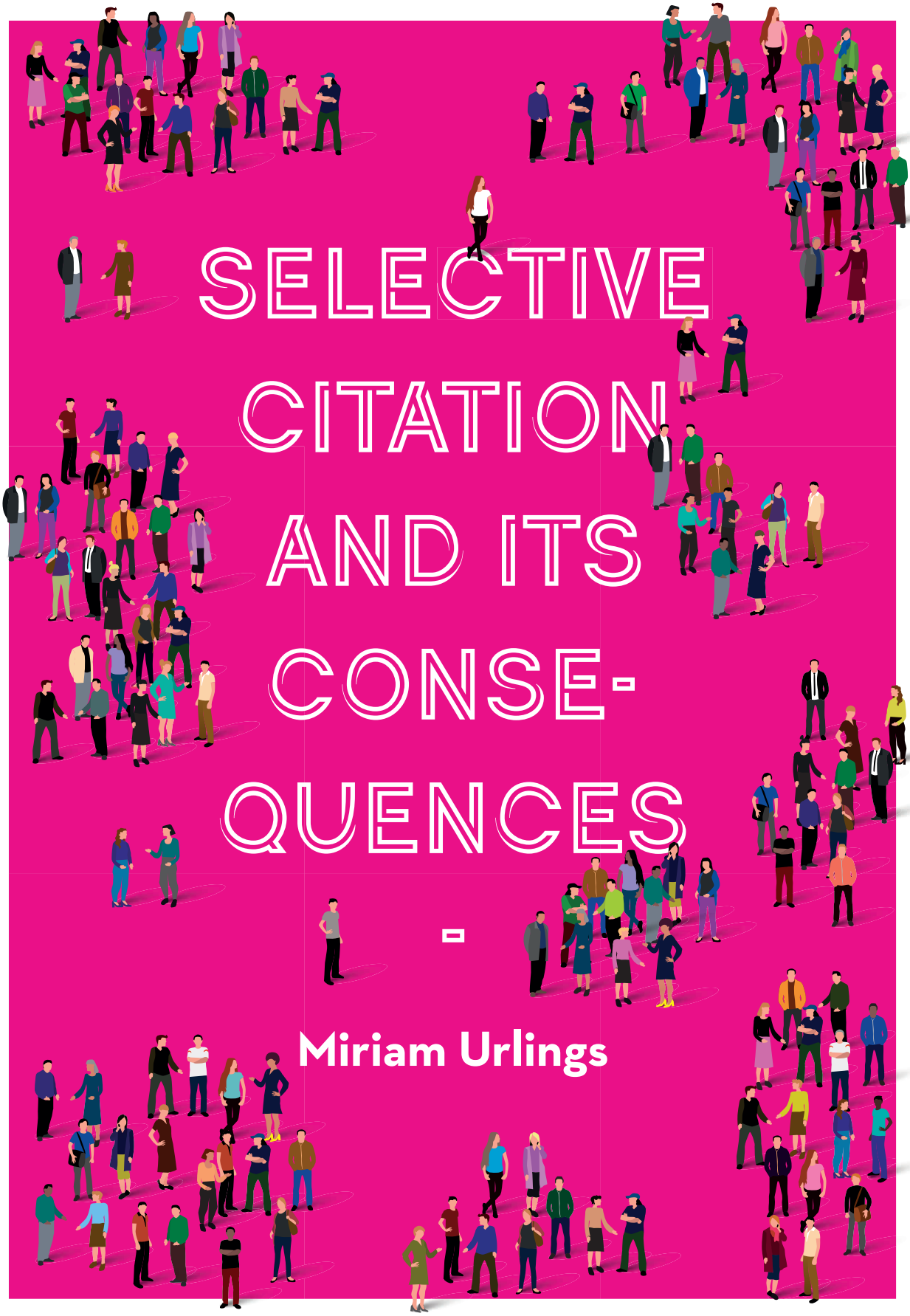

SELECTIVE CITATION AND ITS CONSE-QUENCES
Miriam Urlings

Selective citation and its consequences

Miriam Urlings

Printing: ProefschriftMaken || www.proefschriftmaken.nl

ISBN 978 94 6380 401 1
Cover design: Nathalie Gerrekens

Selective citation and its consequences

DISSERTATION

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof.dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public

on. Wednesday 3 July 2019 , at 14:00 hours

by

Miriam Josephina Elisabeth Urlings

Supervisor(s):
Prof. Dr. M.P.A. Zeegers
Prof. Dr. L.M. Bouter

Co-supervisor:
Dr. G.M.H. Swaen

Assessment Committee:
Prof. Dr. D. Townend (chair)
Prof. Dr. A. Bast
Prof. Dr. A. Knottnerus
Dr. G. ter Riet – Universiteit van Amsterdam
Dr. J. Tijdink – Vrije Universiteit Amsterdam

Financial statement:
Chapters 3, 5 and 6 of this thesis were part of the Sound Science project, a project funded by the Long-range Research Initiative (LRI) of the European Chemical Industry Council (CEFIC).
Chapters 4 and 7 of this thesis were part of the Eatwel project, funded by Maastricht University

Contents

1

Introduction

Meta-research: Research on research

Science is a key driver of progress (1). Scientific knowledge is the basis for a wide range of developments, such as in medical therapies and emerging technologies. Evidence-based medicine, which is a widely accepted approach in the medical discipline nowadays, depends on high quality scientific research (2). With the growing amount of scientific output, we need to ensure that the scientific enterprise remains efficient and reliable (3). The scientific discipline that empirically studies the functioning of science is called meta-research (3). In the Netherlands, meta-research has gained great interest in the last years, following the scientific scandal of professor Diederik Stapel in 2012 (4). For years, he was able to fabricate data and publish the results in high impact factor journals. Fabrication is, together with falsification and plagiarism, considered serious Research Misconduct (RM), which luckily is not very frequent. A meta-analysis of surveys by Fanelli showed that 2% of researchers have engaged in research misconduct (5). Although not very prevalent, these are serious misbehaviors, that lead to scientific output without any value. Furthermore, this is misleading to the scientific community and to other stakeholders and potentially leading to harmful policy or clinical decisions. On a positive note, research misconduct can be resolve, once it is detected. The researcher can be punished and the related publications can be retracted. Not as serious, but still very much damaging the scientific enterprise, are Questionable Research Practices (QRPs) (6). These are more subtle misbehaviors which are not easily recognised and thereby much more threatening to the validity and credibility of science. QRPs can occur in all aspects of performing research; from designing a study to collecting and analysing data and reporting of the research. A number of decisions need to be made during this research process, which might skew the outcomes and interpretation of the findings into a specific direction. These, often arbitrary, decisions in each stage of research, have been referred to as researchers degrees of freedom (7). Meta-research studies the occurrence and determinants of research misconduct and QRPs, with a focus on detection and prevention (1).

Initiatives to map and promote research integrity

As a response to cases of scientific misconduct and with the introduction of the concept of QRPs, several initiatives have been started to improve the credibility and reliability of science. These initiatives approach the issue of misconduct and QRPs from different angles. One approach includes empirical research to map the frequency of QRPs and research misconduct. Other initiatives include the development of skill trainings to teach the new generation of academics about the basic principles of research integrity, as well as setting up codes of conduct to promote responsible conduct of research. A selection of these initiatives will be presented in the following paragraph.

Within the research community, awareness of the importance and impact of QRP is growing. Several collaborations have been started over the last years to promote research integrity research and activities. For example, the organisation of the World Conference on Research Integrity (WCRI), which started in 2007 in Lisbon. Initiated by the European Science Foundation (ESF) and the American Office for Research Integrity (ORI), these conferences function as a forum to study and discuss responsible conduct of research, but also to bring together stakeholders such as policy makers, university boards, publishers and even industry (8). The development and coming of age of the field of meta-research is clearly seen in the conference themes. During the first WCRI in 2007, the focus was on research misconduct (9). The 5th WCRI in 2017 focused on transparency and accountability in research. In the mean time, the WCRI brought forward three important policy documents: the Singapore statement, the Montreal statement and the Amsterdam agenda. The Singapore statement was developed during the second WCRI, with the goal to make a concerted effort to promote global research integrity. It provides guidance for research institutes, governments and academics to develop policies and regulations to promote responsible conduct of research (10, 11). In the Montreal statement, the focus was shifted towards research integrity in cross-boundary research collaborations. In collaborations between different countries, different scientific disciplines or different organisations, there might be differences in research culture and views on responsible conduct of research. The Montreal statement is of help in these situations, by clearly stating collaborative responsibilities at all stages of research (12). The Amsterdam Agenda was the result of the 5th WCRI in 2017. This is focussed on the assessment of efforts to improve scientific integrity and the use of empirical information in developing research integrity policies. The Amsterdam Agenda includes the establishment of a registry for research on responsible conduct of research, that will encourage researchers to plan, conduct, report and share their research.

In an attempt to promote responsible conduct of research among scientists, several codes of conduct have been developed over the past years. The European Code of Conduct on Research Integrity has been initiated by the European Federation of Academies of Sciences and Humanities (ALLEA). They provide guidelines of good research practice in various contexts, such as training and supervision, research procedures, data management and dissemination of results (13). Also on national level, various countries have enforced guidance documents to promote and regulate research integrity, however substantial heterogeneity exists between the different countries (14). Although no legal rights can be derived, the codes of conduct function as a framework for good research practice (15). Since the authors of the various codes of conduct are high level scientific institutions, they also function as a role model to promote adherence to the presented frameworks.

A concrete example of actions that have been undertaken to improve the reporting of research, by researchers and publishers, is the introduction of reporting guidelines.

Guidelines have been developed by the scientific community and have been made specific for different study designs, such as the STROBE guidelines for observational studies (16), the CONSORT statement for clinical trials (17) and the PRISMA statement for systematic reviews (18). Guidelines aim to increase the clarity, transparency and completeness of how a study was conducted (19). This is especially important in being able to judge the credibility of the results and in reproducing the research. Additionally, the use of guidelines leads to a more coherent format of reporting research, which makes it easier to compare separate studies. To promote the use of reporting guidelines, the EQUATOR network was developed in 2009. This international collaboration of researchers aims to promote transparent and accurate reporting of research and provide training and resources to facilitate good research reporting (20). Many journals support the use of reporting guidelines and have integrated it in the submission and peer review process, often as obligatory parts of the publication process.

In 2009, a collaboration of researchers from different countries and academic fields, gathered in a campaign called 'Reduce research Waste and Reward Diligence' (REWARD). They published a series of articles in the Lancet, in which the problem of research waste is described. It was concluded that approximately 85% of the current clinical research can be classified as waste. This waste has different reasons, such as irrelevant research questions (21), inadequate use of research design and analyses (22), inefficient research regulation and management (23), research information that is not fully accessible (24) and biased or unusable research reports (25). The current thesis will focus on this last category: the reporting of research.

Use of metrics in science

Scientific output is often measured by means of several metrics, which can be applied to rank scientists, institutions or journals (26). Metrics are quantitative measures, such as publications in high impact factor journals, number of citations and h-index. As originally intended, these metrics give an indication of the added value that a scientist or journal has on the scientific knowledge development. A highly cited paper has apparently been read by many colleagues in the field and has inspired follow-up research. However, by focussing too much on these metrics, they have become a goal in itself. Metrics are used in the evaluation of researchers' performance, assigning tenure tracks or awarding research grants. Because of these high stakes involved, they have come to function as perverse incentives, which lead to undesirable behavior and can be manipulated relatively easy. Much criticism is heard about the traditional metrics (27, 28). As an alternative, scientific initiatives have started to put less value on the use of traditional metrics, such as the movement Science in Transition (29) and the Leiden Manifesto (30). One of the limitations of traditional metrics is that it only measures impact of scientific work within the scientific community (31). Recently we see a shift in the use of

metrics, towards the use of altmetrics, which also takes into account societal impact (31). Not only scientists comment of the use of metrics, also among publishers it has been recognized that metrics such as journal impact factors give no adequate reflection of the quality of the published research. In 2013, the San Francisco Declaration on Research Assessment (DORA) was signed by editors and publishers in biomedical sciences, as well as individual scientists, as a guidance document to place less focus on journal impact factors in evaluating the work of individual researchers and granting promotions and allocating research funds (32, 33).

Reporting biases

Publishing in peer-reviewed journals remains the main form of communication among scientists (26). Scientific publications make it possible for future researchers to build onto these findings and to create a stronger evidence base. However, scientific writing is hardly ever completely neutral and objective, but rather persuasive and laden with theory and value (34). Different forms of reporting biases can be distinguished, which can all be classified as questionable research practices.

Transparent and complete reporting of research is especially important in enabling the replication of research and thus increasing its credibility. Additionally, publications have a function in making scientific information available for non-scientists, in order to translate scientific findings into practice and policy. In biomedical research, this includes communication towards patients, clinicians, insurance companies and the government, in order to reach the best possible evidence-based care. This paragraph will describe some of the biases that can take place in reporting of scientific research, which negatively impact knowledge development and reduce the quality and credibility of evidence-based medicine.

Publication bias

It is known that many studies do not get published. Already in 1979 this was described by Rosenthal as the 'file-drawer problem' (35). This is mostly a problem, since the studies with positive results are more likely to get published compared to studies with null findings. The published studies will included type-I errors, whereas the type-II errors, which are studies with false negative results, remain unpublished (35). When non-significant findings do get published, it often takes a longer time compared to its significant counterparts (36). Research has shown that the number of non-significant studies has strongly decreased over the last decades in numerous scientific disciplines, showing a clear trend toward publishing positive studies (37). This phenomenon, that the likelihood of publishing research is associated with its study outcome, is known as publication bias (38). Publication bias is well studied in many scientific disciplines and is widely

recognized as a core problem of the scientific enterprise (39). It is a wide known problem that also takes place in non-medical research fields, such as political science (40) and economics (41).

To identify the stakeholders that are mainly responsible for the occurrence of publication bias, it seems obvious to look at journal editors and publishers, since they have the power to accept or reject submitted manuscripts. However, research has shown that publication bias is mostly the result of behavior of the researchers and research funders, since negative results often are not submitted to journals if funders don't demand this (42-44).

Publication bias can have several consequences. First, it will lead to overrepresentation of positive effects and threaten the validity of meta-analyses (39). Especially for meta-analyses this is a problem, since they are expected to give a more accurate evaluation of the available evidence, in comparison to reviews that are more narrative (45). Second, it makes the reported scientific evidence unreliable for decision making, which can be in clinical practices, policy making and development of guidelines (39, 46). Finally, not publishing research is a great waste of resources, which is both financial and in terms of time and trust invested by participants and researchers (24, 47).

With the growing evidence on the existence of publication bias, also the perceived urgency to solve this problem has grown. Some journals have taken the step to include a negative results section, to provide a platform for non-significant results (48).

Outcome reporting bias

Additional to the problem of non-publication of negative studies, there is selective reporting within publications. This selective reporting refers to the publication of a study, with omission of parts of the research results (49). This selection might relate to the studied health outcomes, to the reporting of sensitivity analyses or subgroup analyses (50). Several reasons for the selective reporting of results have been mentioned by scientists, such as limited space in the journal, lack of clinical importance and lack of statistical significance (51). The term outcome reporting bias is used, when the selection of results to be reported is associated with the study outcome (52). In order to assess the occurrence of outcome reporting bias, it is required to know the original research plan as described in the study protocol and the totality of the statistical analyses conducted. By comparing the analysis plan in the protocol to the presented results in the publication, selective reporting can be identified. To assess whether the published selection of results is biased, also the unpublished results should be available. The literature on outcome reporting bias is coming from a broad range of research fields and therefore displays great heterogeneity. However, all available evidence is quite consistent in the sense that selective reporting of outcomes is mostly in favour of positive, significant results (39).

The solution for both publication bias and outcome reporting bias is the publication of the study protocol prior to performing the study (24). The protocol lays down the primary research questions and the methods that will be used to answer this question. By publishing this study protocol in a journal or in an online repository, it is possible to compare the initial research plan with the findings reported in the publication of the study. This can be especially useful for journal editor, peer-reviewers and the funders of the research to check for questionable research practices.

In clinical drug trials the reporting of study protocols has become mandatory. In the United States all researchers who have received funding from NIH are even obligated to publish their research protocol on the repository clinicaltrails.gov. In observational studies on the other hand, publication of study protocols is still quite unusual.

