collecting and screening the articles. The Web of Science search resulted in 95 articles. The number of duplicates was zero, this was expected because only one source was utilized to find relevant studies. All 95 results were screened on the title and the abstract, this screening happened utilizing the inclusion criteria. This initial screening was not extremely strict when looking at the inclusion criterion for topic. This was decided to ensure that articles were not excluded without a good reason and because the full text screening would be strict enough to remove articles that did not meet the criterion. In total 76 articles were excluded by this screening, the reasons for exclusion can be found in Figure 2 together with their frequency. The original 95 articles together with their individual reason for exclusion are reported in Annex B.

Figure 2: Reasons for exclusion during the first screening.
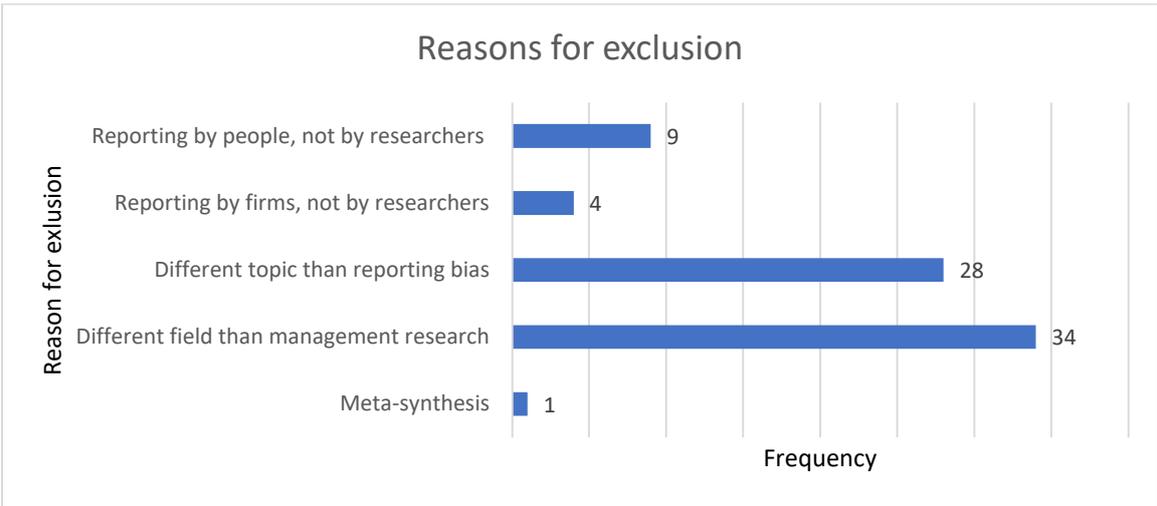
There were five reasons for excluding articles as Figure 2 shows. The two biggest reasons were the field and the topic of the article, together they account for 66 of the exclusions. Another 13 articles discussed a form of reporting bias that was caused by either firms or non-researchers. Lastly, one article was a meta-synthesis. It is important to note that all articles were only excluded for one reason, so it is possible that an article breaks multiple criteria but only one reason is given. In total 19 articles had to be retrieved in full text, two of them were unobtainable, these can be found in Annex C. The 17 remaining articles were read completely, during this two more were excluded, because they did not fit the inclusion criteria. These articles can be found in Annex D together with their reason for exclusion. In total 15 articles met the inclusion criteria and were available, all of them were included in the systematic map.

Figure 3: ROSES flow diagram for this systematic map.
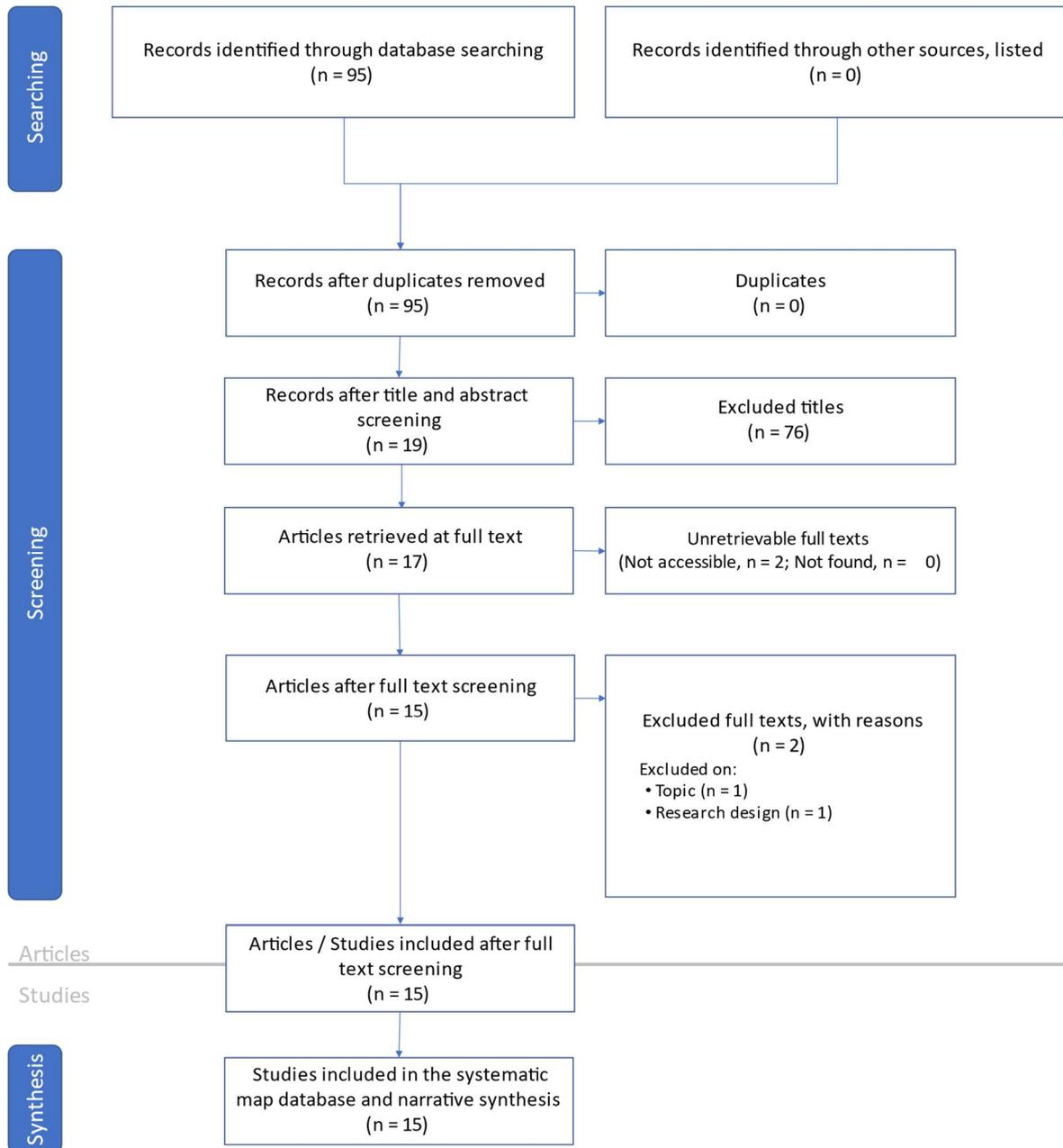
3.2. **Overview of studies**

The following section gives an overview of all 15 articles. First, there is Table 5 that contains the meta-data of all the relevant articles such as title, author, year of publication, research design and research question. The content for year, authors, title, and journal were taken from the Web of Science database, only the layout of these were changed to make it more uniform. The research questions were sometimes taken directly from the articles, but a lot of articles did not include a specific research question. In these cases, a research question was constructed to represent what question the study answered. Some articles did not attempt to answer a question but resembled a combination between an opinion piece and a narrative literature review, which is a good method to share the intuition and experience of a researcher (Pae, 2015), this was the case if the cell is empty. Both the research designs and the main results were made while reading and summarizing the articles, the objective was to give an idea of what the article was about without reading a whole text.

Afterwards, you can find a summary of each article. This is not a common practice within systematic maps, because this approach does not try to give a data synthesis in the same way as the systematic review (CEE, 2013). The summaries were included to offer a better overview of each article which might improve the overall overview of the widely differing studies, and to give an idea of why the article is relevant to this thesis. The low number of articles also allowed the summaries to be written and included.

3.2.1. *Overview of meta-data*

Table 5: Overview of meta-data for each article.

	Year	Author	Title	Journal	Research question	Research design	Main results
1	2020	Tourish, D; Craig, R	Research misconduct in business and management studies: Causes, consequences, and possible remedies	Journal of Management Inquiry	What is the frequency of and what are the reasons for retractions from scholarly journals in business and management studies? How do the experiences of editors, coauthors, and an admitted fraudster illuminate the pressures created by poor research practices?	131 retracted articles were examined to determine the reason for retraction. Six open ended interviews were conducted to get the opinion of three editors, two coauthors of retracted papers and one serial fraudster.	The biggest reason for retraction was data fraud with 51 articles. The editors felt powerless to proof bad practices or to combat them, the coauthors did not even know about the problems in their paper and the serial fraudster saw it as the game he had to play to be successful.
2	2019	Bruns, SB; Asanov, I; Bode, R; Dunger, M; Funk, C; Hassan, SM; Hauschildt, J;	Reporting errors and biases in published empirical findings: Evidence from innovation research	Research Policy	What is the prevalence of reporting errors and reporting bias in innovation research?	Articles from an important scientific magazine about research were examined for the number of reporting errors	the number of reporting errors per article was low, but 45% of the articles contained at least one reporting error. The caliper test indicates significant discontinuities

		Heinisch, D; Kempa, K; Konig, J; Lips, J; Verbeck, M; Wolfschutz, E; Buenstorf, G				and for reporting bias using the caliper test.	around the thresholds, so reporting bias is present.
3	2019	Gall, T; Maniadis, Z	Evaluating solutions to the problem of false positives	Research Policy	What is the best policy to prevent questionable research practices?	A game-theoretic analysis is used to look at different policies and their impact on mild and severe forms of questionable research practices.	Policies that focus on preventing mild forms of questionable research practices are most effective, because the policies for severe forms have a higher cost for the stakeholders.
4	2019	Crede, M; Harms, P	Questionable research practices when using confirmatory factor analysis	Journal of Managerial Psychology	What is the base rate of questionable analytic practices in management journals?	155 articles published in three top-tier organizational journals were reviewed.	Only 14,3% discussed which estimation method was used, none discussed why they used this method or discussed how missing data was handled.
5	2019	Hensel, PG	Supporting replication research in management journals: Qualitative analysis of editorials published between 1970 and 2015	European Management Journal	How was the issue of replication research framed and discussed in editorials by the top management journals	1901 published editorials were analyzed through coding.	Only 3,5% of the editorials mentioned replications studies, but there was a sharp increase in the last years. Most editorials do not reject replication studies,

					between 1970 and 2015?		but they are often seen as inferior to other research.
6	2018	Honig, B; Lampel, J	Reflections on scientific misconduct in management: Unfortunate incidents or a normative crisis?	Academy of Management Perspectives	How do we maintain a set of cultural values and mores governing our scientific activities against the competing pressures of prestige, competition, and entrepreneurial scholarship?	Different scholars from different backgrounds were invited to give their opinion in the form of an essay, nine essays made it to the end.	Many researchers agreed that there were problems with the incentive structure within research. This could be fixed with structural changes that encourage a deeper look at how good the research of a scholar is instead of the quantity of publications, when evaluating a scholar.
7	2017	Bergh, DD; Sharp, BM; Aguinis, H; Li, M	Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings	Strategic Organization	Is there a reproducibility crisis in strategic management research?	Several published studies were reproduced, and the reproduced results compared to the published results.	70% of the studies could not be reproduced at all. In 30% of the studies a hypothesis lost statistical significance when reproduced.
8	2017	Antonakis, J	On doing better science: From thrill of discovery to policy implications	The Leadership Quarterly		Narrative literature review, the used articles were not collected in a systematic way.	There are five main problems within leadership research that cause bias in the field. These can cause HARKing and P-hacking. More

							transparency is needed to prevent these problems.
9	2017	O'Boyle, EH; Banks, GC; Gonzalez-Mule, E	The chrysalis effect: How ugly initial results metamorphosize into beautiful articles	Journal of Management	What is the degree to which questionable research practices effect the stability and veracity of the extant literature?	Changes were tracked in hypotheses, data and results as a manuscript moved from defended dissertation to published article.	In the dissertations 45% of the hypotheses were statistically significant, but in the published articles 66%. There is an increase of 21 percent points, this indicates that questionable research practices were used.
10	2016	Starbuck, WH	60th anniversary essay: How journals could improve research practices in social science	Administrative Science Quarterly		Narrative literature review, the used articles were not collected in a systematic way.	The current editorial practices encourage or ignore questionable research practices like HARKing and P-hacking, to change this better reporting standards are necessary together with a cultural change within the social sciences.
11	2016	van Witteloostuijn, A	REGULAR ISSUE PAPER What happened to Popperian falsification? Publishing neutral and negative findings moving away from	Cross Cultural & Strategic Management		Narrative literature review, the used articles were not collected in a systematic way.	There is probably an abundance of biases within the field of business and management, a big cultural change is needed

			biased publication practices			across all stakeholders to solve this problem.
12	2015	Garud, R	Eyes wide shut? A commentary on the hypothesis that never was	Journal of Management Inquiry	Narrative literature review, the used articles were not collected in a systematic way.	HARKing can lead to untrustworthy scientific knowledge because it gives the impression that a hypothesis has already been tested, while this is not the case.
13	2014	Kepes, S; Bennett, AA; McDaniel, MA	Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice	Academy of Management Learning and Education	Narrative literature review, the used articles were not collected in a systematic way.	Evidence-based management depends on the trustworthiness of the underlying scientific knowledge. This might be a problem because of practices like HARKing and publication bias. To solve this new research practices are necessary.
14	2012	Bettis, RA	The search for asterisks: Compromised statistical tests and flawed theories	Strategic Management Journal	Narrative literature review, the used articles were not collected in a systematic way.	Repeated testing leads to unreliable articles to avoid this there needs to be more honest reporting, non-results should be published and there needs to be research into how big the problem is.

| 15 | 2011 | Rupp, DE | Ethical issues faced by editors and reviewers | Management and Organization Review | | Narrative literature review, the used articles were not collected in a systematic way. | Editors and reviewers should take action when they encounter questionable research practices, and they should phrase their recommendations in such a way that they do not promote HARKing. |

3.2.2. *Summaries*

1. Research misconduct in business and management studies: Causes, consequences, and possible remedies (Tourish & Craig, 2020)

This article is interested in the prevalence of questionable research practices and research misconduct such as an overreliance on corporate funding. A lot of scholars who use these practices rationalize their behavior, for example by the fact that they are widely used and by stating that they are just doing the same as everyone else. Two well-known questionable practices are HARKing and P-hacking, another form is THARKing but this is a better form of HARKing because it involves transparent reporting of the method. In the United States, 90% of the management faculty admits to HARKing. To gain more empirical evidence 131 retracted articles were selected and put in a database together with important variables such as authors, year of retraction and journal. In total, 154 reasons for retraction were found. Some articles were retracted for multiple reasons, and the biggest reason was data-fraud with 51 articles. The authors also conducted six open-ended interviews to assemble different perspectives, three with editors, two with coauthors of retracted papers and one with a serial fraudster. The serial fraudster said that he used these practices because of the nature of scientific community, he found that publication was needed to be successful and questionable practices made this easier. The coauthors of retracted papers experienced a lot of stress when problems were found in their paper, both claimed that they did not know that the problems were present, and that they trusted the other authors too much. The editors felt powerless when they encountered questionable research, seeing as it is very difficult to prove that something was intentionally fabricated. An editor that took action was even pressured by the author to stop the process and felt like his own career was threatened. Potential solutions are better reporting stands such as mentioning the used methods and analyzes, data records should be released on publication together with the methods of collection. Furthermore, the scientific community should be clear that P-hacking is not acceptable, and the term questionable research practices should be changed to deceptive research practices to make this clear.

2. Reporting errors and biases in published empirical findings: Evidence from innovation research (Bruns et al., 2019)

This article claims to be the first to focus on reporting errors in innovation research, at the intersection between economic and management research. A reporting error is defined as a difference between statistical significance indicators such as asterisks and the calculated significance based on the t-value. A difference can sneak it at multiple points of the publication process, such as during the first draft or during the final typesetting by the publisher. But the mistake might also be intentional to present results better than they are. Secondly, the article also adds to the empirical evidence regarding reporting bias in management research using the caliper test. This test works on the premise that there is a 50-50 chance that a p-value is below or above a threshold when the results are close to the threshold. This means that the results around the threshold should be continuously distributed, otherwise reporting bias such as HARKing, P-hacking or publication bias might be present. The examination showed that 15% of the articles contained at least one reporting error, but

the number of reporting errors per article is very low. However, in 25% of the articles one statistically significant result become insignificant or vice versa. The caliper test found big discontinuities at the thresholds, so we can assume that some reporting bias is present. In total, 371 articles from an influential research magazine were sampled, and both reporting errors and reporting bias were prevalent.

3. Evaluating solutions to the problem of false positives (Gall & Maniadis, 2019)

Science has big concerns about the prevalence of false positives in research. This influences how scientific research is viewed and might cause a confidence crisis in published findings. There is a call to improve the transparency on the used methods, this would give more insight and make replications possible. Questionable practices such as P-hacking could be detected if this call was answered. But malicious behavior would be a lot harder to prevent, because in this case scholars are intentionally attempting to present their research better than it is. The problems in scientific research have become so big that the UK parliament has looked if regulations are necessary to deal with the problem. The regulations would aim to increase the cost of using questionable research practices, it is however difficult to determine the best policy because the implications of each policy are hard to predict. This article makes a distinction between mild and severe forms of questionable research practices and attempts to conclude which forms the regulations should try to prevent. A game-theoretic approach is used and shows that it is better to focus on prevention of mild forms, seeing as the more severe policies come with bigger restrictions for scholars while the additional benefits will not compensate for the additional costs.

4. Questionable research practices when using confirmatory factor analysis (Crede & Harms, 2019)

The confirmatory factor analysis (CFA) is a commonly used tool within the organizational sciences, which makes it important to management research as well. CFA is a complicated tool that can be utilized in different ways, so it is important to describe the used model and justify why it was built a certain way. Otherwise, it is almost impossible to judge the quality of the analysis and by virtue the quality of the article. Therefore, this article describes the analytic and reporting practices related to CFA that omit disconfirming evidence from readers. To estimate the base rate of such practices, 155 published in three top-tier organizational journals were reviewed. The research looked if there was enough information provided to judge the quality of the model and to justify why it was used. The review showed that only 14,3% of the articles discussed which estimation model was used, and that none of them justified why the model was used. There was also no article that discussed how missing data were treated. These problems and omissions of information need not be intentional, bad reporting standards and ignorance might be a reason as well. But questionable practices such as HARKing or P-hacking might also contribute to the problem. The authors recommended some practices to prevent these problems such as including raw data and justifications for the used models. Furthermore, editors and reviewers could allow unideal models if the article presented the reasons for why it was used. This could decrease the pressure on authors to present a perfect model to be published.

5. Supporting replication research in management journals: Qualitative analysis of editorials published between 1970 and 2015 (Hensel, 2019)

The value and need for replication studies is well established in the naturel sciences but this is not the case within the social sciences. This causes a lack of replication within the social sciences, and there are good reasons to believe that the field of management is one of the worst. Without replication some published studies might be incorrect without us knowing, this lowers the overall quality of the field and questions the credibility of evidence-based management. A good argument for the need of replication studies, are the widespread questionable research practices such as HARKing and P-hacking. The use of these practices would be discovered if the studies were replicated properly and would also lower the incentives to use them for future research. This article is interested in the opinions of scientific journals towards replication studies, because their attitude towards them can have an influence on whether replications are published. The content of editorials published in 44 top management journals were examined on what they stated about replication studies. In total 67 editorials mentioned replication studies, which is only 3,5% of the used database. But there was a positive upward trend in the last years, so the issue has been gaining attention recently. Generally, these editorials stated that there was no ban or instant rejection of replication studies, so there is a chance for them to be published. However, many editorials indicated that replication research was seen as less valuable than other kinds of research. This shows that journals are partially causing the lack of replication studies in the field of management research.

6. Reflections on scientific misconduct in management: Unfortunate incidents or a normative crisis? (Honig et al., 2018)

Presently, there are a lot of worries about the state of the scientific literature. More and more articles are being retracted, while headlines of research misconduct are being released. Furthermore, publishing many articles is seen as the road to success, which puts a lot of pressure on scholars. All these events are a threat to the validity and trustworthiness of science. Science used to suffer from different problems such as the repression of information by religion. An example of this is Darwin who withheld his findings for 20 years to avoid a clash with the contemporary religious knowledge. However, this study claims that modern scientific problems come from publication pressure and not from totalitarian regimes. Scholars were invited to share their opinion in the form of an essay about this claim. First, invitation letters were sent to potential contributors, these contributors were selected in such a way to ensure a diversity of views. Nine scholars made it through the entire process and the essays went through multiple drafts. Afterwards, the essays were examined, and two distinctions were found between the essays. The first distinction was between essays that took a micro or a macro approach to the problem. The other distinction was that some essays were a lot more extensive in their recommendations to solve the problems. Three essays took a macro approach, while six took a micro approach. Some essays even brought up the problems of P-hacking and HARKing. There were a lot of recommendations such as tenure-recommendations should be reviewed on a deeper level instead on the quantity of publications. Development of a code of conduct

for reviewers and dedicated publishing space for replication studies were other solutions. But the overall message might be that too much is expected from researchers, seeing as researchers are just people who want to have a career as well. Expecting them to behave according to an impossible ethical standard might not be enough.

7. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings (Bergh, Sharp, Aguinis, & Li, 2017)

Currently, the scientific community is concerned that study findings cannot be replicated in fields such as psychology, economics and biomedicine. This article wants to know if this is also a problem in strategic management research, and it examines the reproducibility of articles published in the field. Reproducibility is defined as the ability of other researchers to find the same results when they reexamine and reanalyze the same data. Questions about the validity of the article arise when this is not possible. The empirical models of published articles were reproduced with the same data, then these results were compared with the results in the published articles. First, there was the problem that 70% of the articles did not include enough information for a partial or full reproduction. This indicates that there is a lack of information in a lot of studies. The remaining studies were reproduced, which showed that 30% of the published articles reported significant hypotheses that were not significant in the reproduction. This can be explained by questionable practices such as P-hacking or HARKing during the publishing process, better reporting practices are necessary to prevent this behavior. These results show that there is a chance that researchers are building on other studies that do not accurately represent the underlying data, which questions the credibility of the field. Strategic management needs to recognize the role of reproduction in the scientific process and take action to improve its utilization.

8. On doing better science: From thrill of discovery to policy implications (Antonakis, 2017)

Presently, papers can easily be published because of digital publishing and admissions, this means that flawed studies can avoid detection once they are published. Furthermore, impactful contributions to a field must often be novel and statistically significant, which leads to fierce competition between both authors and journals to publish the most impactful papers. This article claims that there are five specific diseases in scientific research. First, only statistically significant articles get published and once published it becomes very difficult to prove them wrong, which causes publication bias. Secondly, empirical journals expect novelty and innovation while ignoring non-results and replication studies. Furthermore, too many theories get developed but are never tested in subsequent articles. Fourthly, leadership research has too many informal definitions that are not agreed upon by all researchers, some formal definitions are needed to ensure stability in the field. The threshold of 0,05 for p-values is not restrictive enough and should be lowered to 0,001. Lastly, researchers do what they feel incentivized to do to successfully publish articles. These diseases lead to questionable practices such as P-hacking and HARKing. Such practices are preventable by changing the incentive structures. Furthermore, researchers should be evaluated more by how rigorous and useful their

research is instead of quantity of publication. There is also a need for more transparency when it comes to data, methods, and reporting.

9. The chrysalis effect: How ugly initial results metamorphosize into beautiful articles (O'Boyle Jr, Banks, & Gonzalez-Mulé, 2017)

The current reward structure in the management field produces an intense pressure to publish positives results. This entices scholars to engage in questionable research practices such as HARKing and data-manipulation. Furthermore, it leads to suppression of null-effects, and this causes bias in the scientific literature. This article utilizes the general strain theory to outline the means, motives and opportunities to increase the chances of publication independent of rigor and relevance. It also tracks 142 dissertations to examine the prevalence of questionable research practices in the extant literature. As stated above, statistically significant results are extremely important, and authors can achieve such results with various questionable practices. They can delete, add or alter data after the hypothesis tests, and delete or add variables or reverse the direction of the hypothesis to fit the data. Lastly, hypothesis can be created or dropped after testing. Researchers indicate that the anticipation of reviewer reactions and the expectations of journals are two reasons to use these methods. The relatively low reproduction rate and lack of reporting standards also create the opportunity to use them. The examined dissertations showed that 45% of the hypothesis were statistically significant, while this number increased to 66% when the dissertations were published. This 21 percent point increase shows that questionable researcher practices are certainly used, which indicates that there is uncertainty when judging management literature.

10. 60th anniversary essay: How journals could improve research practices in social science (Starbuck, 2016)

Eliminating bad research practices is very difficult as the null hypothesis test shows. The null hypothesis test is still a widely used test, while statisticians and methodologists have been advising other tests for 80 years. Researchers feel like using another test might threaten their success, because others might not understand or approve of these tests. This article focuses on the role of editors and on what they can do to prevent these bad practices and stimulate change and innovation. First, most researchers seem to disagree on the quality of an article, they only seem to agree about the top 20% of articles. This might be because reviewers cannot focus on all the aspects of an article, and some have different priorities in an article. This means that editors could appoint a certain reviewer to an article to get a certain response. This shows that the selected reviewer is an important factor in publishment and an editor should be aware of this when selecting reviewers. They should also insist on honest communication about the research methods and findings, most social science journals do not apply any policy for this. This could prevent methods like HARKing and P-hacking, both give a false appearance that good predictions were made before testing based on already existing theory. At this moment editors and reviewers even encourage these methods when giving feedback. Editors should promote more honest reporting and reliable methods; an example would

be explicitly mentioning whether a hypothesis was made before or after testing. Big changes are necessary but unlikely if textbooks and software packages remain unchanged.

11. REGULAR ISSUE PAPER What happened to Popperian falsification? Publishing neutral and negative findings Moving away from biased publication practices (Van Witteloostuijn, 2016)

Scientific research is conducted by humans, which means that bias will be present both on an individual and collective level. This bias can show up in multiple ways, some of these are explained in the article. First, there is verification bias, which is a form that exists because researchers focus a lot on proving that a theory is right through the generation of positive results. While not even attempting to prove that something is wrong. There is also novelty bias, this means that there is a big emphasis on new and original theories and not on validating older theories. Thirdly, is the normal science bias, which exists because new theories need to fit in with the dominant paradigms in the field. This article claims that management research is becoming a normal science and will be one in the near future. Another form is evidence bias, seeing as we cannot be sure about the quality of published evidence, there is a need for better reporting standards to prevent this bias. Lastly, there is market bias, a bias that exists because authors and journals want to be a part of highly-cited papers. This can cause questionable behavior when doing research. The situation is bad in several fields like economics, psychology and medicine and there are no reasons to believe that business and management research is different. Big cultural changes in the scientific community and all its stakeholders are necessary to prevent these biases. Some of these changes are letting go of the need for novelty and positive results in articles and reporting all the raw data and used analyses.

12. Eyes wide shut? A commentary on the hypothesis that never was (Garud, 2015)

This article is a reaction on an article that was previously published in the same journal named "the hypothesis that never was: uncovering the deceptive use of post hoc hypotheses". This previous article talks about the unethical behavior of presenting a hypothesis that was made after the analyses as if it was known before the analysis. This is deemed unethical behavior because it gives a wrong view of how the research was conducted and it also means that the hypothesis was never tested but presents itself like it has been tested. There is also a chance that the statistical significance is just random or spurious correlation. Nevertheless, these articles can contaminate the pool of knowledge that is based on empirical evidence. This in turn makes the whole field untrustworthy and has serious implications for the reputation and authority of the field. This reaction article talks about different forms of HARKing and about their impact on research. It also addresses that a lot of senior researchers think that this is a valid method and that they encourage new researchers to use it, bringing the problem to a new generation of researchers. Another cause are the high expectations placed on researchers to publish in order to advance their career and reputation within their university. This article is very aware that the problem of HARKing is not the method that is used but the lack of reporting what method is used, better reporting standards can prevent this problem.

13. Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice (Kepes et al., 2014)

The practices of evidence-based management and education are heavily dependent on the quality of the underlying scientific knowledge, seeing as these are the practices in which scientific knowledge is used for decision making and to educate students. This is problematic because the pressure, coming from the reward structure, on scholars to publish new and original research is very high and can lead to questionable research practices. While publication bias prevents the filtering of bad articles because non-results are not published, so not known about in the field. The field of management does not routinely check the trustworthiness of its knowledge, which makes the basis for evidence-based practices unstable. The educators can deal with this by being honest about the shortcomings of the fields. But also, by teaching students how to ask the right questions and by teaching how to be critical of evidence instead of just teaching the current consensus in the field. Managers should be taught to implement decisions based on evidence, but also to analyze the results and to adjust accordingly. Researchers can help as well by replicating more studies and the editorial review process should have more objective standards to guard the quality of published research. To summarize, big changes are necessary in the way that we teach and conduct research in the field of management, otherwise there will be a loss of credibility for evidence-based management and education.

14. The search for asterisks: Compromised statistical tests and flawed theories (Bettis, 2012)

A big problem within the field of strategic management is repeated statistical testing. This happens when multiple tests are done on the same dataset or for the same hypothesis. This article discusses three broad categories when it comes to repeated statistical testing. In the first one several researchers use the same databases across time, so it is very likely that sooner or later something will show up as statistically significant even if it is just random correlation. Secondly, there is the tuning of a model after it has already been tested, this is a common practice but can be problematic when it is not mentioned in the article or when there is no proper reason to do so. Lastly, there is searching for statistical significance and afterwards making a hypothesis that fits the data, in this case the hypothesis is never really tested. These practices make a flawed theory appear as high-quality proven science, which can be used as the basis for other theories. This problem is made even worse by publication bias because it makes it very hard to disprove these flawed theories. This can decrease the credibility and usefulness of the field, so it is in the self-interest of all scholars to fix these issues. The author mentions a couple potential solutions such as publishing non-results in journals. There is also a need for baseline data, to know how big the problem really is. PhD-students should receive better training, so that they do not make the same mistakes as the current researchers. Lastly, better statistical and reporting practices are needed such as reporting the number of conducted tests and the use of split samples.

15. Ethical issues faced by editors and reviewers (Rupp, 2011)

It is important for any field that there is an agreed upon set of ethical research practices, because it ensures that articles are of high quality. The management field still lacks some discussions of complex ethical issues, which are necessary to establish this set of ethical research practices. This article contributes to this problem by looking at the responsibilities of reviewers and editors within the management field. It starts off by explaining how the reviewing process works and by mentioning some already existing guidelines. Afterwards, there is an important section that discusses the responsibility of editors to not promote HARKing through their feedback, this can be avoided by wording their recommendations differently. Then it addresses the problem that editors are responsible for the success of their journal and for scientific progress, this creates tension that can lead to a lesser focus on the scientific progress. Reviewers should realize that they are not authors of the articles but judges of the quality, this means that they should give less specific feedback and recommendations about what should change. To summarize, it is important that editors and reviewers stop chasing their own self-interests and take action when they encounter questionable research practices.

3.3. Narrative synthesis

This section will present the collected data from all 15 articles. The data was turned into frequency tables and graphs to give a quick first glance of the state of management research. Every section starts with a reason for why these variables are relevant and how they were defined in this research. The accompanying text will discuss the results in more detail and will point out relevant and interesting implications or trends that we can find in the data. The different aspects that are presented are the same as those mentioned in the method section.

3.3.1. Agent *of focus*

The first reading showed that many different parties are involved in stimulating and producing reporting bias. Some examples are the universities, governments, the scientific journals and researchers themselves. This means that it is important to study each involved party to understand the problem completely. Therefore, the focus of each article was determined and was included in one of the following categories "authors & journals", authors and journals. A small explanation is given for each category.