Citation bias

Another important aspect in knowledge development is citing previous publications. Previous research described different roles of citations. Fundamentally, citations have the function to assigning priority or ownership of a claim or finding to the authors of the cited publication (53). This would plead for an objective way of selecting citations. Additionally, citations can be used as a tool of persuasion, to underpin the importance of specific findings and motivate the scientific community to integrate this view in the existing knowledge base (53). This might tempt a researcher to cite mostly sources that are in accordance with the reported findings. On the other hand, authors might choose to cite publications with contradicting findings, to explain why the research at issue is different. Already since long time, citation analyses have been used to assess the development of disciplines over time and to score the performance of institutes, individual scientists and journals (54). Recently, the idea that citations are an indicator of study quality has been challenged (55). In a survey among 1300 scientists, it was found that selectively citing publications to enhance ones own findings and selectively citing to please editors and reviewers were one of the most frequent misbehaviors (56). Citation bias refers to the situation where positive, significant studies are more likely to be cited compared to negative, non-significant studies. For correct knowledge development, it is important that citations give a representative overview of the available positive and negative evidence.

Textbox 1. Overview of terminology with respect to reporting biases and incomplete reporting in biomedical research

Publication bias: Non-publication of research, driven by the study outcome. Usually this means positive studies are published more often than negative studies

Outcome reporting bias: Selective reporting of the study findings, driven by study outcome. Usually leading to more positive results to be reported in a publication, while leaving out negative results

Selective outcome reporting: Selective reporting of a part of the study findings, associated with other determinants than study outcome. For example by reporting only a selection of measured health parameters.

Citation bias: Selective citation driven by the study outcome. Usually this means positive studies are cited more than negative studies.

Selective citation: Selective citation driven by other determinants than study outcome. For example by citing more to publically funded studies, compared to industry funded studies.

Table 1. Schematic overview of incomplete reporting in various sections of a publication

		Article paragraph	
Determinants of selection		**Results**	**References**
	Study outcome	Outcome reporting bias	Citation bias
	Other determinants	Selective outcome reporting	Selective citation

Methods for studying citation bias

In order to study whether the likelihood of citation is related to the study outcome, a citation analysis should be performed. Citation analyses can be performed in several ways, for example by studying all publications in a specific field or within a given publication period. Greenberg (2009) performed a network analysis based on a specific scientific claim. After identifying all available literature (primary data articles, model studies and reviews) on a specific scientific claim, he compared the number of citations to critical primary data publications with the number of citations to supportive primary data publications. From this network analysis it was concluded that 94% of the citations was assigned to the supportive publications (57). A risk of such a claim-specific network might be the existence of confirmation bias. The claim under study is often chosen for the particular reason that the researcher expects there to be citation bias, resulting in positive findings. Apart from a claim-specific network, a network analysis can be based

on all publications in a specific journal or time period. Fanelli (2013) performed a citation analysis on hypothesis-testing studies from several research domains, published between 2000 and 2007. It was concluded that positive studies received on average 32% more citations than negative studies (58). A benefit of this approach is the similar time that each publication is 'at risk' for citation. It is known that the chance of citation is highest in the first six years after publication. Studies that have not been cited in these first years, are not likely to receive many citations on the long term (59).

Other determinants of selective citation

Although citation bias specifically describes the association between citation and study outcome, also other determinants can influence the likelihood of being cited. Due to the great amount of publications in each research field, making a selection of articles to be cited is inevitable. However, not every selection will lead to biased knowledge development. This selection can be made on the basis of many determinants, which might relate to the content of the publication, the authors or the journal. Table 1 displays an overview of potential determinants, that have been linked to citation in the literature, and they have been classified based on their expected impact on knowledge development. Opposite to citation bias, study quality is considered to be a justified determinant of citation. Evidence-based medicine would increase in value and credibility if high quality studies were cited more frequently compared to low quality studies. Although it is difficult to score study quality, especially when different research designs are being used, checklists to score study quality are available for a number of study designs (60, 61). Factors that relate to study content, which might be used as proxy for study quality, are study design and sample size. Factors related to the author which might be related to citation are the gender, affiliation and authority of the author. With regard to gender, previous research suggested that the likelihood of being cited might be higher for men compared to women (54). In terms of affiliation, we expect that for-profit affiliations will lead to lower chance of citation compared to universities, in line with lower perceived trust in industry funded research (62). Authority of the author and journal impact factor are expected to have a strong impact on the chance of citation, since they are also often used metrics in promoting researchers and awarding research grants. The number of affiliations is expected to be positively related with the chance of citation for the obvious reason that more affiliations lead to a bigger network of related researcher and thereby potentially more attention for the publication. With regard to the continent where the study is performed, previous research showed that especially North American authors tend to cite each other, while leaving out others (63). As depicted in table 1, most variables are not scored as justified or unjustified determinant of citation. The effect that each of these factors has on knowledge development is expected to depend on the research field and on field-specific circumstances.

Table 1: Overview of potential determinants of selective citation, grouped on the basis of their consequences for the development of science

Justified determinants	Grey area	Unjustified determinants
Study quality	Study design	Study outcome
	Sample size	
	Journal impact factor	
	Number of references	
	Authority of the author	
	Funding source	
	Title of the publication	
	Continent	
	Gender	
	Affiliation	
	Number of affiliations involved	
	Self-citation	

Potential consequences of selective reporting of research

As stated previously, selective reporting is a questionable research practices. QRPs are known to impact scientific knowledge development in several ways. They undermine the validity and reliability of research, they weaken the trust of colleagues in the performed research, they cause a waste of research funds and resources, and ultimately they might lead to harmful decisions (64).

The potential consequences of selective citation are difficult to quantify and can take place on different levels. Partly, the consequences of citation bias can be similar to those of publication bias and outcome reporting bias. By disregarding counter-evidence and lacking communication among scientists, unfounded consensus (65) or polarisation (66) can develop. On the long-term, citation bias might lead to ill-advised research programmes and therefore contribute to research waste (65, 67), by drawing attention to irrelevant problems. Ultimately, citation bias can spread outside of the scientific community, by presenting distorted information in the media (68), and lead to misplaced medical decisions (69). In regulatory science, selective citation is specifically dangerous because in this stage knowledge gets translated into policy.

Figure 1 displays a schematic overview of scientific knowledge development from the production of primary data to making a evidence-based decision. The phase of science-based decision making can refer to many situation, such as the implementation of a new clinical treatment standard, deciding on threshold levels for exposure to chemical substances or implementing new legislation. As a common denominator, these decisions are all intended to be based on the best scientific knowledge available, which largely depends on the quality and completeness of reported evidence. Systematic reviews and meta-analyses are useful for assessing the current scientific consensus,

since they systematically search for all available evidence in a field, summarize these findings and interpret the overview of all evidence. We could argue that citation bias is mostly directly harmful when it happens in review articles, since they have the most direct impact on decision making. When empirical papers cite selectively to other articles, its impact might be not as substantial, because they still produce data of their own and because normally do not directly influence important decisions. On the other hand, selective citations in empirical papers, might drive the agenda-setting in a field into a certain direction by suggesting there might be an interesting association.

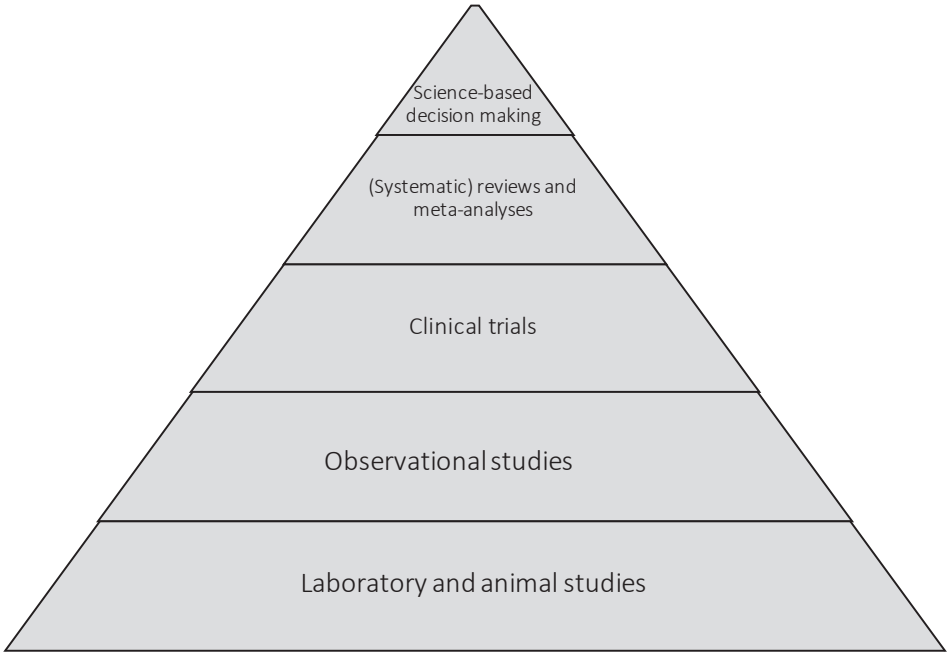


Figure 1. Schematic overview of how knowledge evolves from primary data to decision making in the biomedical sciences

Knowledge dissemination outside of science

A balanced and complete development of scientific knowledge is not only important for the academic world. As stated earlier, scientific information functions as a basis in all kinds of decision-making processes, such as the development of new technologies or medical therapies. Therefore, regulators and health care workers rely on the quality of published information. With the growing concern about questionable research practices and the so-called replication crisis, also the credibility and validity of the evidence that is used as the basis for decision-making processes is debated (70).

One of the processes in which science plays a central role, is policy making. In this dissertation, we will focus on policy decisions in the field of food law. European regulations are in place to assure all foods on the European market are safe. To assure this safety, and to regulate the appearance of new food products on the market, a scientific assessment is being conducted by the European Food Safety Authority (EFSA). The basic principles of this risk assessment is the independent, objective and transparent assessment of scientific literature (71). Clearly, this relies heavily on the reported scientific literature and therefore it will be influenced by publication bias and outcome reporting bias. But also in conducting the risk assessment, there is a risk of citation bias introduced by EFSA. This might be introduced by not using a systematic search strategy, or by excluding certain studies for other reasons.

Aim of the thesis

This thesis aims to discuss various aspects of selective reporting of research and its effect on policy making. The major part of this dissertation will discuss the occurrence and determinants of selective citation. Our main hypothesis was that citation bias is present in biomedical research. As a secondary aim in these studies, we assessed other determinants of selective citation. To put the topic of selective citation and citation bias in a broader perspective of knowledge development, we have also studied outcome reporting bias and selective citation in a risk assessment document by EFSA. By studying these three levels of knowledge development, we aim to create a broader understanding of selective reporting and its potential impact on the development of knowledge.

Study design: Sound Science project

The citation network analyses presented in this dissertation are part of a bigger project, called the Sound Science project. Since citation analysis is a relatively new field of research, no standardized method can be used yet. This thesis includes several case studies to assess the occurrence of citation bias and to identify other determinants of selective citation. Ultimately, these case studies will be combined to identify general determinants of citation, which have been found applicable in all case studies, and topic specific determinant of citation. For the methodology of the citation analyses, we have learned from previously reported studies in integrated them into a new method. Our methods can be described as follows:

1. An extensive literature search is conducted in Web of Science – Core Collection. Although this database will not be sufficient to identify all available evidence, it is the only database that enables us to download the references that are needed to set up the citation network. A broad search strategy is applied, in an attempt to in-

clude as many publications as possible. Article selection takes place based on the title, abstract and where needed the full text, and is done in duplo.
2. After the relevant publications have been identified, each publication needs to be scored on all potential determinants of citation listed in the previous paragraphs. The code book for this is laid down prior to the study in a study protocol. Also this step is performed in duplo, to increase the reliability of the study.
3. As the third step, the citation network has to be created. This is done by specialized software that has been designed to visualize and analyse networks (72). By downloading the publications and intermediate citations from Web of Science, the network of all performed citations gets created.
4. For the statistical analysis also an overview of all *potential* citation pathways is required. A potential citation pathway exists when there is a positive time different between the online publication date of the cited publication and the submission date of the citing publication.
5. With regards to the statistical analysis, a random effect logistic regression is performed. The outcome of interest is the chance of being cited, which is determined by comparing the number of potential citation pathways with the number of actually performed citations. A random effect model is applicable since the data structure consists of two levels: the level of the publication and the level of the citation. Since multiple citations are coming from one common publication, citations are not completely independent from each other. By making use of a random effect model, we deal with this hierarchical data structure.

Overview of dissertation

This dissertation started with a study on outcome reporting bias in the literature on phthalates, as described in **chapter 2**. Study protocols have been compared to the publications that report the study findings, and thereby we aimed to assess the completeness of reporting in the epidemiological studies on phthalates. Additional to the written study protocols, we requested interviews with the researchers to gain insight into the research process and potential discrepancies between the protocol and the publication. In **chapter 3**, we report a systematic review on citation bias. Studies from a variety of domains are displayed, which use different methods to test for citation bias. Building on this previous research, we have developed a new method to study citation bias, where we adjust for the total number of potential citations. Furthermore, we look at selective citation in a broader sense, by including other determinants than study outcome. **Chapters 4, 5 and 6** report citation network analyses on three different research areas. By means of these case studies, we tested our methodology and assessed determinants of selective citation. In chapter 4 we start with studying the association between industrially produced trans fatty acid intake and its effect on serum LDL- and HDL-cholesterol.

Chapter 5 describes a citation analysis on the epidemiological literature on phthalates. In chapter 6 the citation analysis includes the epidemiological literature on bisphenol A. By applying this method to three different research fields, we aimed to assess whether a general pattern of selective citation can be found, or if determinants of selective citation are very field specific. In chapter 7 we shift our attention outside of science, by looking at the occurrence and impact of selection bias in the risk assessment procedure of the European Food Safety Authority.

References

1. Ioannidis JP. Meta-research: Why research on research matters. PLoS biology. 2018;16(3):e2005468.
2. Sackett DL. Evidence-based medicine. Seminars in perinatology; 1997: Elsevier.
3. Ioannidis JP, Fanelli D, Dunne DD, Goodman SN. Meta-research: evaluation and improvement of research methods and practices. PLoS biology. 2015;13(10):e1002264.
4. Levelt WJ, Drenth P, Noort E. Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. 2012.
5. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PloS one. 2009;4(5):e5738.
6. Fanelli D. The black, the white and the grey areas: Towards an international and interdisciplinary definition of scientific misconduct. Promoting research integrity in a global environment. 2011:79-90.
7. Wicherts JM, Veldkamp CL, Augusteijn HE, Bakker M, Van Aert R, Van Assen MA. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. Frontiers in Psychology. 2016;7:1832.
8. Steneck NH, Mayer T, Anderson MS, Kleinert S. The origin, objectives, and evolution of the World Conferences on Research Integrity. Scientific Integrity and Ethics in the Geosciences. 2017;73:1.
9. Mayer T, Steneck N. Final Report to ESF and ORI First World Conference on Research Integrity: Fostering Responsible Research. 2007;4:2010.
10. Resnik DB, Shamoo AE. The singapore statement on research integrity. Accountability in research. 2011;18(2):71-5.
11. Kleinert S. Singapore Statement: a global agreement on responsible research conduct. The Lancet. 2010;376(9747):1125-7.
12. Duarte EF. Montreal statement on research integrity in cross-boundary research collaborations. Epidemiologia e Serviços de Saúde. 2014;23:185-6.
13. ALLEA. The European Code of Conduct for Research Integrity: European Science Foundation; 2011.
14. Godecharle S, Nemery B, Dierickx K. Guidance on research integrity: no union in Europe. The Lancet. 2013;381(9872):1097-8.
15. Drenth PJ. A European code of conduct for research integrity. Promoting Research Integrity in a Global Environment. 2012:161.
16. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS medicine. 2007;4(10):e296.
17. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.
18. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS medicine. 2009;6(7):e1000100.
19. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS medicine. 2010;7(2):e1000217.
20. Simera I, Altman DG. Writing a research article that is" fit for purpose": EQUATOR Network and reporting guidelines. Evidence-based medicine. 2009;14(5):132.
21. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. The Lancet. 2014;383(9912):156-65.
22. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. The Lancet. 2014;383(9912):166-75.
23. Salman RA-S, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. The Lancet. 2014;383(9912):176-85.
24. Chan A-W, Song F, Vickers A, Jefferson T, Dickersin K, Gøtzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. The Lancet. 2014;383(9913):257-66.

25. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. The Lancet. 2014;383(9913):267-76.
26. Neylon C, Wu S. Article-level metrics and the evolution of scientific impact. PLoS biology. 2009;7(11):e1000242.
27. Campbell P. Escape from the impact factor. Ethics in science and environmental politics. 2008;8(1):5-7.
28. Garfield E. Is citation analysis a legitimate evaluation tool? Scientometrics. 1979;1(4):359-75.
29. Dijstelbloem H, Huisman F, Miedema F, Mijnhardt W. Why science does not work as it should. And what to do about it. Science in Transition Position Paper. 2013.
30. Hicks D, Wouters P, Waltman L, Rijcke Sd, Rafols I. Bibliometrics: the Leiden Manifesto for research metrics. 2015.
31. Ravenscroft J, Liakata M, Clare A, Duma D. Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. PloS one. 2017;12(3):e0173152.
32. Hoppeler H. The San Francisco declaration on research assessment. The Company of Biologists Ltd; 2013.
34. Segal JZ. Strategies of influence in medical authorship. Social science & medicine. 1993;37(4):521-30.
35. Rosenthal R. The file drawer problem and tolerance for null results. Psychological bulletin. 1979;86(3):638.
36. Hopewell S, Clarke MJ, Stewart L, Tierney J. Time to publication for results of clinical trials. Cochrane database of systematic reviews. 2007.
37. Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics. 2012;90(3):891-904.
38. Song F, Eastwood A, Gilbody S, Duley L, Sutton A. Publication and related biases: a review. Health technology assessment. 2000;4(10).
39. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS one. 2008;3(8):e3081.
40. Gerber A, Malhotra N. Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. Quarterly Journal of Political Science. 2008;3(3):313-26.
41. Doucouliagos C. Publication bias in the economic freedom and economic growth literature. Journal of Economic Surveys. 2005;19(3):367-87.
42. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010;14(8):1-193.
43. Easterbrook PJ, Gopalan R, Berlin J, Matthews DR. Publication bias in clinical research. The Lancet. 1991;337(8746):867-72.
44. Franco A, Malhotra N, Simonovits G. Publication bias in the social sciences: Unlocking the file drawer. Science. 2014;345(6203):1502-5.
45. Rothstein HR, Sutton AJ, Borenstein M. Publication bias in meta-analysis: Prevention, assessment and adjustments: John Wiley & Sons; 2006.
46. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials. Gov: a cross-sectional analysis. PLoS medicine. 2009;6(9):e1000144.
47. Chalmers I, Glasziou P, Godlee F. All trials must be registered and the results published. British Medical Journal Publishing Group; 2013.
48. Dirnagl U, Lauritzen M. Fighting publication bias: introducing the Negative Results section. SAGE Publications Sage UK: London, England; 2010.
49. Van Den Bogert CA, Souverein PC, Brekelmans CT, Janssen SW, Van Hunnik M, Koëter GH, et al. Occurrence and determinants of selective reporting of clinical drug trials: design of an inception cohort study. BMJ Open. 2015;5(7):e007827.
50. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365.
51. Chan A-W, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. BMJ. 2005;330(7494):753.
52. Hutton J, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2000;49(3):359-70.

53. Nigel Gilbert G. Referencing as persuasion. Social studies of science. 1977;7(1):113-22.
54. Bornmann L, Daniel H. What do citations measure? A review of studies on citing behavior. Journal of Documentation. 2006;64(1):45-80.
55. Seglen PO. Citations and journal impact factors: questionable indicators of research quality. Allergy. 1997;52(11):1050-6.
56. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Research Integrity and Peer Review. 2016;1(1):17.
57. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. BMJ. 2009;339:b2680.
58. Fanelli D. Positive results receive more citations, but only in some disciplines. Scientometrics. 2013;94(2):701-9.
59. Burrell QL. Will this paper ever be cited? Journal of the American Society for Information Science and Technology. 2002;53(3):232-5.
60. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. European journal of epidemiology. 2010;25(9):603-5.
61. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of clinical epidemiology. 1998;51(12):1235-41.
62. DeAngelis CD. Conflict of interest and the public trust. Jama. 2000;284(17):2237-8.
63. Allik J. Bibliometric analysis of the journal of cross-cultural psychology during the first ten years of the new millennium. Journal of Cross-Cultural Psychology. 2013;44(4):657-67.
64. Steneck NH. Fostering integrity in research: Definitions, current knowledge, and future directions. Science and engineering ethics. 2006;12(1):53-74.
65. Greenberg SA. How citation distortions create unfounded authority: Analysis of a citation network. BMJ. 2009 Jul;339.
66. Trinquart L, Johns DM, Galea S. Why do we think we know what we know? A metaknowledge analysis of the salt controversy. International Journal of Epidemiology. 2016 Feb;45(1):251-60.
67. Andrade NS, Flynn JP, Bartanusz V. Twenty-year perspective of randomized controlled trials for surgery of chronic nonspecific low back pain: Citation bias and tangential knowledge. Spine Journal. 2013;13(11):1698-704.
68. Chapman S, Ragg M, McGeechan K. Citation bias in reported smoking prevalence in people with schizophrenia. Aust NZ J Psychiat. 2009;43(3):277-82.
69. Koren G, Nickel C. Perpetuating fears: bias against the null hypothesis in fetal safety of drugs as expressed in scientific citations. J Popul Ther Clin Pharmacol. 2011 2011 (Epub 2011 Jan;18(1).
70. Wallach JD, Gonsalves GS, Ross JS. Research, regulatory, and clinical decision-making: the importance of scientific integrity. Journal of clinical epidemiology. 2018;93:88-93.
71. Commission E. Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal of the European Communities. 2002;31:1-24.
72. Van Eck NJ, Waltman L. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. Journal of Informetrics. 2014;8(4):802-23.

Outcome Reporting Bias in observational epidemiology studies on phthalates

Gerard MH Swaen, **Miriam JE Urlings** and Maurice P Zeegers

This chapter was published as a rapid report in *Annals of Epidemiology*:

Swaen GM, Urlings MJ, & Zeegers MP. Outcome reporting bias in observational epidemiology studies on phthalates. Annals of epidemiology. 2016; 26(8): 597-599.

Abstract

Epidemiological literature on phthalates shows a great variety in the type of health outcomes it is associated with. To interpret this literature adequately, it is important to know what was the initial research plan. In the current study we aimed to make a comparison between the research plan as described in a study protocol and the results presented in the final publication. Firstly, we identified 158 epidemiological publications on phthalates in relation to human health. The corresponding authors of these publications were contacted with a request to share their study protocol and an invitation for a telephone interview. Out of all corresponding authors, 47 authors were willing to participate in the interview. However, only 22 of them were able to share a study protocol and therefore were able to complete the whole study. The chance of sharing a protocol was related to the reported study outcome. Corresponding authors reporting positive associations between phthalates and health outcomes were three times less likely to provide a copy of their protocol compared to authors reporting no positive assocation. Additionally, we scored the quality of the provided protocols and concluded that 21 out of 22 protocols showed too little detail in order to be reproducible.

When a publication reports only a selection of the performed statistical analyses, mostly limited to the significant findings, this is known as outcome reporting bias. Based on the current study we can conclude that epidemiological research on phthalates is at risk for outcome reporting bias, given the low rate of detailed study protocols.

Introduction

Phthalates are used in inter alia cosmetics, flooring adhesives, medical tubings, toys and food packaging materials. A number of reviews of the epidemiology on phthalates have been published with conflicting results (1-10). As also noted by previous reviewers, the phthalate epidemiology literature is very diverse (9). It covers a wide range of health outcomes and a range of research designs are used. In such instances it is essential to know what the investigators had in mind when conducting the study, i.e. what was stated in the research protocol. We therefore contacted all corresponding authors and invited them to participate in a short interview and asked for a copy of the protocol. We clarified that the responses to the interview and the copy of the protocol would remain confidential and would only be used for the purpose of our study. Our study had the following specific aims: 1. Assess the completeness of reporting, 2. Assess the quality of the underlying study protocols, 3. Assess the concordance between the published articles and the underlying protocols and 4. Assess the determinants of protocol provision. We registered our study protocol at the PROSPERO website (registration number CRD42015016017) and our protocol was placed on the website of our Department. The protocol describes the research methods that we applied, including how the corresponding authors were contacted, interviewed and requested for a copy of the protocol. It also provided the scoring sheet and the criteria for assessing the received protocols.

Methods

The literature search yielded 158 journal articles on epidemiological studies on phthalates. Study methodology characteristics and study outcome were scored by GS and MJEU individually and compared and finalized. Via an e-mail message and reminder the corresponding authors, were invited to participate in a telephone interview. The survey covered the study objective, study population, exposure measurement, health outcome parameter, the statistical analysis, but also some items about the corresponding author's career and working environment. If no reply was received they were contacted by telephone. If the author consented a copy of the project proposal, analytical description, grant submission, protocol, project description or ethical committee review submission, all designated here as "study protocol" was requested. The interviews were conducted by GS and MJEU was present. The telephone interview was pre-tested on five unrelated publications and their corresponding authors.

Results

With 45 (28.5%) out of the 158 corresponding authors it was not possible to establish any contact, despite multiple attempts via telephone and two e-mail invitations. With 113 corresponding authors either a telephone conversation or a meaningful e-mail exchange was established (see Table 1).

Table 1: Participation of corresponding authors in the interview and protocol provision of the 158 included observational studies on phthalates

Participation status corresponding author	**N (%)**	
No contact established	**45**	**(28.5%)**
Refused - all combined	**66**	**(41.8%)**
Refused – Specified for reason:		
without providing a reason	9	(5.6%)
Too busy	9	(5.6%)
Had methodological objections to our study	4	(2.5%)
Saw conflict of interest in our study	2	(1.3%)
Had both methodological objections and saw conflict of interest	31	(19.6%)
Stated all information was in the publication	4	(2.5%)
Had already been interviewed for earlier publication	2	(1.3%)
Other reason including that the protocol was confidential	5	(3.2%)
Agreed to participate in the interview	**47**	**(29.7%)**
But had no protocol	16	(10.1%)
First agreed to participate but no response after request for protocol	8	(5.1%)
Full participation but later withdrawal because interviewed author did not feel comfortable	1	(0.6%)
Full participation	22	(13.9%)
Total	**158**	**(100.0)**

Initially 47 corresponding authors agreed to be interviewed. However, after having been requested a copy of the protocol, 16 corresponding authors indicated that they did not have a protocol or could not send a copy. 8 Corresponding authors did not reply to our request for the protocol after having consented to participate. 23 Corresponding authors were interviewed and were willing to share their protocol. One corresponding author out of the 23 withdrew from our project after having been interviewed. This data was excluded from our analysis.

For 43 publications we received information whether a protocol was present or not. For 22 of these a protocol was provided to us. 16 Out of the 43 stated that they did not have a protocol, of which 3 had been lost. Corresponding authors reporting positive associations between phthalates and health outcomes (based on the authors' own conclusions) were three times less likely to provide a copy of their protocol (OR=0.31 95% CI: 0.11-0.86). Associations between other study characteristics and protocol provision, such as year of publication, affiliation, funding source and number of associations tested were not statistically significant. An explorative analysis not foreseen in our

study protocol showed that corresponding authors of publications based on NHANES data were also less likely to provide their protocols, (OR= 0.83 95% CI: 0.77-0.90).

Overall we assessed 21 protocols as having insufficient detail to adequately describe the project. These lacked descriptions of the study population to be selected, how exposure would be measured, what type of statistical analysis would be done, and how confounding factors would be selected and treated, and in most instance combinations of these. Several protocols did not even mention phthalates as exposure variable, but only mentioned wide terms as environmental contaminant or exposure to environmental chemicals. Rule of thumb in the assessment was whether the protocol provided sufficient detail to get an understanding of how the study would be conducted. Given this general lack of detail we refrained from testing our third research aim: assessing concordance between the study protocol and publication.

Discussion

We studied 158 observational epidemiology studies on phthalates with the four following research aims: 1. Assess completeness of reporting, 2. Assess the quality of the underlying protocols, 3. Assess concordance between the published articles and the underlying protocols and 4. Assess the determinants of protocol provision.

Our study yielded insight in determinants of protocol provision. Corresponding authors of publications reporting a positive association study outcome were about three times less likely to participate in our study, which was statistically significant.

To our knowledge this is the first study on observational epidemiology research in which corresponding authors were requested to provide their protocol and participate in a survey on how the research was conducted. In the field of clinical trials some studies have been conducted with similar methodologies. Chan conducted two studies in which trialists were surveyed (11, 12). Both studies provided evidence of selective reporting in clinical trials. Similarly Smyth et al contacted corresponding authors of 268 clinical trials and also found evidence of selective reporting (13). Recently a series of articles was published on increasing the value and reducing waste in biomedical research (14). In one of these articles a strong recommendation was made to make publicly available full protocols, analysis plans and raw data (15). Our study underpins the need for these changes. Following the field of clinical trials we recommend observational epidemiology studies to be based on a detailed protocol that is publicly available. Increased transparency in observational epidemiology studies will contribute to the still high credibility of this type of research. It will also facilitate detecting outcome reporting bias provided protocols contain sufficient detail. Writing a study protocol and setting up a process of good project documentation and archiving are a part of responsible research conduct and it is clear this will require time and effort. Given the selective participation, the limited number of provided protocols and the large portion of studies con-

ducted without a protocol we hesitate to use this literature as a reliable basis for a formal systematic review. Our research could enhance awareness for the need of responsible research conduct in observational epidemiology studies, similar to the clinical trials area and stimulate the discussion about the need of protocoled research in this field as well.

References

1. Meeker JD. Exposure to environmental endocrine disrupting compounds and men's health. Maturitas. 2010; 66:236-41.
2. Tang-Peronard JL, Andersen HR, Jensen TK, Heitmann BL. Endocrine-disrupting chemicals and obesity development in humans: a review. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2011;12:622-36.
3. de Cock M, Maas YG, van de Bor M. Does perinatal exposure to endocrine disruptors induce autism spectrum and attention deficit hyperactivity disorders? Review. Acta paediatrica. 2012;101:811-8.
4. Hauser R, Calafat AM. Phthalates and human health. Occupational and environmental medicine. 2005;62:806-18.
5. Lottrup G, Andersson AM, Leffers H, Mortensen GK, Toppari J, Skakkebaek NE, Main KM. Possible impact of phthalates on infant reproductive health. International journal of andrology. 2006;29:172-80;
6. Jurewicz J, Hanke W. Exposure to phthalates: reproductive outcome and children health. A review of epidemiological studies. International journal of occupational medicine and environmental health. 2011;24:115-41.
7. Cai H, Zheng W, Zheng P, Wang S, Tan H, He G, Qu W. Human urinary/seminal phthalates or their metabolite levels and semen quality: A meta-analysis. Environmental research. 2015;142:486-94.
8. Kay VR, Chambers C, Foster WG. Reproductive and developmental effects of phthalate diesters in females. Critical reviews in toxicology. 2013; 43:200-19.
9. Goodman M, Lakind JS, Mattison DR: Do phthalates act as obesogens in humans? A systematic review of the epidemiological literature. Critical reviews in toxicology. 2014;44:151-75.
10. Ejaredar M, Nyanza EC, Ten Eycke K, Dewey D. Phthalate exposure and childrens neurodevelopment: A systematic review. Environmental research. 2015;142:51-60.
11. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. Jama. 2004;291:2457-65.
12. Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. Canadian Medical Association journal. 2004;171:735-40.
13. Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. BMJ. 2011;342:c7153.
14. Kleinert S, Horton R. How should medical science change? Lancet. 2014;383:197-8.
15. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383:166-75.

3

Scientific Citations Favor Positive Results: A Systematic Review and Meta-analysis

Bram Duyx, **Miriam J.E. Urlings**, Gerard M.H. Swaen, Lex M. Bouter, Maurice P. Zeegers

This chapter is published in *Journal of Clinical Epidemiology*:

Duyx B, Urlings MJ, Swaen GM, Bouter LM & Zeegers MP. Scientific citations favor positive results: a systematic review and meta-analysis. Journal of clinical epidemiology, 2017;88:92-101.

Abstract

Background and Objectives: Citation bias concerns the selective citation of scientific articles based on their results. In this systematic review we brought together all available evidence on citation bias across scientific disciplines and quantified its impact.

Methods: An extensive search strategy was developed and applied to the Web of Science Core Collection and Medline, yielding 52 studies in total. We classified these studies on scientific discipline, selection method and other variables. We also performed random effects meta-analyses to pool the effect of positive versus negative results on subsequent citations. Finally, we checked for other determinants of citation as reported in the citation bias literature.