Authors: This category includes all articles that mainly focus on the actions of researchers and authors of articles that can lead to reporting bias. But also, the reasons why they may engage in these actions. It is still possible and likely that journals were also brought up in these articles to discuss the problem, but they played a much less prominent role in the article.

Journals: This category includes the articles that looked mainly at editors or reviewers of scientific journals and at how they contributed to and caused reporting bias. This category was included because editors and reviewers can have a large influence on an article before it is published

(Lawrence, 2003; Rupp, 2011). Their demands or suggestions can lead authors to change the article or to omit information if it will improve the likelihood of publishment.

Authors & journals: This category contains all the articles that give an equal amount of attention to both authors of articles and the editors or reviewers at scientific journals.

Figure 4: The frequencies of each agent of focus.
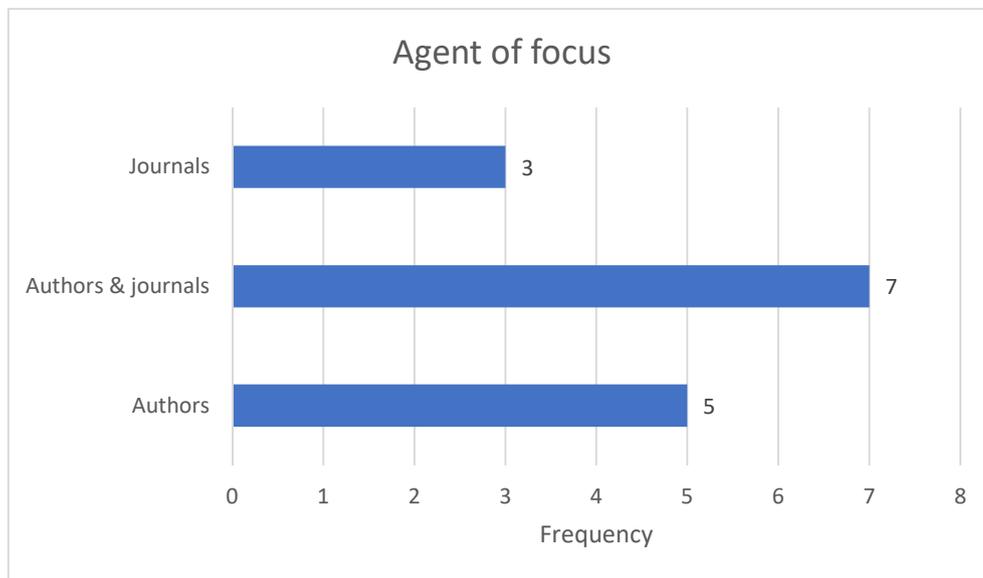
The focus of the article is only on the journal in three of the 15 articles as Figure 4 shows. For authors this number is five, so there is no big difference between the frequency of both categories. The biggest category is authors & journals, which has a frequency of seven articles. This indicates that the current research understands that both authors, editors and reviewers play a role in causing reporting bias, and that the behavior of all of them must be researched.

3.3.2. *Discipline*

Management research is a broad field, and it is better viewed as a class of professional disciplines such as medicine, law and engineering (Squires, 2001). It is for this reason that this research also looked at the discipline of each article. The discipline of the article was based on how the authors themselves defined it, so there is no hard barrier between the different disciplines. In total seven different disciplines were found in the relevant articles; these can be found in Figure 5 together with their frequency.

Figure 5: The frequencies of each discipline.
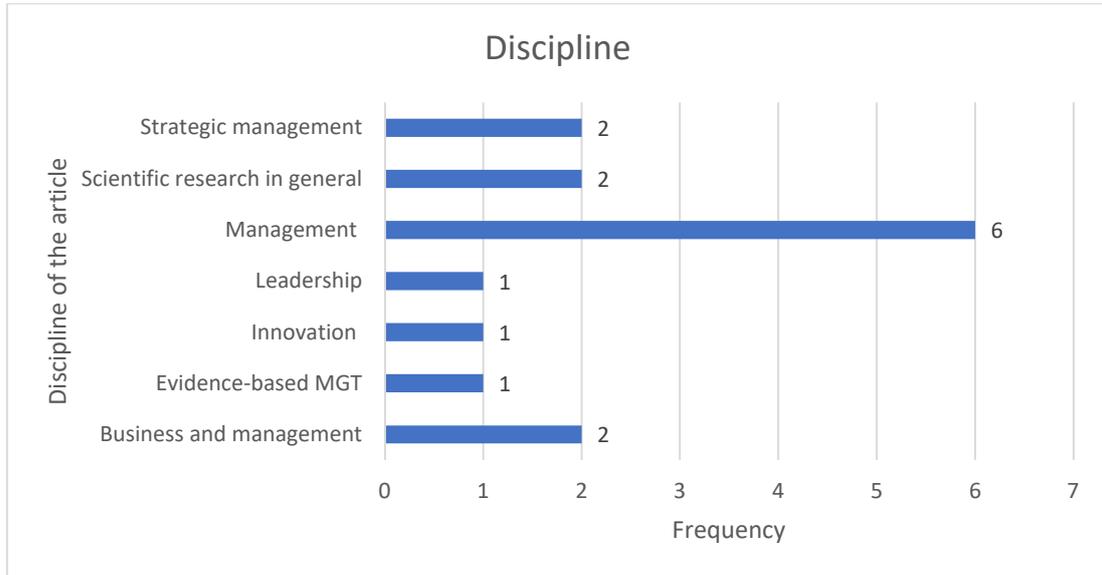
Three of the disciplines only showed up once, these were innovation research, evidence-based management and leadership. Then there were the disciplines that showed up twice: business and management, scientific research in general and strategic management. Lastly there is the discipline that was most used by the authors, this was management. This means that most articles that are relevant to reporting bias in management research use the term of management to classify their article. This is unsurprising because this is the broadest terms to describe the field of management, unlike strategic management or evidence-based management.

3.3.3. *Year of publication*

Another aspect of the articles that was looked at is the year of publication. This aspect was chosen because it gives a good overview of when the relevant literature was published and an idea of whether the topic is still studied and researched until this day.

Figure 6: The number of articles published every year.
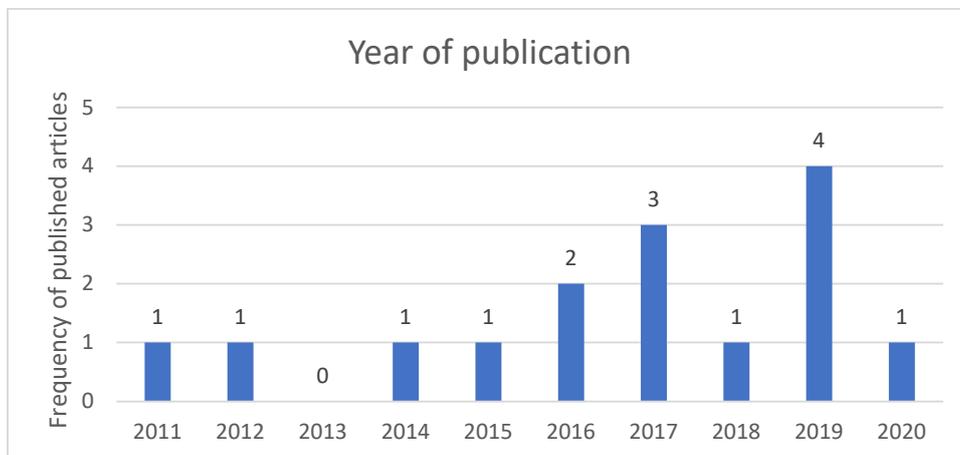
The first relevant article in this sample was published in 2011 (see Figure 6), this is quite noticeable because this research placed no restriction on when the articles were published. This would lead one to expect that there would be older articles because the problem of bias has been known in many fields for a long time (Dickersin, 1990; Sterling, 1959). So, research into reporting bias in management research is a relatively new phenomenon compared to other fields. In the first five years there were only four articles published, not even one a year. After this the number of publications started to rise from 2016 to 2019 when it reaches a peak of four articles in one year. But in 2020 it went back down to one a year and there has not been one so far in 2021, although the articles were collected in July, so half the year had already passed. One reason for this might be the pandemic of Covid-19, which started in March 2020 and might have caused a lot of attention to go to online teaching and research into more relevant topics like managing people who work from home.

3.3.4. *Theory developing and testing*

During the first reading of the articles, it became clear that a distinction could be made between two types of articles. Articles can either develop theories or they could test theories. This is an important distinction because we cannot be sure that an untested theory is true. It is also important to know the extend and scale of the problem certainly in this relatively new field of reporting bias in management research, so empirical evidence is needed. A balance between these two is important to make sure that the established theory is right and anchored in reality, while also pushing the theory forward. Therefore, the two following categories were made. A lack of theory testing has also been present in management research for quite a while (Edwards & Berry, 2010; Kacmar & Whitfield, 2000).

Theory developing: This category includes articles that only developed new theory based on the already existing literature, no further research was done, such as examining articles to estimate the prevalence of reporting bias or a regression. This means that without other theory testing articles these articles might be wrong and a bad diagnosis of the problem.

Theory testing: The articles in this category need to test a theory, this theory can be from another article or be developed in the same article. This might be done in different ways like a regression or through interviews or some literature might be scoured to determine the prevalence of reporting mistakes in it. These articles are of great importance, because they test whether a theory is based on actual evidence in the real world, and they deliver empirical evidence (Corbetta, 2003).

Figure 7: The frequencies of theory developing and testing articles.
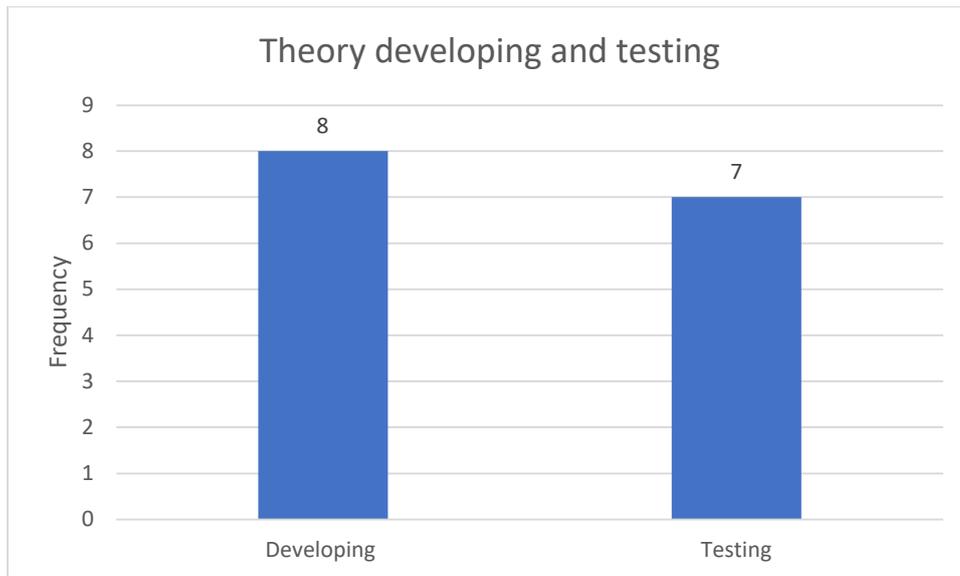
Figure 7 shows that the frequency of both articles is almost the same. Theory developing occurred eight times and theory testing occurred seven times. After reading all 15 articles and reading a lot of articles about the quality of research, I am inclined to say that this is not a good balance. First, there are fewer testing articles than developing articles, this would imply that every testing article would on average test more than one developing article. This would also mean that only one testing article would be enough to establish a theory as robust and well-documented science. Lastly, it would also imply that all these testing articles can be trusted, this goes against the premise of this thesis. So, there might be a strong need for more testing articles in the field of reporting bias in management research, before even more theory gets developed. Figure 8 was made to check how this ratio is evolving over time.

Figure 8: The frequencies of theory developing and testing articles over time.
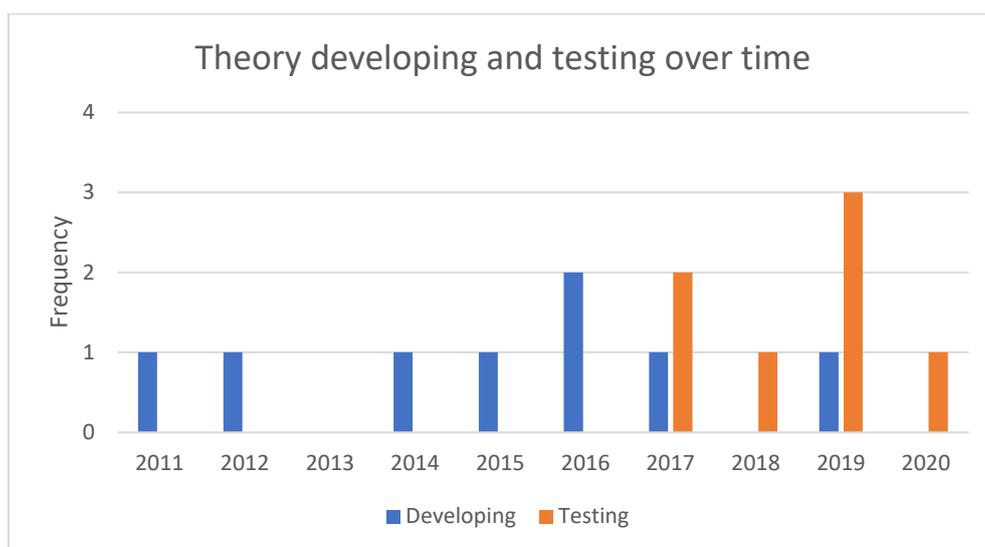
The first articles that were published about reporting bias were predominantly theory developing articles. This is to be expected because a problem needs to be established before it can be tested. It was only in 2017 when the first theory testing article was published but since than six more have appeared in only four years. This indicates that the trend is going in the right direction, and that the empirical evidence for reporting bias in management research has grown a lot in the last years.

3.3.5. *Direct or indirect research into reporting bias*

Reporting bias is a very broad kind of bias and can be divided in many different kinds (McGauran et al., 2010). It involves any kind of bias that exists because some form of information was not reported or was reported badly. This broadness leads to reporting bias being a part of many other problems. Research into these other problems can cause indirect research into reporting bias, which means that reporting bias was not the main focus, but implications can be made of its prevalence or impact. The following categories were used.

Direct: This category consists of the articles that directly investigated the problem of reporting bias or into a smaller part of it, like HARKing or P-hacking. It was not necessary that the entire article was about reporting bias but reporting bias had to be a part of the researched problem and had to add some new perspective or theory or had to quantify the problem of reporting bias.

Indirect: This contains the articles that might recognize some form of reporting bias as a part of their researched problem, but it does not directly try to research its occurrence or add some new theory. A lot of these articles might bring up standard solutions to the problems of reporting bias, but they do not really go deeper into the problem. This category was added because these articles frame reporting bias into bigger problems and might still be valuable to read to understand the complete problem.

Figure 9: The frequencies of articles that conduct direct and indirect research on reporting bias
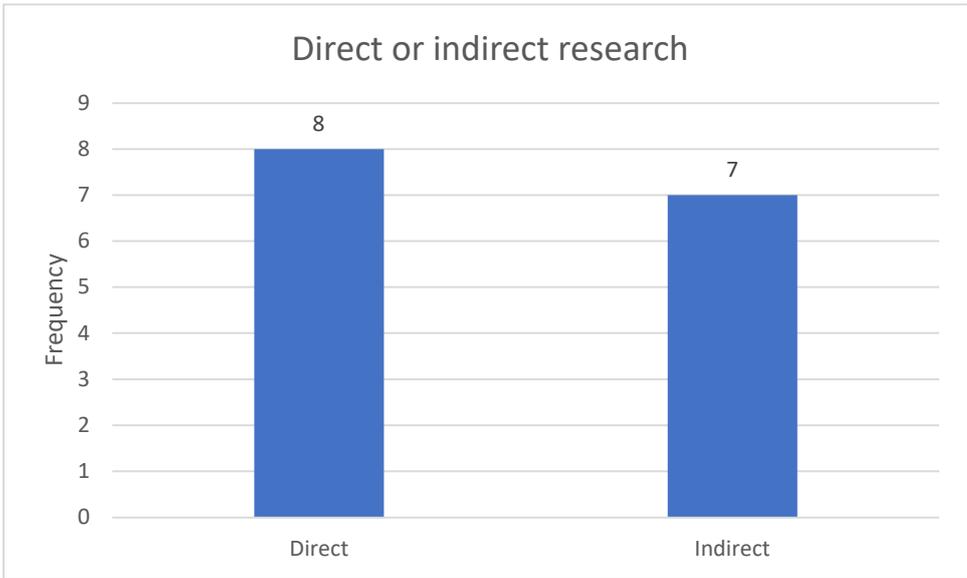
As Figure 9 shows the articles were almost evenly distributed. Eight articles looked directly at reporting bias and seven indirectly. This means that almost half the articles only bring up reporting bias a part of a bigger problem without drawing their own conclusions to solve the problem. A bigger focus on direct research might be more effective to tackle the problem directly.

3.3.6. *Topic*

Reporting bias can occur in many ways and has many different names. To bring some structure into what the research is about five categories were made. Five of these categories are very prevalent in the already existing literature and are well known as forms of reporting bias. The last category was discovered after reading the articles for the first time but could be seen as forms of reporting bias and is based on scientific research. The coding of these topics in each article can be found in Annex E.

Publication bias: A form of bias that occurs at the study-level, either because authors decide not to publish studies without statistically significant results or when reviewers and editors reject papers without statistically significant results (Rosenthal et al., 1979).

HARKing: First looking for statistically significant results and then coming up with a fitting theory to explain the results without reporting this process (Bettis, 2011), commonly known as HARKing (Hypothesizing After the Results are Known) (Kerr, 1998).

P-hacking: This form of bias happens at the analysis-level when researchers try multiple analyses and only report the estimates that are statistically significant. There might not always be a deliberate intention to mislead readers (Simonsohn et al., 2014).

Reporting error: A form of reporting bias where the reported statistical information in an article is inconsistent. For example, when asterisks that denote significance levels differ from the significance levels based on the calculated t-value (Bruns et al., 2019).

External pressure: Not publishing certain findings or theories because they do not fit in with the current way of thinking. In the past this would occur because there would be friction with the contemporary religious knowledge or because of the prevailing ideology like fascism or communism (Honig et al., 2018). In the modern world, a better example would be the normal science bias. This means that new theories need to fit in with the current paradigms to be accepted by the field (Van Witteloostuijn, 2016). This could be a form of reporting bias when authors decide not to make articles or not to publish them because they expect them to be rejected by the journals or that the article will not advance their career. The idea is very similar to publication bias.

Figure 10: The frequencies of each topic.
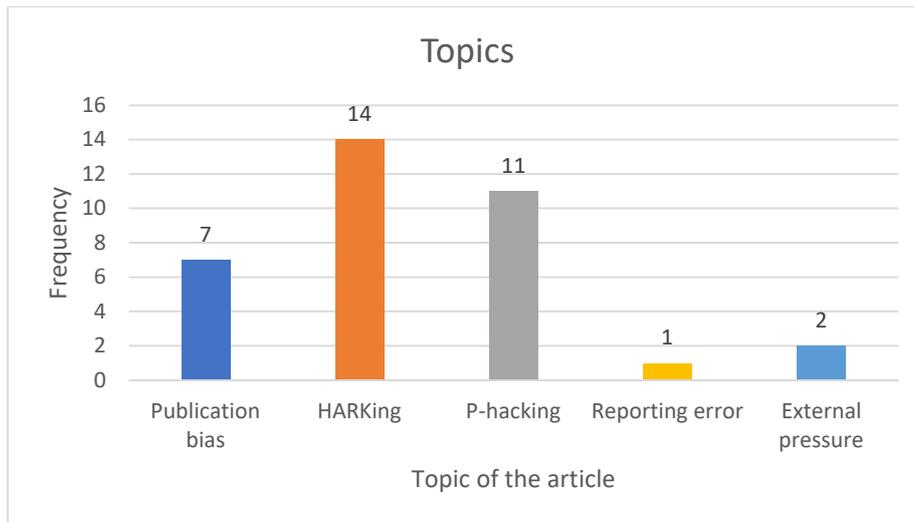
Figure 10 shows how many times each form of reporting bias showed up across all 15 articles. The most researched form is clearly HARKing, only one article did not mention it as a problem or did not research it. The second biggest category was P-hacking with 11 occurrences over 15 articles. These two categories were clearly bigger than the other ones. This means that the awareness of both HARKing and P-hacking is very high in the current literature of reporting bias in management research, it might even be concluded that there is too much focus on these two. Publication bias was mentioned in seven of the 15 articles, so almost in half of them. This is still a lot for one form of bias to appear, seeing as not every article can talk about every form of bias. The two remaining categories received a lot less attention in the examined articles. External pressure only showed up twice, and only one article dealt with reporting errors in management research. This indicates that the awareness of these problems is very low compared to the other forms such as HARKing, P-hacking and publication bias.

4. Discussion

After the data synthesis we can answer the research questions that this thesis asked. The first secondary question is interested in the topics, and the current literature has a nice division of topics. The literature seems to be aware of the role of both journals and authors in the creation of reporting bias and focuses mainly on forms such as HARKing, P-hacking and publication bias. Two other forms are researched very little, these are external pressure and reporting errors. Lastly, it might be better for the field if more focus goes to testing already existing theories and directly researching reporting bias instead of developing new theories.