Results: Evidence for the occurrence of citation bias was most prominent in the biomedical sciences, and least in the natural sciences. Articles with statistically significant results were cited 1.6 times more often than articles with non-significant results. Articles in which the authors explicitly conclude to have found support for their hypothesis were cited 2.7 times as often. Article results and journal impact factor showed an effect on citation more often than any other reported determinant.

Conclusion: Similar to what we already know on publication bias, also citation bias can lead to an over-representation of positive results and unfounded beliefs.

Introduction

Citations are key elements in the evolution of knowledge. They enable particular research findings to survive over time and to develop into academic consensus. Given the large body of scientific literature, it is often unfeasible to cite all published articles on a specific topic, and so, some selection needs to take place. If this selection is influenced by the actual results of the article, then citation bias occurs (1).

Citation bias is considered to be a questionable research practice (QRP). QRPs are scientific misbehaviors that lie on the continuum between research misconduct (fabrication, falsification and plagiarism), and responsible conduct of research. QRPs are often not deliberate, and their individual effects are assumed to be less severe than those of research misconduct.

Nevertheless, questionable research practices are believed to occur frequently and may have a strong negative impact on the development of knowledge (2). A well-known example is publication bias, which leads to an over-representation of positive results in the scientific literature. According to a meta-analysis of surveys (3), researchers report to engage in QRPs (about 34%) much more often than in research misconduct (about 2%). Similarly, in a recent survey among researchers, selective citation was ranked as the most frequently occurring research misbehavior (4). In order to assess the potential consequences of citation bias, a proper understanding of its ubiquity is required.

Citation bias has been documented for several fields and disciplines, but to our knowledge, no systematic review exists. Our first aim was therefore to identify and assess all published evidence on citation bias, regardless of scientific discipline. Our second aim was to quantify the overall impact of article results on the likelihood of being cited.

Materials and Methods

Search strategy

All publications reporting empirical evidence on the association between article results and citation frequency were included. In order to identify these publications, we developed an extensive search strategy. Roughly, it consists of three angles:

a) "citation bias";
b) *"publication bias"* (with the restriction that it should be related to citation);
c) the combination of "research outcome" and "citation frequency".

This search strategy was applied to the Web of Science Core Collection. Because the majority of the studies on citation bias turned out to be conducted in the biomedical field, we extended this search strategy to Medline, as was laid down in our research protocol (5). Both searches were performed on 20 November 2016. Reference lists of

included publications were also checked. There was no restriction with regards to year of publication. Only published research written in English was included. The selection process was done in duplicate (BD and MJEU). Disputes were resolved by a third researcher (GMHS).

Data extraction

The following characteristics were extracted for each included study: first author, publication year, scientific discipline (social sciences, biomedical sciences, natural sciences, or multiple disciplines), article selection method (claim-specific, review-based, or journal-based), type of article included (trial, any primary data study, meta-analysis, or any type of article), operationalization of article results, other potential determinants of citation included in analysis, conclusion on the occurrence of citation bias (citation bias found, no citation bias, mixed results, or unclear), total number of articles (sample size), total number of citations, and total citation time. With citation time we mean the time period over which the citations have been accumulated. Data extraction was performed in duplicate (BD and MJEU).

For the meta-analysis we extracted or calculated additional information: the number of positive articles, the number of negative articles, the number of citations to positive articles, the number of citations to negative articles, the citation time of all positive articles together and the citation time of all the negative articles together. If necessary, we approached the authors of the citation bias studies at least twice in order to retrieve missing information.

Meta-analyses

Citation data are non-parametric. Therefore we used rate ratios to pool these data. The rate is the total number of citations within a certain time frame. The rate ratio is the ratio of the citation rates in the positive outcome articles versus the negative outcome articles. We used the inverse-variance method with random effects for pooling of the natural logarithms of the rate ratios. Four meta-analyses were performed, one for each of the following operationalizations of the article results:

1. Relationship between *statistical significance* of the results (regardless of their direction) and citation frequency. Articles with statistically significant results ($\alpha = 0.05$) are considered positive, articles with statistically non-significant results as negative.
2. Relationship between *direction* of the results (regardless of their significance) and citation frequency. Articles with results in the expected direction are considered positive, articles with results in the opposite direction as negative.
3. Relationship between *hypothesis conformity* (results being significant and in the expected direction) and citation frequency. Articles with results that are statistically significant in the expected direction are considered positive, articles with non-

significant results or with significant results in the opposite direction are considered as negative.

4. Relationship between *authors' conclusion* in the individual articles (regardless of the actual data) and citation frequency. Articles in which the authors conclude to have found support for the tested hypothesis are considered positive, articles in which the authors conclude not to have found support as negative.

Authors were contacted multiple times to request any missing information. If we could not retrieve the necessary information, we either used more specific methods to infer it, or else excluded the study from the meta-analysis.

Supporting Information

More information on the search strategy, details about the citation bias studies, methods to calculate the rate ratio, and results of sensitivity analyses can be found in the Supporting Information and in our review protocol (5). More information on the terminology we use can be found in Figure 1.

Our review can be considered as meta-meta-research. It includes different levels of research. We discern between these levels by using the following terminology throughout our manuscript:

Level 1 - An ***article*** refers to the original published work. Each article has a specific outcome (called *article results*) and *citation frequency*.

Level 2 - A ***publication*** is a published work that studies citation bias in the network of included articles. (Publications that are not primarily about citation bias but measure both article results and citation frequency, are also included.) A publication can report multiple ***studies***.

Level 3 - Our systematic ***review*** investigates all *publications* on citation bias. (Our meta-analyses use *study* as the unit for analysis, as different studies within a publication can yield different rate ratios.)

Figure 1. Adopted terminology and levels of research.

Results

Our search strategy identified 47 publications (Figure 2). Three of these publications comprised two or more empirical studies, yielding a total of 52 separate studies on citation bias, and including the citation data of more than 13,000 articles on various topics. Most of the 52 studies found evidence for citation bias in their field: 29 showed a clear effect of outcome on citation against 11 studies that showed no effect (and 12 with mixed results). The direction of citation bias was fairly consistent: with some ex-

ceptions (6, 7), most studies reported that positive articles were cited more often than negative articles (Table 1).

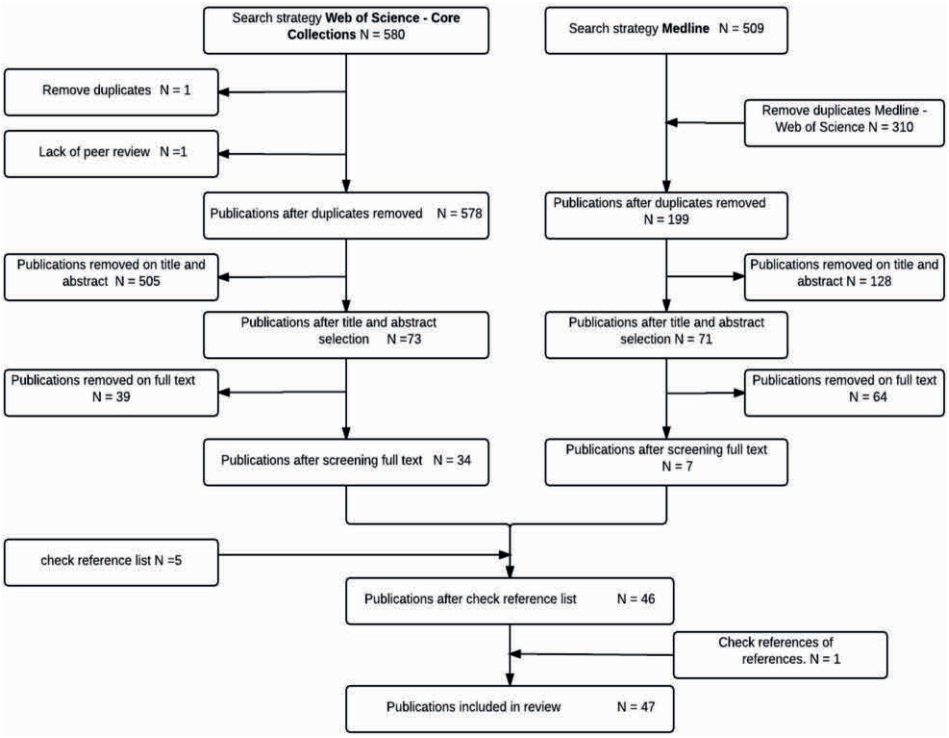


Figure 2. Flow diagram of the article selection process.

Table 1. Number of studies on citation bias, by discipline, selection method, and outcome (number of studies in meta-analyses)

Scientific discipline / Article selection method	Found support for citation bias ?			Total number of studies in review (meta-analysis)
	Yes	No	Mixed/Unclear	
Social	6 (2)	0 (0)	1 (1)	7 (3)
Biomedical	21 (14)	8 (4)	9 (6)	38 (24)
Natural	2 (0)	3 (0)	1 (0)	6 (0)
Multiple	0 (0)	0 (0)	1 (1)	1 (1)
Claim-specific	10 (6)	2 (1)	2 (2)	14 (9)
Review-based	11 (6)	4 (2)	5 (3)	20 (11)
Journal-based	6 (3)	1 (0)	2 (1)	9 (4)
Other selection	2 (1)	4 (1)	3 (2)	9 (4)
Total *	29 (16)	11 (4)	12 (8)	52 (28)

Notes: **support for citation bias** as stated by the authors of the included publications. Some publications present multiple studies with different results; therefore we present the number of studies in this table. ***** 28 of the 52 studies were eligible to be included in at least one of the meta-analyses. Inclusion in the meta-analyses does not seem to depend on support for citation bias ($\chi^2(2) = 2.2$, $p = .34$).

The majority of the studies are biomedical (7-42), but some also concern the social (43-49) and natural sciences (6, 50, 51), or a combination of these (52). The biomedical studies ranged from highly specific fields - such as the relationship between job strain and cardiovascular disease (16), or the treatment of chronic non-specific lower-back pain (8) - to broader categories like cardiovascular medicine (10). Most of these studies provided clear evidence for citation bias. Citation bias was also identified within the psychological (44, 46-49) and economic (43, 45) literature, but the evidence for citation bias in the natural sciences (mostly ecology) (6, 50, 51) seemed less convincing. This difference between scientific disciplines was not statistically significant ($\chi^2(4) = 5.7$, $p = .22$, Table 1).

Apart from scientific discipline, these studies also differ in their article selection approach. 14 of the 52 studies have used a claim-specific approach to study citation bias (8, 12, 19-25, 34, 40, 46, 47, 49). Their aim was to identify all the relevant literature about a specific claim and to study citation behavior within that network of articles. Another approach is to select all the articles from a specific journal or database for one or more years. Nine studies used this approach (10, 17, 18, 26, 29, 31, 43, 45, 52), whereas 20 other studies based their selection on a previously published review or reviews (6, 11, 15, 16, 25, 27, 30, 33, 35, 36, 38, 39, 41, 44, 48, 50, 51).

Claim-specific research on citation bias could be prone to selection bias as the studied claims might have been chosen according to an already existing concern of selective citation. This could potentially lead to an overestimation of the citation bias prevalence. However, this is unlikely as the journal-based selection studies showed very similar results (67% showing clear support for citation bias against 71% of the claim-specific studies). Evidence from the review-based selection studies was slightly less convincing (55% showing clear support for citation bias, and 25% showing no citation bias). This difference between selection methods was not statistically significant ($\chi^2(4) = 1.2$, $p = .88$, Table 1).

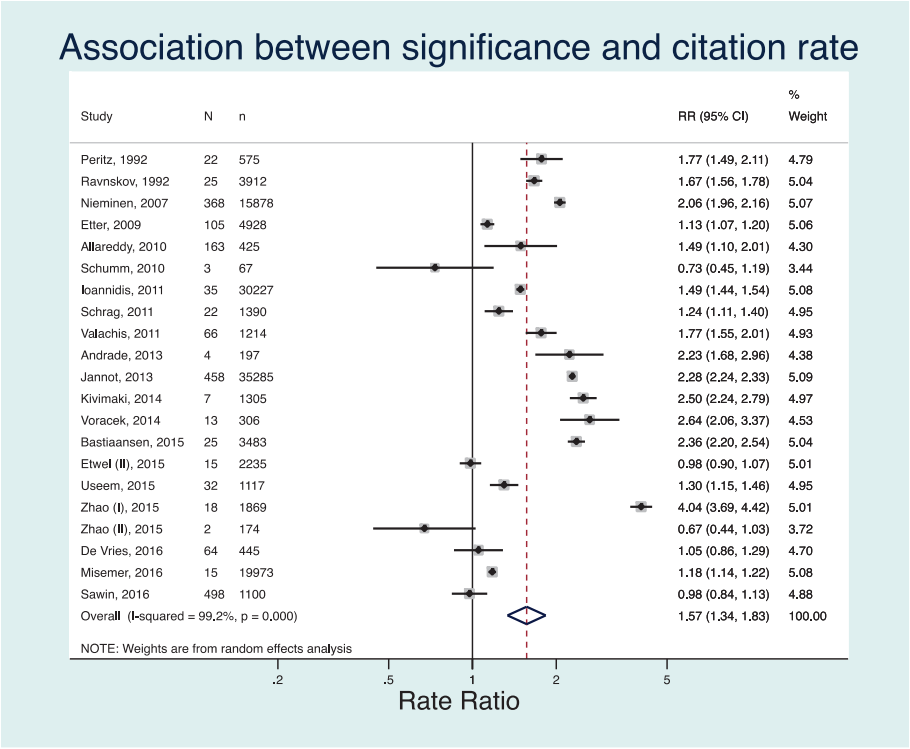
Meta-analyses

Next to identifying and assessing the published evidence on citation bias, our second aim was to quantify the overall impact of the results reported in an article on how often it is cited. If available, we used data already present in the publications. For the remaining 35 publications we contacted the authors to provide the necessary information. Despite several attempts, only 15 authors were able and willing to comply.

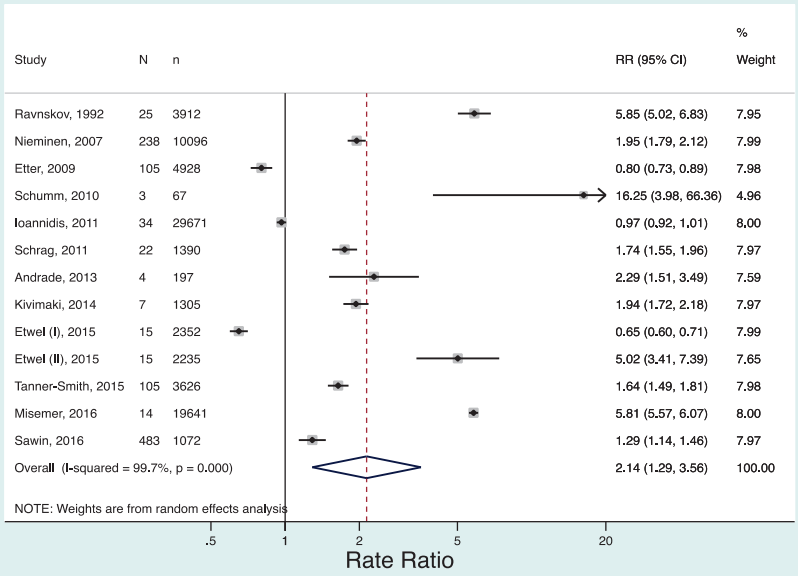
Twenty one studies provided sufficient data to calculate a citation rate ratio for *statistical significance*, and to pool their results in a random effects meta-analysis (Figure 3a). This analysis showed that statistically significant studies were cited 1.6 times as often as non-significant studies. Sensitivity tests did not reveal any differences between article selection methods or article type (Figures S5 and S6). Although the heterogeneity

was high and the rate ratio varied between studies, almost all included studies showed a positive relationship for articles with significant results.