The other secondary question asks when the literature was published. The results show that the literature is quite new compared to other fields, seeing as the first relevant article in this search is from 2011. This is late compared to fields such as psychology, medicine and economics, where a google search can easily find published articles about publication bias in medicine from 1986, 1990 and still in 2006 (Dickersin, 1990; Ospina, Kelly, Klassen, & Rowe, 2006; Simes, 1986). Psychology already published about the phenomenon of publication bias in 1959 and still in 2001 (Spence & Blanchard, 2001; Sterling, 1959) and you can find an article in economics in 1999 (Ashenfelter et al., 1999). Luckily, the results also show that reporting bias in management research has been gaining more and more attention in recent years, and that it started testing theories in 2017.

Lastly, a more general view that also answers the primary research question. The current literature is mainly interested in HARKing, P-hacking, and publication bias. It is a relatively new topic in the field that is receiving more and more attention, but the number of research is still low. The database search only found 95 articles, while the same search string with a different field tag leads to 348 results in psychology, 419 results in medicine and 215 results in economics. So, the literature about reporting bias in management research is small compared to the other fields, but this also means that theories and solutions might already be present in other fields that might also be applicable in management. Such as the reproducibility project in psychology (Baker, 2015).

5. Conclusion

A systematic map was conducted on the literature regarding reporting bias in management research. One database was used, and the search resulted in 95 potentially relevant articles, these were screened down to 15 obtainable and relevant articles. All of them were used in the data synthesis. The map shows us that the awareness of reporting bias is rising and that the most researched topics are HARKing and P-hacking. While it is good that more and more attention is being paid to the issue it still does not solve the problem right now. This means that currently the quality of management research is not very well known and that a lot of research is necessary to tackle the problem. So, scholars might even be building further on theories that were confirmed by practices like P-hacking or HARKing, which means that they have never properly been tested and confirmed. Scandals such as happened in the field of psychology since 2010 might even be lurking if no actions are taken (Shea, 2012). These findings also lead to questions regarding the validity of evidence-based management and even the education about management. The practices that evidence-based management proposes to use might be based on theories that were confirmed with questionable practices, so they might not reflect reality and by virtue might not be helpful when managing people. But an equally big problem might be that universities are teaching future managers these theories and practices. They might try to apply them during their careers or even teach them to colleagues, which spreads these faulty theories.

There are however several shortcomings present in this thesis. First, the biggest shortcoming is that the systematic map was not completely comprehensive, because time constraints made it impossible to search all the relevant databases or grey literature. Only one source was utilized, this was the Web of Science Core Collection. Further research could search other databases to include all the potential literature on reporting bias out there. The term reporting error could also be included in the search string, but this would likely not change much seeing as Bruns et al. (2019) thinks that it was the first to examine this bias in management research. The normal science bias could also be included in the search string to look more into external pressure. But this is not seen as a form of reporting bias in the management literature, some discussions in the field of management could determine whether it should be. Furthermore, small literature reviews could be conducted to review the literature on HARKing, P-hacking and the role of authors and journals, this could summarize and appraise the relevant articles in management research and show exactly what is already known regarding these topics. A better method might even be a living systematic review, these reviews are updated on a regular basis. Such a review could keep a nice overview of and grow with the knowledge (Elliott et al., 2017). Lastly, the articles were screened and coded by only one person, this is not common in systematic reviews or maps which are often conducted by multiple researchers.

References

Andrade, C. (2021). HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *The Journal of Clinical Psychiatry, 82*(1), 0-0.

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly, 28*(1), 5-21.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour economics, 6*(4), 453-470.

Baker, M. (2015). First results from psychology's largest reproducibility test. *Nature, 30*(10.1038).

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (2017). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization, 15*(3), 423-436.

Bettis, R. A. (2012). The search for asterisks: Compromised statistical tests and flawed theories. *Strategic Management Journal, 33*(1), 108-113.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics. *American Economic Review, 110*(11), 3634-3660. doi:10.1257/aer.20190687

Bruns, S. B., Asanov, I., Bode, R., Dunger, M., Funk, C., Hassan, S. M., . . . König, J. (2019). Reporting errors and biases in published empirical findings: Evidence from innovation research. *Research Policy, 48*(9), 103796.

CEE. (2013). *Guidelines for systematic reviews in environmental management*.

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature, 56*(3), 920-980.

Clarivate. (n.d.). Web of Science Core Collection. Retrieved from https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection/

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi:10.1126/science.aac4716

Corbetta, P. (2003). *Social research: Theory, methods and techniques*: Sage.

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

Dickersin, K. (1990). The Existence of Publication Bias and Risk Factors for Its Occurrence. *JAMA, 263*(10), 1385-1389. doi:10.1001/jama.1990.03440100097014

Dickersin, K. (1997). How important is publication bias? A synthesis of available data. *AIDS education and prevention, 9*, 15-21.

Edwards, J. R., & Berry, J. W. (2010). The Presence of Something or the Absence of Nothing: Increasing Theoretical Precision in Management Research. *Organizational Research Methods, 13*(4), 668-689. doi:10.1177/1094428110380467

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., . . . Pearson, L. (2017). Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of Clinical Epidemiology, 91*, 23-30. doi:https://doi.org/10.1016/j.jclinepi.2017.08.010

Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research, 60*(4), 411-415. doi:https://doi.org/10.1016/j.jbusres.2006.12.003

Gall, T., & Maniadis, Z. (2019). Evaluating solutions to the problem of false positives. *Research Policy, 48*(2), 506-515.

Garud, R. (2015). Eyes wide shut? A commentary on the hypothesis that never was. *Journal of Management Inquiry, 24*(4), 450-454.

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence, 7*(1), 7.

Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H., & Short, J. (2017). Publication Bias in Strategic Management Research. *Journal of Management, 43*(2), 400-425. doi:10.1177/0149206314535438

Hensel, P. G. (2019). Supporting replication research in management journals: Qualitative analysis of editorials published between 1970 and 2015. *European Management Journal, 37*(1), 45-57.

Hollenbeck, J. R., & Wright, P. M. (2017). Harking, Sharking, and Tharking:Making the Case for Post Hoc Analysis of Scientific Data. *Journal of Management, 43*(1), 5-18. doi:10.1177/0149206316679487

Honig, B., Lampel, J., Baum, J. A., Glynn, M. A., Jing, R., Lounsbury, M., . . . Walsh, J. P. (2018). Reflections on scientific misconduct in management: Unfortunate incidents or a normative crisis? *Academy of Management Perspectives, 32*(4), 412-442.

Hubbard, R., & Armstrong, J. S. (1994). Replications and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing, 11*(3), 233-248. doi:https://doi.org/10.1016/0167-8116(94)90003-5

Hubbard, R., & Vetter, D. E. (1996). An empirical comparison of published replication research in accounting, economics, finance, management, and marketing. *Journal of Business Research, 35*(2), 153-164. doi:https://doi.org/10.1016/0148-2963(95)00084-4

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine, 2*(8), e124. doi:10.1371/journal.pmed.0020124

James, K. L., Randall, N. P., & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environmental Evidence, 5*(1), 7. doi:10.1186/s13750-016-0059-6

Joel, A. C. B., & Brian, S. S. (2004). Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. *Journal of Business Venturing, 19*(3), 411-436. doi:https://doi.org/10.1016/S0883-9026(03)00038-7

Kacmar, K. M., & Whitfield, J. M. (2000). An Additional Rating Method for Journal Articles in the Field of Management. *Organizational Research Methods, 3*(4), 392-406. doi:10.1177/109442810034005

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

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review, 2*(3), 196-217. doi:10.1207/s15327957pspr0203_4

Lawrence, P. A. (2003). The politics of publication. *Nature, 422*(6929), 259-261. doi:10.1038/422259a

McGauran, N., Wieseler, B., Kreis, J., Schüler, Y.-B., Kölsch, H., & Kaiser, T. (2010). Reporting bias in medical research - a narrative review. *Trials, 11*(1), 37. doi:10.1186/1745-6215-11-37

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

Ospina, M. B., Kelly, K., Klassen, T. P., & Rowe, B. H. (2006). Publication Bias of Randomized Controlled Trials in Emergency Medicine. *Academic Emergency Medicine, 13*(1), 102-108. doi:https://doi.org/10.1197/j.aem.2005.07.039

Pae, C.-U. (2015). Why Systematic Review rather than Narrative Review? *Psychiatry investigation, 12*(3), 417-419. doi:10.4306/pi.2015.12.3.417

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy, 98*(5, Part 2), S71-S102. doi:10.1086/261725

Rosenthal, R., Kleid, J. J., & Cohen, M. V. (1979). Abnormal mitral valve motion associated with ventricular septal defect following acute myocardial infarction. *American heart journal, 98*(5), 638-641.

ROSES. (2017). ROSES for systematic map protocols. Retrieved from https://www.roses-reporting.com/systematic-map-protocols

Rost, K., & Ehrmann, T. (2015). Reporting Biases in Empirical Management Research: The Example of Win-Win Corporate Social Responsibility. *Business & Society, 56*. doi:10.1177/0007650315572858

Rupp, D. E. (2011). Ethical issues faced by editors and reviewers. *Management and Organization Review, 7*(3), 481-493.

Ryan, P. (2012). *How venture capital works*: The Rosen Publishing Group, Inc.

Shea, C. (2012). Scandals in Social Psychology Spreading. Retrieved from https://www.wsj.com/articles/BL-IMB-3627

Simes, R. J. (1986). Publication bias: the case for an international registry of clinical trials. *Journal of clinical oncology, 4*(10), 1529-1541.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General, 143*(2), 534.

Spence, J. C., & Blanchard, C. (2001). Publication bias in sport and exercise psychology: The games we play. *International Journal of Sport Psychology*.

Squires, G. (2001). Management as a professional discipline*. *Journal of Management Studies, 38*(4), 473-487. doi:https://doi.org/10.1111/1467-6486.00245

Starbuck, W. H. (2016). 60th anniversary essay: How journals could improve research practices in social science. *Administrative Science Quarterly, 61*(2), 165-183.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American statistical association, 54*(285), 30-34.

Stupple, A., Singerman, D., & Celi, L. A. (2019). The reproducibility crisis in the age of digital medicine. *npj Digital Medicine, 2*(1), 2. doi:10.1038/s41746-019-0079-z

Tourish, D., & Craig, R. (2020). Research misconduct in business and management studies: Causes, consequences, and possible remedies. *Journal of Management Inquiry, 29*(2), 174-187.

Van Witteloostuijn, A. (2016). What happened to Popperian falsification? Publishing neutral and negative findings: Moving away from biased publication practices. *Cross Cultural & Strategic Management*.