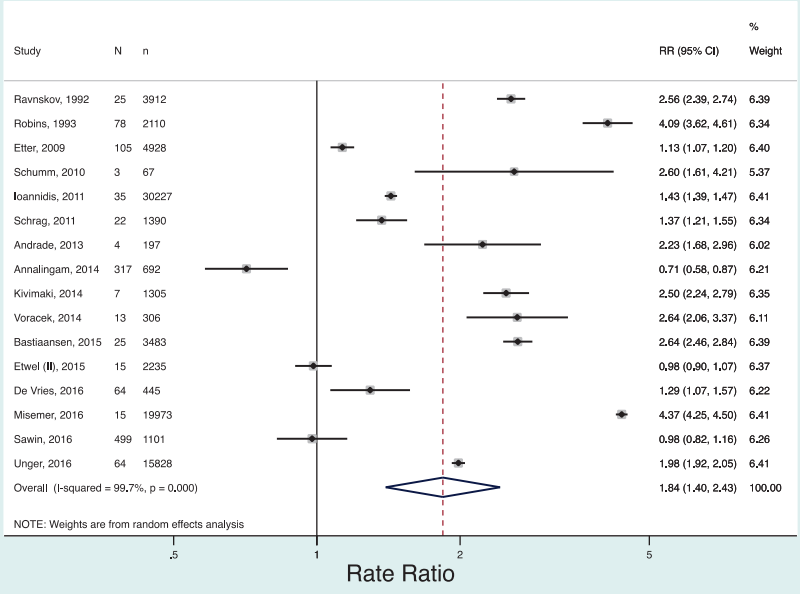
Statistical significance in itself is not enough to imply support for a tested hypothesis, as this would also depend on the direction of the findings. In order to check if some aspects of article results drive citation more than others, additional meta-analyses were performed. These analyses, one on the *direction of results* and one on *hypothesis conformity*, showed similar estimates as the one on statistical significance (with pooled ratio ratios of 2.1 and 1.8 respectively, Figures 3b and 3c).



Association between direction and citation rate



Association between hypothesis conformity and citation rate



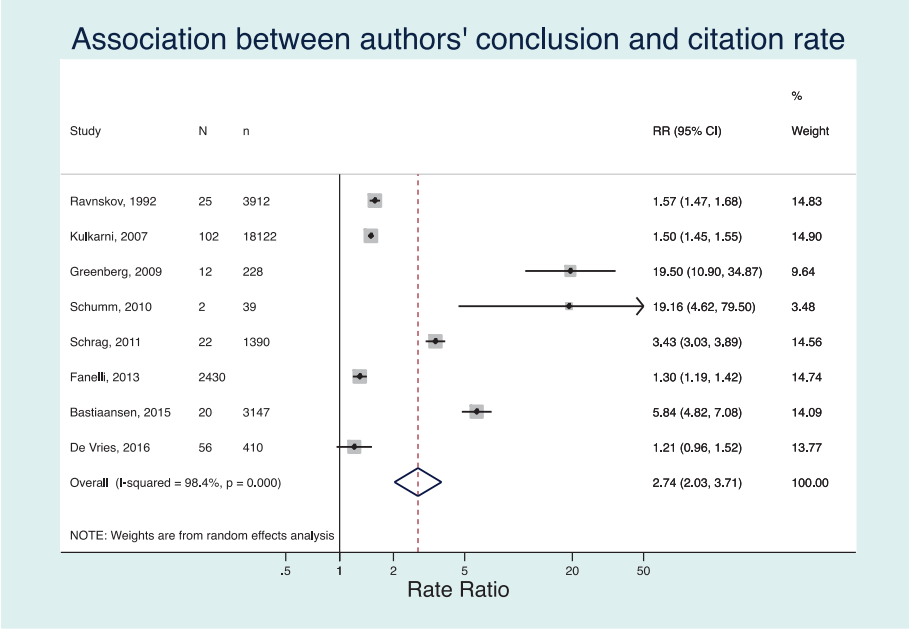


Figure 3. Forest plot of association between article results and citation rate.

NOTES: **RR** Rate Ratio, **CI** Confidence Interval, **N** number of articles, **n** number of citations.

Authors' Conclusion. The previous operationalizations of article results are all data-driven. The decision to cite an article could also be based on the authors' interpretation of the results rather than on the results themselves. There were in total eight studies on citation bias that looked at the conclusion as stated by the original authors. A meta-analysis including all these eight studies showed that original articles with a positive conclusion were cited 2.7 times more often (Figure 3d).

All our meta-analyses demonstrated that positive articles were cited about 1.5 to 2.5 times more often than negative articles. To check whether this is representative for all published research on citation bias, we looked again at the 23 studies that were not included in any of the meta-analyses. 52% of these studies showed evidence for citation bias (versus 59% of the included studies), while 30% (14%) concluded there was no evidence for citation bias and 17% (28%) provided mixed or unclear evidence (Table 1). The difference between studies that were included in the meta-analyses and those that were not included, was small and not statistically significant ($\chi^2(2) = 2.4$, $p = .31$). We therefore believe the double citation rates for positive studies to be representative for all published research in our systematic review.

Other determinants of citation

To evaluate which other factors determine the number of citations, we identified all potential determinants of citation as analyzed in the 47 publications of our review, and scored how often they showed an impact on citation frequency (Table 2). In these publications, article results (76%) and also journal impact factor (89%) were more often associated with citation frequency than justifiable determinants such as research quality (17%), sample size (29%) and research design (50%). Future multivariate analyses are needed to test if this result can be generalized.

Table 2. Determinants of citation

Determinant	Number of publications with determinant included in analysis*	Number of publications with determinant showing a significant effect on citation count			Percentage of publications in which determinant shows an effect on citation count**
		Confirms	Mixed/Unclear	Rejects	
Article Results	46 ***	26 ****	12	8	76 %
Impact Factor	19	16	1	2	89 %
Sample Size	19	4	5	10	29 %
Research Design	11	4	3	4	50 %
Research Topic	10	6	3	1	86 %
Country of Author(s)	10	5	1	4	56 %
Research Quality	8	1 *****	2	5	17 %
Number of Authors	7	4	2	1	80 %
Funding Source	7	4	2	1	80 %
Affiliation of Author(s)	3	0	1	2	0 %
Authority of Author(s)	2	1	0	1	50 %

Notes: The classification of these determinants is based on findings from the 47 publications included in this review because it was not always possible to distinguish these determinants for each separate study (e.g. (6)) . ***** Mostly based on univariate analyses; ****** Mixed and unclear results are ignored in the calculation of this percentage. E.g. Perc(Article Results) = 26/(26+8)*100% = 76%; ******* One publication had measured the outcome and citation frequency of the included articles, but did not analyze the relationship between them; ******** One publication confirmed citation bias but in opposite direction, with negative articles being cited more often (7); ********* Only one publication showed an effect of quality-related measures; it showed that lower quality was associated with a higher citation frequency (43).

Discussion and Conclusion

Citation bias seems to exist throughout the sciences. It is most prominent in the biomedical sciences with many studies in different fields showing evidence for citation bias. The evidence in the social sciences is also convincing, although it is based on fewer studies. The evidence in the natural sciences is scarcer and so far less convincing. Our meta-analyses show that positive articles are cited about two times more often than

negative ones. Our results suggest that citations are mostly based on the conclusion that authors draw rather than the underlying data.

To our knowledge, this is the first time that all empirical literature on the relationship between article results and citation has been systematically investigated, and that the magnitude of citation bias has been summarized in a pooled estimate. There is one earlier review, but no search strategy had been specified and only a few publications were included (1).

There is one other study that compared the occurrence of citation bias in multiple scientific disciplines (52). This empirical study by Daniele Fanelli is also included in our review. His approach was to randomly select a number of articles published between 2000 and 2007 and score them on outcome, number of citations, and discipline. This what we call journal-based approach is powerful, but it has its caveats compared to the claim-specific approach described before. To give a fictional example, let us look at the health effects of fruits. Study A, on the health effects of apples, shows promising, positive results and this gives rise to a high number of additional studies on apples. Study B, on the health effects of oranges, shows negative results instead, and does not inspire more studies on oranges. It is likely that Study A will be cited more often, but is this because of the positive results? Or is it because there are more follow-up studies on the same topic that are likely to cite each other? A journal-based selection approach cannot rule out this alternative explanation for citation bias, because, basically, it compares apples with oranges. In addition to Fanelli's study, our review has allowed us to check whether the occurrence of citation bias depends on the article selection approach. It turned out that this is not the case.

The majority of citation bias studies are performed in the biomedical sciences. This might reflect a higher awareness for this kind of biases compared to other disciplines rather than a higher prevalence. In fact, the biomedical field seems generally more advanced in employing initiatives to counter reporting bias and publication bias as reflected in the use of research protocols and preregistration of clinical trials (e.g. (53-55)).

The scientific process stands or falls by a balanced representation of the available research. Citation bias distorts this balanced representation and may lead to false beliefs (e.g. (56)). The good news is that there is a self-correcting mechanism in the form of systematic reviews, which ideally take all published evidence into account regardless of whether it has been cited before or not. Still, even though systematic reviews and meta-analyses are often regarded as providing the best form of evidence, they can be flawed and even misleading (e.g. (57)). Furthermore, when there is no decent systematic review available, citation bias can have serious consequences that are similar to other questionable research practices (e.g. (58)).

To give some examples, studies included in our review showed that biased exclusion of previous evidence leads to distorted information in the media (44), to incorrect risk perceptions, and to unwarranted decisions such as withholding from treatment in case

of a serious medical condition (33). Also, citation bias has led to research waste because it steered the focus of research into a wrong direction (8, 12). Furthermore, it has been shown that the conclusions of reviews (both narrative and systematic) can be predicted from the choice of which literature was cited in those reviews (59). In other words, if this cited literature is biased, wrong conclusions can be drawn.

Our review has a few limitations. One limitation is the large heterogeneity of our meta-analyses. This is due to the large variety of studies included. We have performed several sensitivity analyses but could not identify the source of this heterogeneity. We therefore performed random-effects meta-analyses to take the heterogeneity into account. Nevertheless, we have to be prudent in drawing a generalized conclusion about the magnitude of citation bias across the sciences.

Further, we used rate ratios in order to pool effects of the included studies. The use of citation rates assumes a linear effect over time and this is unlikely to be the case. In fact, citation generally follows an inverted U-shape with the maximum number of citations often accumulated a couple of years after publication (e.g. (60)). Also, the citation time over which citations have been gathered often varies between the studies that are included. But *within* the majority of these studies the positive and negative rates are based on the exact same publication time, yielding rate *ratios* that can in principle safely be pooled. However, the pooling of rate ratios also assumes a normal distribution, and this assumption is unlikely to be met. Most articles generate just a moderate number of citations while some seminal articles are cited in abundance. This may have lead to overdispersion and an underestimation of our standard errors and confidence intervals.

Finally, this review has focused on the association between article results and citation, but it has not controlled for potential confounders. It is theoretically possible that positive articles are of a higher quality. If this is the case, then research quality may be the actual determinant of citation frequency rather than research outcome. This would imply that high-quality articles would receive more attention, and this could in fact be beneficial for the scientific process.

However, our analysis has shown that quality was not related to the number of citations (Table 2). This is consistent with previous research that showed no association of citation frequency with research quality (e.g. (61-64)), although there is some evidence for an association with research design (65) which is related to research quality. Only journal impact factor showed a consistent effect on citation. However, we believe this factor to mediate the effect of results on citation (e.g. (15)). It is more likely to publish an article in a high impact journal if its results are positive, and this may be part of the explanation why high impact factor journals and articles receive more citations. All in all, it seems improbable that the impact of the article results on the number of citations, as established in this review, can be explained by other factors.

Conclusion

The negative consequences of citation bias can be similar to those of other questionable research practices like publication bias. They may occur with the best of intentions and their individual effects may be small, but all together they lead to an over-representation of positive findings in the scientific literature. This hampers the scientific process, leads to wrong conclusions and decisions, and will eventually harm the reputation of science.

Citation bias could be avoided by citing only systematic reviews but these are not always available or suitable. Alternatively, we could cite all the relevant literature on a topic but this is not realistic. In fact, even in our systematic review, which presents an exhaustive overview of the literature on citation bias, we may have indulged in selective citation ourselves when it comes to side topics. We have used some references to back up an argument, and we did so to the best of our knowledge but without systematically checking the available literature on each of these side topics. By preceding these ad hoc references with 'e.g.', we aimed to clarify that they are merely an example of all the available literature. Likewise, journals could adopt the policy to include a statement on the representativeness of the cited literature, similar to statements on funding and author contributions. Such statement could increase the awareness for selective citation, and an increased awareness could reduce its potential harm.

References

1. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010;14(8).
2. Steneck NH. Fostering integrity in research: Definitions, current knowledge, and future directions. Science and Engineering Ethics. 2006;12(1):53-74.
3. Fanelli D. How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. PLoS One. 2009;4(5):e5738.
4. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Research Integrity and Peer Review. 2016;1(1):17.
5. Duyx B, Urlings MJE, Swaen G, Bouter LM, Zeegers MP. Protocol for Systematic Review of Citation Bias 2015 [Available from: Dataverse http://hdl.handle.net/10411/20710.]
6. Leimu R, Koricheva J. What determines the citation frequency of ecological papers? Trends Ecol Evol. 2005;20(1):28-32.
7. Hutchison BG, Oxman AD, Lloyd S. Comprehensiveness and bias in reporting clinical-trials - study of reviews of pneumococcal vaccine effectiveness. Can Fam Physician. 1995;41:1356-60.
8. Andrade NS, Flynn JP, Bartanusz V. Twenty-year perspective of randomized controlled trials for surgery of chronic nonspecific low back pain: Citation bias and tangential knowledge. Spine J. 2013;13(11):1698-704.
9. Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA. 2002;287(21):2847-50.
10. Conen D, Torres J, Ridker PM. Differential citation rates of major cardiovascular clinical trials according to source of funding - A survey from 2000 to 2005. Circulation. 2008;118(13):1321-7.
11. Etter JF, Stapleton J. Citations to trials of nicotine replacement therapy were biased toward positive results and high-impact-factor journals. J Clin Epidemiol. 2009;62(8):831-7.
12. Greenberg SA. How citation distortions create unfounded authority: Analysis of a citation network. BMJ. 2009;339.
13. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. JAMA. 2005;294(2):218-28.
14. Ioannidis JPA, Panagiotou OA. Comparison of Effect Sizes Associated With Biomarkers Reported in Highly Cited Individual Articles and in Subsequent Meta-analyses. JAMA. 2011;305(21):2200-10.
15. Jannot AS, Agoritsas T, Gayet-Ageron A, Perneger TV. Citation bias favoring statistically significant studies was present in medical research. J Clin Epidemiol. 2013;66(3):296-301.
16. Kivimäki M, Batty GD, Kawachi I, Virtanen M, Singh-Manoux A, Brunner EJ. Don't let the truth get in the way of a good story: An illustration of citation bias in epidemiologic research. Am J Epidemiol. 2014;180(4):446-8.
17. Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. J Clin Epidemiol. 2002;55(4):407-10.
18. Nieminen P, Rucker G, Miettunen J, Carpenter J, Schumacher M. Statistically significant papers in psychiatry were cited more often than others. J Clin Epidemiol. 2007;60(9):939-46.
19. Perino AC, Hoang DD, Holmes TH, Santangeli P, Heidenreich PA, Perez MV, et al. Association between success rate and citation count of studies of radiofrequency catheter ablation for atrial fibrillation possible evidence of citation bias. Circ Cardiovasc Qual Outcomes. 2014;7(5):687-92.
20. Ravnskov U. Cholesterol lowering trials in coronary heart-disease - frequency of citation and outcome. BMJ. 1992;305(6844):15-9.
21. Schrag M, Mueller C, Oyoyo U, Smith MA, Kirsch WM. Iron, zinc and copper in the Alzheimer's disease brain: A quantitative meta-analysis. Some insight on the influence of citation bias on scientific opinion. Prog Neurobiol. 2011;94(3):296-306.