Annex

Annex A. Search strings tested for number of results

This List shows all the tested search strings and their results when tested on Web of Science

Tested search strings	Results
(TS= ("publication bias*" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	18
(TS= "reporting bias*" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	22
(TS= (data-snooping OR "data snooping") NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	24
(TS= (p-hacking OR "p hacking") NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	8
(TS= "selective reporting*" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	3
(TS= "credibility crisis*" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	5
(TS= harking NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	11
(TS= (non-results OR "non results") NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	1
(TS= ("data fabrication*" OR data-fabrication*) NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	2
(TS= ("statistical methodology" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	57
(TS= ("research malpractice*" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)	0

			23
(TS= ("questionable research practice*" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)			23
(TS= ("theory testing" NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)			128
(TS= (falsification NOT ("meta-analy*" OR "meta analy*" OR "systematic review")) AND WC= management) AND LANGUAGE (English) AND DOCUMENT TYPES (Article)			58

Annex B. All articles from the search and the first screening

An empty cell in the fourth column means that the article was not excluded in this title and abstract screening.

Year	Article title	Authors	Reason for exclusion
	The consequences of self-reporting biases: Evidence from the crash preventability program	Scott, A; Balthrop, AT	reporting by firms, not by researchers
	The contribution of human capital to foreign direct investment inflows in developing countries	Abbas, A; Moosa, I; Ramiah, V	different field than management research
2021	Practice what you preach: Credibility-enhancing displays and the growth of open science	Kraft-Todd, GT; Rand, DG	different topic than reporting bias
2021	A framework for open policy analysis	de la Guardia, FH; Grant, S; Miguel, E	different field than management research
2021	Technical trading and cryptocurrencies	Hudson, R; Urquhart, A	different topic than reporting bias
2021	Data snooping in equity premium prediction	Dichtl, H; Drobetz, W; Neuhierl, A; Wendt, VS	different topic than reporting bias
2020	Data mining fool's gold	Smith, G	different topic than reporting bias
2020	Actionable recommendations for narrowing the science-practice gap in open science	Aguinis, H; Banks, GC; Rogelberg, SG; Cascio, WF	different topic than reporting bias
2020	Using retracted journal articles in psychology to understand research misconduct in the social sciences: What is to be done?	Craig, R; Cox, A; Tourish, D; Thorpe, A	different field than management research
2020	Information Aggregation and P-Hacking	Rytchkov, O; Zhong, X	different field than management research
2020	Research Misconduct in Business and Management Studies: Causes,	Tourish, D; Craig, R	

	Consequences, and Possible Remedies		
2021	Factors affecting orderly parking of dockless shared bicycles: an exploratory study	Wang, YC; Jia, SQ; Zhou, HY; Charlton, S; Hazen, B	different topic than reporting bias
2020	AN A IS AN A: THE NEW BOTTOM LINE FOR VALUING ACADEMIC RESEARCH	Aguinis, H; Cummings, C; Ramani, RS; Cummings, TG	different topic than reporting bias
2020	Flexibility in the selection of patent counts: Implications for p-hacking and evidence-based policymaking	Bruns, SB; Kalthaus, M	different field than management research
2021	Using Outliers for Theory Building	Gibbert, M; Nair, LB; Weiss, M; Hoegl, M	different topic than reporting bias
2021	Why the Increasing Use of Complex Causal Models Is a Problem: On the Danger Sophisticated Theoretical Narratives Pose to Truth	Saylors, R; Trafimow, D	different topic than reporting bias
2019	Factors affecting the academic performance of real estate students in a specialized Federal University of Technology in Nigeria	Ojetunde, I; Sule, AI; Kemiki, OA; Olatunji, IA	different field than management research
2020	Evaluation of current research on stock return predictability	Reschenhofer, E; Mangat, MK; Zwatz, C; Guzmics, S	different field than management research
2019	Reporting errors and biases in published empirical findings: Evidence from innovation research	Bruns, SB; Asanov, I; Bode, R; Dunger, M; Funk, C; Hassan, SM; Hauschildt, J; Heinisch, D; Kempa, K; Konig, J; Lips, J; Verbeck, M; Wolfschutz, E; Buenstorf, G	
2019	Adoption and use of technology with low litigation risk - the case of financial reporting on Twitter by ASX companies	Xiong, F; Chapple, L; Xu, S; Lin, WW	reporting by firms, not by researchers
2019	Race, threat and workplace sexual harassment: The dynamics of harassment in the United States, 1997-2016	Cassino, D; Besen-Cassino, Y	different topic than reporting bias
2019	Calendar Anomalies in The Indian Stock Market - An Emperical Study	Wats, S	different field than management research
2019	The First 20 Years of Organizational Research Methods: Trajectory,	Aguinis, H; Ramani, RS; Villamor, I	different topic than reporting bias

	Impact, and Predictions for the Future		
2019	How does environmental performance map into environmental disclosure? A look at underlying economic incentives and legitimacy aims	Tadros, H; Magnan, M	reporting by firms, not by researchers
2019	Towards a taxonomy of research misconduct: The case of business school research	Hall, J; Martin, BR	different topic than reporting bias
2019	Evaluating solutions to the problem of false positives	Gall, T; Maniadis, Z	
2019	Questionable research practices when using confirmatory factor analysis	Crede, M; Harms, P	
2019	Supporting replication research in management journals: Qualitative analysis of editorials published between 1970 and 2015	Hensel, PG	
2019	Research on Work as a Calling ... and How to Make It Matter	Thompson, JA; Bunderson, JS	different topic than reporting bias
2018	REFLECTIONS ON SCIENTIFIC MISCONDUCT IN MANAGEMENT: UNFORTUNATE INCIDENTS OR A NORMATIVE CRISIS?	Honig, B; Lampel, J	
2018	Knowledge dissemination in clinical trials: Exploring influences of institutional support and type of innovation on selective reporting	Salandra, R	different field than management research
2018	Above a swamp: A theory of high-quality scientific production	Kiri, B; Lacetera, N; Zirulia, L	
2018	Retraction statements and research malpractice in economics	Cox, A; Craig, R; Tourish, D	different field than management research
2018	Time series forecasting using functional partial least square regression with stochastic volatility, GARCH, and exponential smoothing	Kim, JM; Jung, HJ	different field than management research
2018	Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations	Srivastava, SB; Goldberg, A; Manian, VG; Potts, C	reporting by people, not by researchers
2018	WHAT YOU SEE IS WHAT YOU GET? ENHANCING	Aguinis, H; Ramani, RS; Alabduljader, N	

	METHODOLOGICAL TRANSPARENCY IN MANAGEMENT RESEARCH		
2018	Interplay between P-O fit, transformational leadership and organizational social capital	Raja, U; Bouckenooghe, D; Syed, F; Naseer, S	different topic than reporting bias
2018	To Grasp Cognition in Action, Combine Behavioral Experiments with Protocol Analysis	Reypens, C; Levine, SS	different topic than reporting bias
2017	Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings	Bergh, DD; Sharp, BM; Aguinis, H; Li, M	
2017	Degrees of Freedom in SEM: Are We Testing the Models That We Claim to Test?	Cortina, JM; Green, JP; Keeler, KR; Vandenberg, RJ	different field than management research
2017	Exploratory data analysis as a foundation of inductive research	Jebb, AT; Parrigon, S; Woo, SE	different topic than reporting bias
2017	Multiperiod portfolio investment using stochastic programming with conditional value at risk	Chen, HH; Yang, CB	different field than management research
2017	Two Tales of Return Predictability: The Case of Asia-Pacific Equity Markets	Shynkevich, A	different field than management research
2017	The Gray Zone: Questionable Research Practices in the Business School	Butler, N; Delaney, H; Spoelstra, S	different field than management research
2017	Solutions to the Credibility Crisis in Management Science	Byington, EK; Felps, W	
2017	A Call for Openness in Research Reporting: How to Turn Covert Practices Into Helpful Tools	Schwab, A; Starbuck, WH	
2017	The Chrysalis Effect: How Ugly Initial Results Metamorphosize Into Beautiful Articles	O'Boyle, EH; Banks, GC; Gonzalez-Mule, E	
2017	On doing better science: From thrill of discovery to policy implications	Antonakis, J	
2017	BOOTSTRAP TESTING OF TRADING STRATEGIES IN EMERGING BALKAN STOCK MARKETS	Radovanov, B; Marcikic, A	different field than management research
2016	Forecasting Errors, Directional Accuracy and Profitability of	Costantini, M; Cuaresma, JC; Hlouskova, J	different field than management research

	Currency Trading: The Case of EUR/USD Exchange Rate		
2016	Defining and operationalizing theory	Cortina, JM	different topic than reporting bias
2016	Bias and the Commitment to Disclosure	Heinle, MS; Verrecchia, RE	reporting by firms, not by researchers
2016	60th Anniversary Essay: How Journals Could Improve Research Practices in Social Science	Starbuck, WH	
2016	De-Biasing the Reporting Bias in Social Media Analytics	Chen, HY; Zheng, Z; Ceran, Y	reporting by people, not by researchers
2016	The effectiveness of R&D subsidies: A meta-regression analysis of the evaluation literature	Dimos, C; Pugh, G	different topic than reporting bias
2016	Critical Essay: Building new management theories on sound data? The case of neuroscience	Lindebaum, D	different field than management research
2016	Mediating role of organizational learning on the relationship between market orientation and innovativeness	Raj, R; Srivastava, KBL	different topic than reporting bias
2016	Emergent risks in business process change projects	Jurisch, MC; Rosenberg, Z; Krcmar, H	different topic than reporting bias
2016	REGULAR ISSUE PAPER What happened to Popperian falsification? Publishing neutral and negative findings Moving away from biased publication practices	van Witteloostuijn, A	
2016	Adaptive Evolutionary Neural Networks for Forecasting and Trading without a Data-Snooping Bias	Sermpinis, G; Verousis, T; Theofilatos, K	different field than management research
2015	Eyes Wide Shut? A Commentary on the Hypothesis That Never Was	Garud, R	
2015	How reliable are self-report measures of mileage, violations and crashes?	af Wahlberg, AE; Dorn, L	reporting by people, not by researchers
2015	Training comprehensiveness: construct development and relation with role behaviour	Srivastava, AP; Dhar, RL	different field than management research
2015	Time-varying industry beta in Indian stock market and forecasting errors	Das, S; Barai, P	different field than management research
2014	Conceptual and empirical confounds in the organizational	Martinko, MJ; Harvey, P; Mackey, JD	different topic than reporting bias