22. Sheth U, Simunovic N, Tornetta P, Einhorn TA, Bhandari M. Poor citation of prior evidence in hip fracture trials. J Bone Joint Surg Am. 2011;93A(22):2079-86.
23. Valachis A, Mauri D, Neophytou C, Polyzos NP, Tsali L, Garras A, et al. Translational medicine and reliability of single-nucleotide polymorphism studies: Can we believe in SNP reports or not? Int J Med Sci. 2011;8(6):492-500.
24. Voracek M. No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D:4D): A comprehensive meta-analysis and critical evaluation of research. Evol Hum Behav. 2014;35(5):430-7.
25. Zhao J, Schooling CM. Promotion of "Low T" and citation bias in testosterone studies. Int J Cardiol. 2015;184:510-1.
26. Kostoff RN. The difference between highly and poorly cited medical articles in the journal Lancet. Scientometrics. 2007;72(3):513-20.
27. Bastiaansen JA, de Vries YA, Munafo MR. Citation Distortions in the Literature on the Serotonin-Transporter-Linked Polymorphic Region and Amygdala Activation. Biol Psychiatry. 2015;78(8):E35-E6.
28. Useem J, Brennan A, LaValley M, Vickery M, Ameli O, Reinen N, et al. Systematic Differences between Cochrane and Non-Cochrane Meta-Analyses on the Same Topic: A Matched Pair Analysis. PLoS ONE. 2015;10(12):17.
29. Kulkarni AV, Busse JW, Shams I. Characteristics associated with citation rate of the medical literature. PLoS ONE. 2007;2(5).
30. Peritz BC. On the heuristic value of scientific publications and their design - a citation analysis of some clinical-trials. Scientometrics. 1994;30(1):175-86.
31. Allareddy V, Shah A, Lin C-Y, Elangovan S, Lee MK, Chung K, et al. Randomized trials published in the journal of dental research are cited more often compared with those in other top-tier non-specialty-specific dental journals. J Evid Based Dent Pract. 2010;10(2).
32. Annalingam A, Damayanthi H, Jayawardena R, Ranasinghe P. Determinants of the citation rate of medical research publications from a developing country. SpringerPlus. 2014;3.
33. Koren G, Nickel C. Perpetuating fears: bias against the null hypothesis in fetal safety of drugs as expressed in scientific citations. J Popul Ther Clin Pharmacol. 2011;18(1).
34. Zambon M, Biondi-Zoccai G, Bignami E, Ruggeri L, Zangrillo A, Landoni G. A comprehensive appraisal of meta-analyses focusing on nonsurgical treatments aimed at decreasing perioperative mortality or major cardiac complications. J Anesth. 2012;26(4).
35. Sawin VI, Robinson KA. Biased and inadequate citation of prior research in reports of cardiovascular trials is a continuing source of waste in research. J Clin Epidemiol. 2016;69.
36. Etwel F, Koren G. When positive studies of novel therapies are subsequently nullified: cumulative meta-analyses in preeclampsia. Clinical and Investigative Medicine. 2015;38(5):E274-E84.
37. Nieri M, Clauser C, Franceschi D, Pagliaro U, Saletta D, Pini-Prato G. Randomized clinical trials in implant therapy: relationships among methodological, statistical, clinical, paratextual features and number of citations. Clin Oral Implants Res. 2007;18(4):419-31.
38. Peritz BC. The use of meta-analyses for citation studies: an illustrative example. Infometrics. 1992:358-76.
39. Misemer BS, Platts-Mills TF, Jones CW. Citation bias favoring positive clinical trials of thrombolytics for acute ischemic stroke: a cross-sectional analysis. Trials. 2016;17.
40. Trinquart L, Johns DM, Galea S. Why do we think we know what we know? A metaknowledge analysis of the salt controversy. Int J Epidemiol. 2016;45(1):251-60.
41. de Vries YA, Roest AM, Franzen M, Munafo MR, Bastiaansen JA. Citation bias and selective focus on positive findings in the literature on the serotonin transporter gene (5-HTTLPR), life stress and depression. Psychol Med. 2016;46(14):2971-9.
42. Unger JM, Barlow WE, Ramsey SD, LeBlanc M, Blanke CD, Hershman DL. The Scientific Impact of Positive and Negative Phase 3 Cancer Clinical Trials. JAMA oncology. 2016;2(7):875-81.

43. Aguinis H, Dalton DR, Bosco FA, Pierce CA, Dalton CM. Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. J Manage. 2011;37(1):5-38.
44. Chapman S, Ragg M, McGeechan K. Citation bias in reported smoking prevalence in people with schizophrenia. Aust NZ J Psychiat. 2009;43(3):277-82.
45. Eisend M, Tarrahi F. Meta-analysis selection bias in marketing research. Int J Res Mark. 2014;31(3):317-26.
46. Robins RW, Craik KH. Is there a citation bias in the judgment and decision literature. Organ Behav Hum Decis Process. 1993;54(2):225-44.
47. Christensen-Szalanski JJ, Beach LR. The citation bias: Fad and fashion in the judgment and decision literature. AmP. 1984;39(1):75-8.
48. Tanner-Smith EE, Polanin JR. A retrospective analysis of dissemination biases in the brief alcohol intervention literature. Psychol Addict Behav. 2015;29(1):49-62.
49. Schumm WR. Evidence of pro-homosexual bias in social science: Citation rates and research on lesbian parenting. Psychol Rep. 2010;106(2):374-80.
50. Barto EK, Rillig MC. Dissemination biases in ecology: Effect sizes matter more than quality. Oikos. 2012;121(2):228-35.
51. Lortie CJ, Aarssen LW, Budden AE, Leimu R. Do citations and impact factors relate to the real numbers in publications? A case study of citation rates, impact, and effect sizes in ecology and evolutionary biology. Scientometrics. 2013;94(2):675-82.
52. Fanelli D. Positive results receive more citations, but only in some disciplines. Scientometrics. 2013;94(2):701-9.
53. Kim MA, Lee HS. Duo-Trio Difference-Preference Test with Two Replications: Use of Psychological Biases for Measuring Meaningful Preference. J Sens Stud. 2015;30(3):211-24.
54. Kumar GK, Kumar KK, Baumgarten G, Ramkumar G. Validation of MERRA reanalysis upper-level winds over low latitudes with independent rocket sounding data. Journal of Atmospheric and Solar-Terrestrial Physics. 2015;123:48-54.
55. Kunz M, Yu Z, Frangakis AS. M-free: Mask-independent scoring of the reference bias. J Struct Biol. 2015;192(2):307-11.
56. Greenberg SA. Understanding belief using citation networks. J Eval Clin Pract. 2011;17(2):389-93.
57. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. The Milbank Quarterly. 2016;94(3):485-514.
58. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. JAMA. 1990;263(10):1392-5.
59. Zhou X, Wang Y, Tsafnat G, Coiera E, Bourgeois FT, Dunn AG. Citations alone were enough to predict favorable conclusions in reviews of neuraminidase inhibitors. J Clin Epidemiol. 2015;68(1):87-93.
60. Yu G, Li YJ. Parameter identification of the observed citation distribution. Scientometrics. 2007;71(2):339-48.
61. Akcan D, Axelsson S, Bergh C, Davidson T, Rosen M. Methodological quality in clinical trials and bibliometric indicators: no evidence of correlations. Scientometrics. 2013;96(1):297-303.
62. Nieminen P, Carpenter J, Rucker G, Schumacher M. The relationship between quality of research and citation frequency. BMC Med Res Methodol. 2006;6:42.
63. West R, McIlwaine A. What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality. Addiction. 2002;97(5):501-4.
64. Giuffrida MA, Brown DC. Association between Article Citation Rate and Level of Evidence in the Companion Animal Literature. J Vet Intern Med. 2012;26(2):252-8.
65. Bhandari M, Busse J, Devereaux PJ, Montori VM, Swiontkowski M, Tornetta P, et al. Factors associated with citation rates in the orthopedic literature. Can J Surg. 2007;50(2):119-23.

4

Citation bias in the literature on dietary trans fatty acids and serum cholesterol

Miriam J.E. Urlings, Bram Duyx, Gerard M.H. Swaen, Lex M. Bouter, Maurice P.A. Zeegers

This chapter has been published in *Journal of Clinical Epidemiology*:

Urlings MJ, Duyx B, Swaen GM, Bouter LM & Zeegers MP. Citation bias in the literature on dietary trans fatty acids and serum cholesterol. Journal of clinical epidemiology, 2019; 106, 88-97.

Abstract

Objective: Balanced citations are a necessary condition for a sound development of scientific knowledge, whereas selective citations may bias scientific consensus. In this study, we assess which determinants influenced the likelihood of being cited in the literature on trans fatty acids and cholesterol.

Study design: We conducted a citation network analysis of the literature concerning trans fats and LDL- and HDL-cholesterol. Each publication was scored on various potential determinants of citation, such as study outcome, study design, sample size, journal impact factor and funding source. We applied random effect logistic regression to identify determinants of citation.

Results: A network of 108 publications was identified, containing 5041 potential citation paths and 669 utilized citation paths. Reporting statistically significant results was found to be a strong predictor of citation, together with sample size, journal impact factor and the authority of the authors.

Conclusion: Within the literature on trans fat intake and cholesterol, selective citations are based on several grounds. Especially the effect of reporting significant results on citation requires special attention, since disproportionate attention is paid to publications suggesting a harmful effect of trans fat on cholesterol.

Introduction

Citations are a key element in the development of knowledge and are highly valued within the scientific community. Citations have been suggested as a measure of (intellectual) influence (1, 2), persuasiveness (3, 4) and socially defined quality (5, 6). With the growing amount of literature in biomedical research, it is impossible to cite every relevant source and therefore only a selection of the relevant literature is cited in each publication. It might be questioned whether citation practices are solely driven by research quality, or whether factors such as the study results also influence selection of citations. Previous research, on clinical trials from a variety of fields, showed that most trials only cite a very small percentage of available previous trials and thereby neglect evidence (7, 8). Song (2010) defined the phenomenon when the selection of citations is related to the study result, as citation bias (9). Citation bias is considered to be a Questionable Research Practice (QRP). QRPs are a variety of scientific misbehaviors that negatively influence the development of knowledge (10). A recent survey among scientists showed that selective citation is one of the most frequent QRPs (11).

The occurrence of citation bias has been studied in a number of research areas. Recently, a systematic review and meta-analysis summarized the literature on citation bias (12). 47 Publications on citation bias were identified from different research areas and using different methods. Overall, our systematic review concluded that citation bias is present, with the most convincing evidence from the field of biomedical sciences (12). A meta-analysis in our review showed that positive studies are approximately two times more likely to be cited than negative studies. However, heterogeneity was very high due to the different methodologies and topics under study. Some publications studied citation bias in a specific research area (13-16), whereas others studied citation bias within one or more journals or databases (17-19) or assessed citation bias in systematic reviews (20-22). Additionally, the way in which citation bias was measured varied across publications. For example, Greenberg (2009) compared the percentage of positive citations to the percentage of negative citations in a certain field (23). Ioannidis (2011) followed a different approach, by taking a cohort of publications that received more than 400 citations and comparing their reported effect sizes with the effect sizes in meta-analyses in the same field (24). Both approaches assess the relationship between study outcome and citation, but each from a different perspective.

Citation bias focuses on study outcome as the determinant of selective citation. However, selective citation can be driven by other factors as well. Determinants that have shown a positive association with citation rate in multiple studies are: sample size, study design, journal impact factor and the number of references (19, 25, 26). Furthermore, privately funded studies are often believed to be less credible compared to publicly funded research (27). Research has indicated that for-profit studies receive higher numbers of citations and thereby potentially skew knowledge development (28). Other factors that have been incidentally linked to citation count, but with uncertain results,

are gender of the author, number and type of affiliations involved in a publication, the authors' reputation and whether the title of the publication includes its conclusion or not (25, 26). Finally, the language of a publication might influence the likelihood of citation, e.g. by formulating firm conclusions or not. This can be measured as 'hedging'. Hedging refers to the use of vague language and therefore attenuates the strength of a claim (29). Via specialized software (29), each publication can be given a hedging score between 0 to 5, with a higher value corresponding to the use of more uncertain language. To which extent these potential determinants of selective citation actually influence knowledge development is unclear and will most likely be subtle.

In the current study, we apply a new methodology to study the occurrence and determinants of selective citation. Inspired by previous literature, we have combined several approaches to assess the impact of different determinants on the likelihood of being cited. In this citation network analysis we are not interested in the content or the correctness of citations, but solely in the occurrence of selective citations in the light of skewed knowledge development. We aim to add to previous literature by providing a broader overview of selective citation by including all previously mentioned determinants into one study.

As an example topic, we will apply our methodology to the literature on the effect of industrially produced trans fatty acid (IP-TFA) intake on LDL cholesterol (LDL-c) and HDL cholesterol (HDL-c). IP-TFA is known to be associated with increased serum LDL-c and decrease serum HDL-c (30-33). This scientific consensus has been strong enough for policy makers and industry to take action and limit the amount of IP-TFA in food (34, 35). The reason for choosing this research area as an example is that no obvious signs of citation bias are present in this field. Therefore, we can study the occurrence of citation bias and the determinants of selective citation in an objective way, without expectations with regard to the outcome. Additionally, the fact that the discussion has been settled makes it interesting to see how selective citation might have influenced the development of consensus in this field.

Methods

Prior to performing the study, the citation network analysis method was described in a study protocol (see supplemental material). The main activities in the citation network analysis are also presented in this paragraph.

Search strategy and article selection

A systematic search strategy was performed in Web of Science – Core Collections. The following search terms were used to identify relevant studies on IP-TFA, LDL-c and HDL-c: ("Trans fat*" OR "Hydrogenated oil*" OR "Elaidic acid*") AND ("Low density lipopro-

tein" OR "High density lipoprotein" OR "Cardiovascular" OR "Coronary heart disease" OR "LDL" OR "HDL" OR "CVD"). The search strategy was based on broad search terms, to make sure no relevant publications would be missed. Reference lists of the identified articles were not checked for missing publications, since this would interfere with the study aim. This would have caused an overrepresentation of articles cited within the network, whereas articles neglected by the network would less likely be found.

Publications were included if one of their study aims concerned the effect of IP-TFA intake on LDL-c and/or HDL-c and if the study design was an observational or an intervention study, a systematic review or another type of synthesis article. Only studies on human subjects were included, no restrictions with regard to language or publication year were used.

The search, conducted by MJEU in January 2016, identified 1027 publications (Figure 1). After the title selection, performed in duplo by MJEU and BD, 332 publications remained. A second selection round, based on abstract, performed by MJEU and GMHS, limited the number of publications to 118. Ultimately, the network contained 108 publications, since the full text of ten publications could not be identified, despite multiple attempts to find them via Google scholar, the corresponding author and ResearchGate.

Figure 1. Flow diagram of the article selection procedure

Data extraction

All publications in the network were scored on a number of potential determinants of citation, listed in Table 1. Data extraction was performed in duplo by MJEU and GMHS.

Study outcome was operationalized in two ways. First, we scored whether the publication reported a statistically significant effect of trans fat intake on LDL-c, HDL-c or the ratio of total cholesterol:HDL-cholesterol (TC:HDL-c ratio). The latter was not described in our study protocol but added as an outcome measure, because this appeared to be an important and often used biomarker of cardiovascular disease during the 1990's. Second, study outcome was measured by scoring whether or not the findings of the publication were in line with the hypotheses that IP-TFA intake increases LDL-c, decreases HDL-c and increases the TC:HDL-c ratio. When the findings were in line with the above hypotheses, scoring was based on the reported point estimate, without taking into account the confidence interval. For publications that did not report statistical results, such as narrative reviews, scoring was based on the author's conclusion. The study designs presented in this network were observational studies (cohort and cross-sectional studies), intervention studies, systematic reviews, narrative reviews and editorials. Sample size was measured as the total number of participants included in a study. For systematic reviews the sample size was measured as the sum of participants of all included studies. Narrative reviews or other documents where the number of participants was not clearly described were coded as having zero participants, in order to avoid missing values in the analysis. The determinant 'authority of the corresponding author' was measured by the publication's level and varies over time. All co-authors of all publications were assigned an 'authority score', which was the number of citations received within this IP-TFA network, at each year the network was active. The authority of the author was measured as the total number of citations received at the moment before a new potential citation, to avoid interference between the authority and the citation score.

Other variables that were measured are: hedge factor, number of affiliations, journal impact factor at the moment of publication, funding source, number of references, whether or not the title of the publication described the conclusion of the publication, as well as the gender and the affiliation of the corresponding author.

Statistical analysis

Each publication in the network plays the role of a citing and a cited publication. We were solely interested in the effect of the characteristics of the cited publication on the likelihood of being cited. The unit of analysis in this study was therefore the citation path itself. A potential citation path existed between one publication and every other publication in the network that was published at least two years earlier. This limit of two years was set to allow the citing publication sufficient time to be written, reviewed and published. In the constructed data set, each row represented a potential citation path followed by an indication whether the potential citation path had actually been utilized or not and the characteristics of the cited publication of that citation path.

A single publication could cite to multiple other publications. Therefore, a multilevel approach was required, in which the citation paths were nested under the citing publication. Random effect logistic regression was modeled to assess the effect of characteristics of the cited article on the likelihood of being cited. Fixed effect logistic regression analysis was used to test concordance between the cited and citing publication in relation to citation behavior.

First, univariate analyses were performed to test every previously described potential determinant of citation in the cited publication as a predictor of the likelihood of being cited. Second, each analysis was adjusted for study design and sample size. These two variables were used as a proxy for study quality, which we considered to be an acceptable driver of selective citation. All statistical analyses were performed in Stata 13.

The outcomes of the logistic regression are reported in the results section of this publication as odds ratios. According to the literature, the odds ratio gives an overestimation of the risk compared to the relative risk in studies where the outcome is a common condition (36). In our network, the prevalence of being cited is 13% (669 actual citations of 5041 potential citations). With this prevalence, we do not consider 'being cited' as a very common condition and hence the overestimation of the odds ratio over the relative risk will be within considerable limits (36). Ultimately, the odds ratio gives an accurate estimation of the direction of the effect; only the exact magnitude of the effect should be interpreted with some caution. For the readability of the publication, we interpret these values as if they are relative risks and therefore speak about 'the likelihood of being cited'.

Results

In total 108 publications, published between 1990 and 2015, on the effect of IP-TFA on cholesterol were identified in the network. Among these publications, 5041 potential citation paths existed; 669 of them have been utilized. Characteristics of the network are displayed in Table 1. Table 2 displays the crude odds ratios for all potential determinants of the likelihood of being cited. Figure 2 visualizes of part of the network, namely the empirical studies (circles) and systematic reviews (squares). From the visualization, the development of the amount of evidence over time can be seen together with the timing of the systematic reviews. Additionally, the visualization shows that some publications are highly cited within the network (e.g. Mensink, 1990), whereas others are less popular (e.g. Mutalib, 1999).

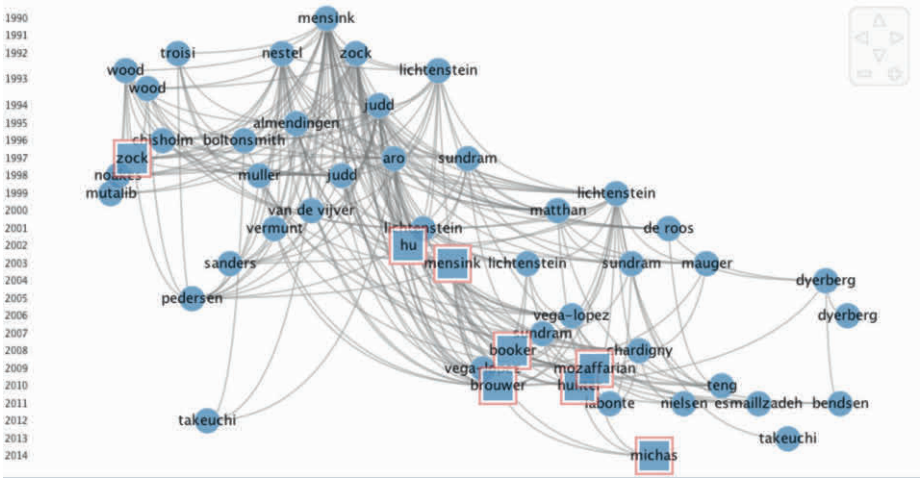


Figure 2. Visualization of the empirical studies and systematic reviews in the network*

* Circles reflect the empirical studies, squares reflect systematic reviews. Each line depicts a performed citation. The y-axis shows the time line; the x-axis is only for visualization purposes.

Table 1. Distribution of article characteristics and potential determinants of citation in the literature on trans fats and cholesterol of 108 articles and 5041 citation relations*

		N Publications	N Citation paths utilized (%)	N Citation paths not utilized (%)
Study outcome				
Statistical significance LDL-c	Significant increase LDL-c	26	431 (29)	1036 (71)
	No signficant increase LDL-c	16	118 (13)	811(87)
Statistical significance HDL-c	Significant decrease HDL-c	21	308 (27)	852 (73)
	No significant descrease HDL-c	22	243 (19)	1053 (81)
Statistical significance ratio TC:HDL-c	Significant increase TC:HDL-c	12	123 (25)	378 (75)
	No significant increase TC:HDL-c	6	61 (18)	279 (82)
Hypothesis LDL-c	In line with hypothesis – increase LDL-c	86	513 (14)	3209 (86)
	Not in line with hypothesis	16	147 (15)	862 (85)
Hypothesis HDL-c	In line with hypothesis – decrease DHL-c	86	498 (13)	3311 (87)
	Not in line with hypothesis	9	90 (16)	457 (84)
Hypothesis TC:HDL-c	In line with hypothesis – increase TC:HDL-c	38	211 (16)	1136 (84)
	Not in line with hypothesis	2	11 (14)	65 (86)
Article characteristics – content related				
Study design	Observational study	6	18 (7)	236 (93)
	Experimental study	36	475 (23)	1584 (77)
	Systematic review	9	49 (21)	183 (79)
	Narrative review	54	122 (5)	2329 (95)
	Other	3	5 (11)	40 (89)
Sample size ***	10-40	22	280 (20)	1115 (80)

		N Publications	N Citation paths utilized (%)	N Citation paths not utilized (%)
	40-80	13	190 (31)	419 (69)
	>80	15	91 (13)	586 (87)
Number of extra determinants***	0	34	120 (7)	1519 (93)
	1	21	83 (13)	534 (87)
	>1	52	466 (17)	2319 (83)
Number of extra outcomes ***	0-3	43	222 (9)	2186 (91)
	4-6	29	267 (20)	1102 (80)
	>6	36	180 (14)	1084 (86)
Article characteristics – not content related				
Number of affiliations***	1	36	207 (11)	1698 (89)
	2	28	153 (13)	998 (87)
	>2	43	309 (15)	1685 (85)
Journal Impact Factor ***	0-2	31	34 (2)	1439 (98)
	2-4	45	311 (14)	1886 (86)
	>4	31	322 (24)	1020 (76)
Funding source	Not-for-profit	39	248 (12)	1730 (88)
	For-profit	13	133 (19)	373 (81)
	Both	9	113 (23)	373 (77)
	Not reported	45	175 (9)	1699 (91)
Number of references***	< 30	22	206 (17)	992 (83)
	30-50	43	276 (14)	1674 (86)
	> 50	42	187 (10)	1706 (90)
Title of publication	Not suggesting a conclusion	91	541 (13)	3734 (87)
	Suggesting a conclusion	17	128 (17)	638 (83)
Language	English	104	669 (13)	4291 (87)
	Other	4	0 (0)	81 (100)
Author characteristics				
Gender	Male	74	514 (15)	2983 (85)
	Female	33	155 (10)	1325 (90)
Affiliation corresponding author	University	87	508 (12)	3712 (88)
	Government	7	87 (27)	232 (73)
	Industry	7	26(13)	168 (87)
	Hospital	3	1 (4)	22 (96)
	Other	4	4 (2)	181 (98)
Authority***	<10	NA**	128 (7)	1610 (93)
	11-60	NA**	254 (17)	1281 (83)
	> 60	NA **	287 (16)	1481 (84)

* Missing values, due to unclear findings or lack of reporting, are not reported in this table, ** Values are not measured on publications level, but vary over time, *** Variables are measured on continuous scale and decoded into categories based on their tertiles

Study outcome

In this network, each narrative review was in line with the hypotheses that IP-TFA increases LDL-c, decreases HDL-c, and increases TC:HDL-c ratio (data not shown).

Reporting a significant effect of trans fat intake on LDL-c, regardless of the direction of the effect, is a strong predictor of citation and leads to an approximately three times higher likelihood of being cited. After adjusting for study design and sample size this effect remained, suggesting that the effect cannot be explained by study quality. For the relation between significant findings on HDL-c and the ratio TC:HDL-c and citation, this effect is also present but of a lower magnitude of approximately 1.5 and 2 respectively. Taking into account the direction of the effect, being in line with the hypothesis on LDL-c and HDL-c or not, is a significant determinant of citation only after adjusting for study design and sample size (Table 2). Reporting results that are in line with the hypothesis that TC:HDL-c is increased by IP-TFA intake or not, has no impact on the likelihood of being cited. Since the variables "significance" and "being in line with hypothesis" are related to each other, a Bonferroni adjustment was performed to correct for multiple comparisons. These two variables are tested in six hypotheses, regarding LDL-c, HDL-c and TC:HDL-c. After applying a Bonferroni correction, an α of 0.008 was taken into account. This did not change the interpretation of the results.

Article characteristics - content related

Large variation was observed in the sample sizes of the publications, ranging from 10 to 140,390 participants. To deal with the large range and non-parametric distribution, sample size was used as a categorical variable in the analysis (Table 1). Naturally, sample size highly correlates with study design, as the lowest category of sample size (0 participants) mainly contains narrative reviews, whereas observational studies only appear in the highest category (more than 80 participants). Nevertheless, there was no proof of collinearity between sample size and study design and the likelihood of being cited increased with a higher number of participants (Table 2).

Although narrative reviews made up half of the network, they get cited rarely (5%) compared to every other research design (Table 1). The likelihood of being cited for intervention studies and systematic reviews is respectively 6 and 7 times higher compared to narrative reviews (Table 2). However, the effect of experimental study design on citation is explained completely by the sample size, as the odds ratio changed from 6.0 to 0.84 after correcting for sample size (Table 2).

Many publications included additional determinants and outcomes besides IP-TFA and LDL-c and HDL-c (Table 1). The likelihood of being cited increased with both additional determinants and study outcomes, shown by a p-trend of <0.001 for the crude ORs of both variables. This can be explained because less specific publications can be cited for multiple reasons, by different research areas. The effect of additional out-

comes and determinants on the likelihood of being cited disappeared and even reversed, after adjusting for study design and sample size respectively.

Article characteristics – not content related

Hedge factors were measured for all publications written in the traditional format (i.e. having separate sections for introduction, method, results and discussion), which were 42 publications in total. Values of the hedge factor ranged from 0.57 to 3.26, on a 5-point scale, with a median value of 1.35. The likelihood of being cited increased significantly by almost three times with an increasing hedge factor (crude OR: 2.92 95%CI: 2.1-4.0). As explained before, a higher hedge factor referred to more uncertain statements.

The journal impact factor (JIF) at the moment of publication of an article ranged from 0 to 53 and followed a non-parametric distribution. Therefore, JIF was categorized based on its tertiles (Table 1). Having an impact factor higher than 4 increased the likelihood of being cited by approximately 18 times, compared to an impact factor lower than 2. Because the JIF was measured at the moment of publication, the JIF did not interfere with the dependent variable of citation.

Having a title stating the conclusion significantly raised the likelihood of being cited (OR: 1.46, 95% CI: 1.2-1.8). This effect reduced to a non-significant OR of 1.06 after adjusting for sample size and study design. Regarding language, only four publications were identified that were written in a language other than English and none of them received any citations in this network (Table 1). Therefore, language could not be included in the analysis as a potential determinant of citation. Looking at funding of the study, which the publication reports, a large part of the network (45 publications) did not report any funding source. Nevertheless, no significant effect of reporting funding on the likelihood of being cited was found (Table 2). When comparing different types of funding, for-profit funding had a 1.64 times higher likelihood of being cited compared to not-for-profit funding in this network (Table 2).

The number of references in a publication is inversely related to the likelihood of being cited. This contradicts our hypothesis that a higher number of references would lead to more credibility and therefore more citations. The number of references is often limited by the journal, and this rule might be different for high impact factor journals than low impact factor journals. However, the Pearson's correlation between the JIF and number of citations is only 0.0296.

Author characteristics

The majority of publications in this network had a male corresponding author (Table 1). The corresponding authors of 87 publications were affiliated to a university. The likelihood of being cited is significantly smaller for female corresponding authors compared to male corresponding authors (OR: 0.72; 95%CI: 0.6-0.9), whereas the affiliation of the corresponding author did not make a difference (OR: 1.37; 95%CI: 0.9-2.1). After cor-

recting for study design and sample size, both gender (OR: 1.19; 95%CI: 0.9-1.6) and affiliation (OR: 0.83; 95%CI: 0.5-1.3) did not significantly influence the likelihood of being cited anymore. The authority of the authors, which is determined by the total number of citations received each year in the network, ranged from 0 to 255, and followed a skewed distribution (median: 30; IQR: 3-93). Based on the tertiles, the variable was recoded into a categorical variable with three categories (<10, 11-60, >60). As expected, the likelihood of being cited increased three to four fold with an increase in the authority of the authors. Also after adjusting for study design and sample size, this significant effect of authority on the likelihood of being cited remained and even increased to an OR of 5 (Table 2).

Table 2. Crude and adjusted odds ratios (95% CI) for the likelihood of being cited

	Crude OR	Adjusted OR
Study outcome		
Significance LDL-c *	3.08 (2.4-4.0)	3.15 (2.4-4.2)
Significance HDL-c	1.55 (1.3-1.9)	1.67 (1.3-2.1)
Significance TC:HDL-c*	1.59 (1.1-2.3)	2,21 (1.4-3.5)
Hypothesis LDL-c	1.06 (0.9-1.3)	3.30 (2.6-4.2)
Hypothesis HDL-c	0.86 (0.7-1.1)	2.09 (1.6-2.8)
Hypothesis TC:HDL-c	1.12 (0.6-2.2)	0.35 (0.1-2.4)
Article characteristics – content related		
Study design * **		
Narrative review	1 (ref)	1 (ref)
Observational study	1.15 (0.7-2.0)	0.40 (0.2-0.8)
Experimental design	6.00 (4.8-7.5)	0.84 (0.3-2.4)
Systematic review	7.20 (4.9-10.6)	3.41 (2.1-5.5)
P for trend	<0.001	<0.001
Sample size ** **		
0 participants	1 (ref)	1 (ref)
1-40 participants	5.60 (4.4-7.1)	7.05 (2.5-20.1)
41-80 participants	9.80 (7.5-12.9)	12.34 (4.3-35.5)
>80 participants	3.42 (2.5-4.6)	3.38 (2,2-5.1)
P for trend	0.025	<0.001
Number of determinants		
0 additional determinants	1 (ref)	1 (ref)
1 additional determinant	2.54 (1.9-3.5)	0.38 (0.2-0.6)
>1 additional determinant	2.70 (2.2-3.4)	0.43 (0.3-0.6)
P for trend	<0.001	0.001
Number of outcomes		
0-3 additional outcomes	1 (ref)	1 (ref)
4-6 additional outcomes	2.57 (2.1-3.2)	1.21 (0.9-1.5)
>6 additional outcomes	2.17 (1,7-2.7)	0.92 (0.7-1.2)
P for trend	<0.001	0.701

	Crude OR	Adjusted OR
Article characteristics – not content related		
Hedge factor	2.64 (1.9-3.6)	3.27 (2.1-5.1)
Number of affiliations		
1 affiliation	1 (ref)	1 (ref)
2 affiliations	1.36 (1.1-1.7)	1.41 (1.1-1.8)
>2 affiliations	1.78 (1.5-2.2)	1.12 (0.9-1.4)
P for trend	<0.001	0.254
Journal impact factor		
0-2	1 (ref)	1 (ref)
2-4	7.58 (5.2-11.0)	5.51 (3.6-8.3)
>4	18.00 (12.3-26.4)	10.88 (7.1-16.6)
P for trend	<0.001	<0.001
Funding source		
For-profit vs not-for-profit	1.64 (1.3-2.1)	1.18 (0.9-1.6)
Reported vs not reported funding	0.88 (0.7-1.1)	1.1.0 (0.9-1.4)
Number of references		
<30 references	1 (ref)	1 (ref)
30-50 references	0.87 (0.7-1.1)	0.74 (0.6-0.9)
>50 references	0.60 (0.5-0.8)	0.55 (0.4-0.7)
P for trend	<0.001	<0.001
Title of publication	1.46 (1.2-1.8)	1.06 (0.8-1.4)
Author characteristics		
Gender		
Female vs male	0.72 (0.6-0.9)	1.19 (0.9-1.6)
Affiliation corresponding author		
Public vs private sector	1.37 (0.9-2.1)	0.83 (0.5-1.3)
Authority		
<10	1 (ref)	1 (ref)
11-60	3.15 (2.4-4.1)	2.70 (1.9-3.8)
>60	4.76 (3.6-6.2)	5.06 (3.5-7.4)
P for trend	<0.001	<0.001

* Odds ratio of chance to be cited for significant vs not significant findings, ** model adjusted for study design and sample size, both as categorical variables, *** Adjusted model is adjusted for sample size, **** Adjusted model is adjusted for study design

To assess the impact of the large number of narrative reviews in this network, a sensitivity analysis was performed leaving out the narrative reviews as cited publications. This did not result in a substantial shift in outcomes, compared to the overall crude and adjusted analysis. Results of this analysis are presented in Table S1 in the digital supplement.

Results of the concordance analyses are presented in Table S2 of the digital supplement. It appeared that concordance between the citing and cited publication did not greatly influence the likelihood of citation, for any of the study characteristics.