Year	Title	Authors	Reason
	sciences: An explication and discussion		
2014	Evidence-Based Management and the Trustworthiness of Our Cumulative Scientific Knowledge: Implications for Teaching, Research, and Practice	Kepes, S; Bennett, AA; McDaniel, MA	
2014	Illusory profitability of technical analysis in emerging foreign exchange markets	Kuang, P; Schroder, M; Wang, Q	different field than management research
2014	Building Technical Trading System with Genetic Programming: A New Method to Test the Efficiency of Chinese Stock Markets	Qu, H; Li, XD	different field than management research
2014	Guidelines for Neuroscience Studies in Information Systems Research	vom Brocke, J; Liang, TP	different field than management research
2013	Meta-Synthesis of Qualitative Case Studies: An Approach to Theory Building	Hoon, C	meta-synthesis
2013	Economic significance of market timing rules in the Forward Freight Agreement markets	Nomikos, NK; Doctor, K	different field than management research
2013	Funds of Hedge Funds versus Do-It-Yourself Funds of UCITS	Sender, S	different field than management research
2012	The doping self-reporting game: The paradox of a 'false-telling' mechanism and its potential research and policy implications	Petroczi, A; Haugen, KK	reporting by people, not by researchers
2012	THE SEARCH FOR ASTERISKS: COMPROMISED STATISTICAL TESTS AND FLAWED THEORIES	Bettis, RA	
2011	Ethical Issues Faced by Editors and Reviewers	Rupp, DE	
2010	Profitability of technical analysis in financial and commodity futures markets - A reality check	Yen, SMF; Hsu, YL	different topic than reporting bias
2010	Port safety and the container revolution: A statistical study on human factor and occupational accidents over the long period	Fabiano, B; Curro, F; Reverberi, AP; Pastorino, R	different topic than reporting bias
2010	Nonlinearity, data-snooping, and stock index ETF return predictability	Yang, J; Cabrera, J; Wang, T	different field than management research

2009	SELECTIVE STATUS REPORTING IN INFORMATION SYSTEMS PROJECTS: A DYADIC-LEVEL INVESTIGATION	Iacovou, CL; Thompson, RL; Smith, HJ	reporting by people, not by researchers
2009	On the existence of sports sentiment: the relation between football match results and stock index returns in Europe	Klein, C; Zwergel, B; Heiden, S	different field than management research
2009	The Value of Alpha Forecasts in Portfolio Construction	Fong, K; Gallagher, DR; Lee, AD	different topic than reporting bias
2009	Job satisfaction and subjective health among sales managers	Gustainien, L; Endriulaitiene, A	different topic than reporting bias
2008	The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias	Dellarocas, C; Wood, CA	reporting by people, not by researchers
2008	Out-of-sample forecasting of unemployment rates with pooled STVECM forecasts	Milas, C; Rothman, P	different topic than reporting bias
2007	Stock market trading rule discovery using pattern recognition and technical analysis	Wang, JL; Chan, SH	different topic than reporting bias
2007	The effects of optimistic and pessimistic biasing on software project status reporting	Snow, AP; Keil, M; Wallace, L	reporting by people, not by researchers
2007	Forecasting inflation using economic indicators: The case of France	Bruneau, C; de Bandt, O; Flageollet, A; Michaux, E	different field than management research
2007	Investment timing and trading strategies in the sale and purchase market for ships	Alizadeh, AH; Nomikos, NK	different topic than reporting bias
2006	Publication bias: A case study of four test vendors	McDaniel, MA; Rothstein, HR; Whetzel, DL	different field than management research
2005	The continuity-change duality in narrative texts of organizational identity	Chreim, S	reporting by people, not by researchers
2005	Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries	Awartani, BMA; Corradi, V	different field than management research
2004	Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood	Gonzalez-Rivera, G; Lee, TH; Mishra, S	different field than management research

2004	Biases in incident reporting databases: an empirical study in the chemical process industry	van der Schaaf, T; Kanse, L	different field than management research
2002	The challenge of accurate software project status reporting: A two-stage model incorporating status errors and reporting bias	Snow, AP; Keil, M	reporting by people, not by researchers
2001	A cascaded inference model for evaluation of the internal audit report	Krishnamoorthy, G	different field than management research

Annex C. Unobtainable articles

Year	Article title	Authors
2017	Solutions to the Credibility Crisis in Management Science	Byington, EK; Felps, W
2017	A Call for Openness in Research Reporting: How to Turn Covert Practices Into Helpful Tools	Schwab, A; Starbuck, WH

Annex D. Articles excluded during the full text screening

Year	Article title	Authors	Reason for exclusion
2018	Above a swamp: A theory of high-quality scientific production	Kiri, B; Lacetera, N; Zirulia, L	different topic than reporting bias
2018	WHAT YOU SEE IS WHAT YOU GET? ENHANCING METHODOLOGICAL TRANSPARENCY IN MANAGEMENT RESEARCH	Aguinis, H; Ramani, RS; Alabduljader, N	systematic literature review

Annex E. Quotes for the coding of the topic

This annex shows a table that gives quotes that were used to code the topics of each article.

Article number	Quote	topic
1	"P-hacking involves reporting only results that deliver a desired p value" (Tourish & Craig, 2020, p 176)	P-hacking

1	"HARKing presents hypotheses as if they were developed a priori when in fact they have been developed after the results are known." (Tourish & Craig, 2020, p 176)	HARKing
2	"This "p-hacking" (Simonsohn et al., 2014) need not be based on a deliberate intention to mislead the audience" (Bruns et al., 2019, p 2)	P-hacking
2	"… differs from the significance level denoted by asterisks, then a reporting error is present" (Bruns et al., 2019, p 2)	Reporting errors
2	"This procedure is commonly known as HARKing (Hypothesizing After the Results are Known)" (Bruns et al., 2019, p 2)	HARKing
2	"Publication bias occurs at the study level when authors do not submit studies with statistically non-significant results for publication, or when reviewers and editors reject studies with non-significant results." (Bruns et al., 2019, p 2)	Publication bias
3	"In essence, these requirements aim at reducing or eliminating behavior that seeks to "improve" research results by adopting favorable methods, for instance "p-hacking" by unreported multiple testing, censoring of data or adjustment in the number of trials" (Gall & Maniadis, 2019, p 506)	P-hacking

4	"Unfortunately, perceived or actual journal norms haveresulted in a tendency for researchers in the organizational science to hide potentiallydisconfirming evidence and primarily present evidence that is supportive of a theory," (Crede & Harms, 2019, p 18)	Publication bias
4	"In this paper, we describe seven analytic practices and reporting practices relating to thetesting of measurement models via confirmatory factor analysis (CFA) that reduce thedegree to which readers are exposed to disconfirming evidence" (Crede & Harms, 2019, p 18)	P-hacking
4	"encouraged by reviewers and editors to engage in QRPs such as HARKing" (Crede & Harms, 2019, p 25)	HARKing
5	"Multiple tests conducted in search for asterisks (Bettis, 2012) allowing for hypothesizing after results are known" (Hensel, 2019, p 46)	HARKing
5	"p-hacking, that is, adjusting the statistical analysis and reporting to show "statistically significant" results" (Hensel, 2019, p 46)	P-hacking
6	"Selectively omitting (or including) variables, observations, and/or statistical analyses until non-significant results become significant at standard levels, or "p-	P-hacking

	hacking" (Simmons, Nelson, & Simonsohn,2011)" (Honig et al., 2018, p 423)	
6	"Presenting post hoc hypotheses as if they were a priori hypotheses, or "HARKing"" (Honig et al., 2018, p 423)	HARKing
6	"While Charles Darwin withheld his initial findings for20 years out of fear of offending contemporary religious views." (Honig et al., 2018, p 414)	External pressure
7	"cherry-picked their findings from a larger set of models" (Bergh et al., 2017, p 431)	P-hacking
7	"In addition, articles may experience a metamorphosis during the review process whereby authors may engage in post hoc alterations of hypotheses and data as well as engage in questionable research practices" (Bergh et al., 2017, p 431)	HARKing
8	"The problem of publication bias is ubiquitous and evident across many fields (Pfeiffer, Bertram, & Ioannidis, 2011) making it difficult to reconstruct the distribution of effect sizes." (Antonakis, 2017, p 7)	Publication bias
8	"When only statistically significant and novel results are published, individuals may game the system expressly; they may p-hack or engage in other questionable research practices." (Antonakis, 2017, p 9)	P-hacking

8	"HARKing— "hypothesizing after the results are known."" (Antonakis, 2017, p 9)	HARKing
9	"It is the lack of information and in some cases, such as presenting post hoc findings as a priori hypotheses, misinformation that data mining or "peeking" becomes a QRP." (O'Boyle Jr et al., 2017, p 392)	HARKing
9	"The result of these QRPs is an increase in Type I errors and a suppression of null effects, which biases the literature" (O'Boyle Jr et al., 2017, p 377)	P-hacking
10	"The essay also identifies some troublesome properties of prevalent methodology, such as statistical significance tests, HARKing, and p-Hacking, and proposes editorial policies to mitigate these detrimental behaviors." (Starbuck, 2016, p 165)	HARKing
10	"Data mining, p-Hacking, or data dredging involves subjecting data to many calculations or manipulations in search of an equation or classification system that captures strong patterns" (Starbuck, 2016, p 171)	P-hacking
11	"This malpractice of so-called HARKing (=hypothesizing after the results are known; Kerr, 1998) may be inadvertently stimulated by the behavior of many editors and reviewers." (Van Witteloostuijn, 2016, p 484)	HARKing

11	"If only positives are published without any serious attempt at replication, the only thing we can be sure of is that false positives will abound." (Van Witteloostuijn,2016, p 489)	Publication bias
11	"In normal science, findings that seriously go against the prevailing paradigm are not welcomed as a step toward further progress, but rather are put aside as mistakes of the researcher" (Van Witteloostuijn,2016, p 489)	External pressure
12	"Several scholars have commented on the practice of presenting post hoc hypotheses as if they were known a priori." (Garud, 2015, p 451)	HARKing
13	"One plausible explanation is that the reward structure inhibits scientific progress and publication bias prohibits compiling all data that exists (Kepes & Mc- Daniel, 2013)" (Kepes et al., 2014, p 448)	Publication bias
13	"As an example, O'Boyle, Banks, and Gonzalez-Mule (in press) have documented the extensive use of questionable research practices (e.g., HARKing, hypothesizing after the results are known, Kerr, 1998) in transforming dissertation research into journal articles (see also Pigott, Valentine, Polanin, Williams, & Canada, 2013)." (Kepes et al., 2014, p 458)	HARKing

14	"Replication of this hypothesis test on other samples would very likely provide counterevidence, but would not be publishable because professional norms generally preclude publication of replication studies and what are usually called 'non-results.'" (Bettis, 2012, p 109)	Publication bias
14	"Data snooping or searching for asterisks is the most damaging form of repeated testing, since the aim is to reject null hypotheses while consciously ignoring the many models and tests that have been conducted and, thus, reporting greatly exaggerated levels of significance." (Bettis, 2012, p 110)	P-hacking
14	"When such models were found, he helped his mentors propose theories and hypotheses on the basis of which the 'asterisks' could be explained." (Bettis, 2012, p 109)	HARKing
15	"Other papers in this special issue have discussed HARKing: hypothesizing after results are known (see Leung, 2011)" (Rupp, 2011, p 486)	HARKing