Discussion

In this network, we found that several factors are determinants for selective citation. First, citation bias was found, since the adjusted analysis showed a significantly higher likelihood of citation for statistically significant studies on the relationship between LDL-c, HDL-c and TC:HDL-c and serum cholesterol. Citation bias can affect the scientific process in several ways. For example, it might drive the reader, consciously or unconsciously, in a certain direction. Especially when this happens systematically, the reader might develop a certain belief system that is not in line with the available evidence. The effect of citation bias is especially dangerous, because it cannot easily be detected within one publication. By systematic overrepresentation of and exposure to positive results, gradually a shift towards an unfounded belief system might be made. Greenberg (2009) showed in his citation analysis that citation bias can amplify certain claims by citing to reviews, which lack data addressing this claim (23). As a kind of mere exposure effect, the reader will accept this claim more easily. The harmful effect of citation bias was shown clearly by Ioannidis (2011). This study showed that highly cited publications show exaggerated effect sizes when compared to effect sizes of meta-analyses in the same field (24). Consequently, these exaggerated effect sizes will be of great influence in the research field because of their high number of citations whereas they do not reflect the true effect.

Citation bias has been studied together with other types of selective reporting, such as outcome reporting bias and publication bias (37). It is seen as a general phenomenon that negative findings receive less attention, which leads to less publication, reporting and citation of negative or non-significant results (37). All types of selective reporting lead to an overrepresentation of certain evidence, making it very difficult to indicate the true effect and even potentially threatening the validity of meta-analyses (38, 39). Although meta-analyses will be primarily influenced by publication bias and outcome reporting bias, also citation bias might play a role in this. Identification of literature via reference checking and contacting experts in the field are fixed components of systematic reviews and meta-analyses. Negative studies that are not often cited might be missed in this process, compared to positive, highly cited publications. Particularly in narrative reviews, citation bias might affect the discussion. Scientific knowledge development and agenda setting might be seriously influenced by over-reporting certain results, while neglecting other evidence. This becomes particularly problematic when scientific consensus is being used as a basis for clinical guidelines, legislation, industry decisions or future research funding.

Additional to citation bias, our study identified other determinants of selective citation. Especially sample size, JIF and the authority of the author were positively correlated with the likelihood of being cited. These findings were in line with previous literature. Onodera (2013) showed that the influence of JIF is greater than author-related factors, such as the number of citations received by each author, with regard to citation rates (26). Callaham (2002) analyzed which factors influenced the citation of publications for which abstracts had been submitted to a scientific meeting (40). In this body of literature, sample size was the biggest predictor of citation, whereas study outcome had no effect on citation. Opposed to previous research, factors such as funding and the title of the publication were not associated with citation in the current network.

The methodology used in this work is different from earlier citation analyses in several ways. Previous studies have often compared the citation *count* towards positive and negative study (e.g. 41). We have however assessed the likelihood of being cited for publications, in relation to a number of determinants, by taking into account all potential citation pathways. Additionally, the statistical method, a multilevel logistic regression, has not been used on previous citation analyses. With this approach we take into account the fact that citations are clustered in publications and therefore, not all citation pathways are independent of each other. Finally, we studied not only the characteristics of the cited publication in relation to the chance of citation, but we also looked at concordance. The rational behind this analysis was that authors use citations to amplify their own findings, and therefore cite mostly to sources that share the same characteristics. This would be different form of selective citation.

Our study also has several limitations. First, the search strategy to map the network was only conducted on the Web of Science database, making it impossible to use MeSH-terms in the search strategy and potentially leaving relevant publications unidentified. Especially the fact that only four non-English publications were identified, makes us suspect that not all relevant publications have been found. The reason for limiting the search to the WoS database was that this is the only database that enabled us to download the publications together with all corresponding citation paths, which was required for developing the data set. Second, the study outcome of each publication, which could be either statistical significance or being in line with the hypothesis, was scored based on the evidence reported in the result section of the publication. This does not necessary correspond with the way the publication has been cited by other publications. Although it is likely that citations are based on the presented evidence, it might also be based on the author's overall conclusion. The third limitation relates to the statistical analysis. Initially, it was planned to perform a multivariate analysis, to adjust for all determinants that might relate to citation frequency. In this way, we aimed to develop a model to predict the chance of being cited. In the prespecified study protocol, it was laid down that the multivariate model would include all determinants that showed a significant effect on the likelihood of being cited in the crude analysis. Unfortunately, analyzing this multivariate model created a very unstable model. As it was

impossible to draw meaningful conclusions from this analysis, we decided to limit our study to the analysis adjusted for study design and samples size, as a proxy for study quality. Also in the adjusted analysis we should be aware that there might be overlap in the explained variation by both study design and sample size, as these variables are closely related. Nevertheless, no evidence for collinearity between these two variables was found.

In conclusion, with this study we have shown a new methodology to assess the occurrence of citation bias, as well as created a broader insight into the influence of other determinants of selective citation. From previous research, citation bias has shown to be a problem in several research fields. Also in the example study on trans fatty acid intake and serum cholesterol, citation bias was found. Sample size, JIF and authority of the author were identified as other important determinants of citation. Further studies are needed to confirm these findings in other research areas and assess the consequences of selective citation for knowledge development and science-based decision-making.

References

1. Merton RK. The Matthew effect in science. Science. 1968;159(3810):56-63.
2. Bornmann L, Daniel H. What do citations measure? A review of studies on citing behavior. Journal of Documentation. 2006;64(1):45-80.
3. Zuckerman H. Citation analysis and the complex problem of intellectual influence. Scientometrics. 1987;12(5-6):329-38.
4. Gilbert GN. Referencing as persuasion. Social Studies of Science. 1977:113-22.
5. Cole J, Cole S. Measuring the Quality of Sociological Research: Problems in the Use of the" Science Citation Index". The American Sociologist. 1971:23-9.
6. Moed HF. Citation analysis in research evaluation: Springer Science & Business Media; 2006.
7. Robinson KA, Dunn AG, Tsafnat G, Glasziou P. Citation networks of related trials are often disconnected: implications for bidirectional citation searches. Journal of clinical epidemiology. 2014;67(7):793-9.
8. Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. Annals of internal medicine. 2011;154(1):50-5.
9. Song F, Parekh S, Hooper L, Loke YK, Ryder J. Dissemination and publication of research findings : an updated review of related biases. Health Technology Assessment. 2010;14(8):234.
10. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PloS one. 2009;4(5):e5738.
11. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Research Integrity and Peer Review. 2016;1(1):17.
12. Duyx B, Urlings MJ, Swaen GM, Bouter LM, Zeegers MP. Scientific Citations Favor Positive Results: A Systematic Review and Meta-analysis. Journal of Clinical Epidemiology. 2017.
13. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. BMJ 1992 ;305.
14. Trinquart L, Johns DM, Galea S. Why do we think we know what we know? A metaknowledge analysis of the salt controversy. International journal of epidemiology. 2016;45(1):251-60.
15. Schumm WR. Evidence of pro-homosexual bias in social science: citation rates and research on lesbian parenting. Psychological reports. 2010;106(2).
16. Andrade NS, Flynn JP, Bartanusz V. Twenty-year perspective of randomized controlled trials for surgery of chronic nonspecific low back pain: citation bias and tangential knowledge. The spine journal : official journal of the North American Spine Society. 2013;13(11).
17. Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. Journal of clinical epidemiology. 2002;55(4).
18. Nieminen P, Rucker G, Miettunen J, Carpenter J, Schumacher M. Statistically significant papers in psychiatry were cited more often than others. Journal of clinical epidemiology. 2007;60(9):939-46.
19. Kostoff R. The difference between highly and poorly cited medical articles in the journal Lancet. Scientometrics. 2007;72(3):513-20.
20. Jannot A-S, Agoritsas T, Gayet-Ageron A, Perneger TV. Citation bias favoring statistically significant studies was present in medical research. Journal of clinical epidemiology. 2013;66(3).
21. Leimu R, Koricheva J. What determines the citation frequency of ecological papers? Trends in ecology & evolution.2005;20(1).
22. Etter J-F, Stapleton J. Citations to trials of nicotine replacement therapy were biased toward positive results and high-impact-factor journals. Journal of clinical epidemiology. 2009;62(8).
23. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. BMJ. 2009 2009;339.
24. Ioannidis JPA, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. Jama. 2011;305(21).

25. Bornmann L, Daniel H-D. What do citation counts measure? A review of studies on citing behavior. Journal of documentation. 2008;64(1):45-80.
26. Onodera N, Yoshikane F. Factors affecting citation rates of research articles. Journal of the Association for Information Science and Technology. 2015;66(4):739-64.
27. Sismondo S. How pharmaceutical industry funding affects trial outcomes: causal structures and responses. Social science & medicine. 2008;66(9):1909-14.
28. Conen D, Torres J, Ridker PM. Differential citation rates of major cardiovascular clinical trials according to source of funding: a survey from 2000 to 2005. Circulation. 2008;118(13):1321-7.
29. Kilicoglu H, Bergler S. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. BMC bioinformatics. 2008;9(11):1.
30. Mensink RP, Katan MB. Effect of dietary trans fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. New England Journal of Medicine. 1990;323(7):439-45.
31. Booker C, Mann J. Trans fatty acids and cardiovascular health: Translation of the evidence base. Nutrition, Metabolism and Cardiovascular Diseases. 2008;18(6):448-56.
32. Hu FB, Willett WC. Optimal diets for prevention of coronary heart disease. Jama. 2002;288(20):2569-78.
33. Hunter JE, Zhang J, Kris-Etherton PM. Cardiovascular disease risk of dietary stearic acid compared with trans, other saturated, and unsaturated fatty acids: a systematic review. The American journal of clinical nutrition. 2010;91(1):46-63.
34. Okie S. New York to trans fats: you're out! New England Journal of Medicine. 2007;356(20):2017-21.
35. Astrup A. The trans fatty acid story in Denmark. Atherosclerosis Supplements. 2006;7(2):43-6.
36. Cummings P. The relative merits of risk ratios and odds ratios. Archives of pediatrics & adolescent medicine. 2009;163(5):438-45.
37. Fanelli D. Positive results receive more citations, but only in some disciplines. Scientometrics. 2013 Feb;94(2):701-9.
38. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. Journal of clinical epidemiology. 2000;53(2):207-16.
39. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS one. 2008;3(8):e3081.
40. Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. Jama. 2002 ;287(21).
41. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. BMJ. 2009;339:b2680.

Selective citation is present in epidemiological studies on phthalates: a citation analysis

Miriam J.E. Urlings, Bram Duyx, Gerard M.H. Swaen, Lex M. Bouter, Maurice P.A. Zeegers

This chapter has been submitted for publication to the journal *Epidemiology*

Abstract

Introduction: Selective citation can lead to skewed knowledge development and biased scientific consensus. If the selection of citations is associated with study outcome this is called citation bias. We will study selective citation in a broader sense, including other factors that can influence the selection of citations, e.g. study design, journal impact factor or the funding source of the publication. As a case study we assess which factors drive citation in the human literature on phthalates, specifically the metabolite mono(2-ethylhexyl) phthalate (MEHP).

Methods: A systematic literature search identified all relevant publications on human health effect of MEHP. Data on potential determinants of selective citation were extracted in duplo. Specialized software was used to create a citation network, including all potential citation pathways. We applied random effect logistic regression to assess whether these determinants influence the likelihood of citation.

Results: 112 Publications on MEHP were identified, with 5684 potential citation pathways of which 551 citations were performed. Reporting of a harmful point estimate, the journal impact factor, authority of the author, a male corresponding author, research performed in North America and self-citation were positively associated with the likelihood of being cited.

Conclusion: In the literature on MEHP, citation is mostly driven by a number of factors that are not related to study outcome. Although the identified determinants do not necessarily lead to bias, it shows selective use of published literature.

Introduction

The frequency of being cited is often used to assess the quality and scientific impact of a publication (1). The rationale behind this is that high quality work will lead to more citations by peer scientists compared to low quality work (2). However, it can be questioned whether it is scientific merit only, that drives the number of citations, or other factors such as study outcome or the number of authors involved. It is known that only a few publications reach high citation counts whereas the majority of publications is cited only a few times at most (3). Consequently, publications are forgotten over time whereas others remain to be cited, leading to potentially skewed knowledge development. Since scientific evidence is used as the basis for clinical and policy decisions, as well as for setting the future research agenda, selective citations can have serious consequences. Authors have different motivations to base the selection of their citations on, which can be either justified or unjustified. Study outcome is an unjustified determinant of citation, leading to citation bias (4). Citation analyses have been performed in a large variety of research fields and with different methods. Citation analysis can be done in different contexts, for example within a demarcated research field (5), within a specific journal (6, 7), or based on the individual publications included in a meta-analysis (8). Also different methods can be distinguished to assess the relationship between study outcome and the chance of being cited. Ioannidis (2011) compared the effect size of 35 highly cited publications with the effect sizes in meta-analyses on the same topics (9), whereas Greenberg (2009) compared the citation count towards positive and negative studies in the same field (5). Both studies found proof for the existence of citation bias. A recent systematic review and meta-analysis showed that citation bias is present in different research areas, with an average of approximately twice the number of citations for positive studies compared to negative studies (4).

Contrary to study outcome, high quality of research is considered a justified determinant of citation. Most other determinants of citation will hold the middle between justified and unjustified determinants, e.g. journal impact factor, funding source or affiliation of the corresponding author. Journal impact factor is calculated on the basis of the total number of received citations by a journal in the previous two years (10, 11). Although this cannot directly be translated to the chance of being cited for single publications, several studies did show a correlation between journal impact factor and the chance of citation (2). Also a correlation between study design and citation was reported, where case reports were least likely to receive high number of citations and meta-analysis were most likely to be highly cited (12). This seems legitimate, comparing the level of evidence prestented in meta-analyses compared to case reports. Industry funded research or industry affiliated researchers have been suggested to have a lower chance of being cited, as a result of lower perceived credibility of a for-profit organisation (13, 14). Furthermore, the gender of the author, number of references in a publication, number and type of affiliations included in a publication, the authors' reputation

and whether the title of the publication includes its conclusion or not have been studied in relation to citation, with varying results for different research fields (2, 13).

As a case study, we will study determinants of citation within the human literature on the harmful health effects of phthalates. Phthalates are used in plastic products, such as food packaging, toys and medical tubes and increase the flexibility of the plastic (15). Phthalates are a group of different parent compounds such as diethyl phthalate (DEP), di-n-butyl phthalate (DBP), di-isonyl phthalate (DINP) and di-2-ethylhexyl phthalate (DEHP) (15). They have been studied in relation of a variety of health outcome, such as human reproduction (16, 17), obesity (18) and ADHD in children (19). Due to the fast breakdown of phthalates in the body, phthalate exposure is mostly measured by means of metabolites in urine or blood. Since each parent compound breaks down into a number of metabolites, there is a great variety of phthalate metabolites that might be studied. In terms of citation analysis all these metabolites can be studied as separate subnetworks. In the current study we will focus on one metabolite: mono(2-ethylhexyl) phthalate (MEHP), which is one of the main metabolites of the parent compound DEHP and that can detected in 75% of the US population (20, 21).

In the current citation analysis we combine all of the previously mentioned potential determinants of citation, to get a broader insight into the determinants of citation, specifically in the literature on phthalates and human health effects. Our main research question is therefore: Which determinants influence the likelihood of being cited in the scientific literature on harmful health effects of MEHP in humans?

Methods

The design of this study was described in a study protocol, which was finalised prior to the data collection and data-analysis and was published online (https://bit.ly/2xhTrj1). The main steps of the citation network analysis will be described in the following paragraphs.

Search strategy and article selection

The network was composed via a systematic search in Web of Science – Core Collections. A broad search strategy was applied to prevent missing important publications, namely “phthalate*” AND “human*”. No limitations with regard to the health outcomes under study were applied. Identification of publications by checking the reference lists was not applied, since this would interfere with the research question. Reference checking would result in an overrepresentation of articles that are cited within the network, whereas articles that have been neglected by the network would still be missed.

The article selection was carried out in two phases. The first selection round was based on publication title, to limit the number of publications. The second selection

round included reviewing abstracts, figures and tables, to finalize the article selection. The article selection was conducted individually by two researchers, MJEU and BD, followed by several consensus meetings.

Data extraction

All publications in the network were scrutinized for a number of characteristics that may be potential determinants of citation (see Table 1). Data extraction was performed independent by MJEU and BD. In all cases consensus was reached. Study outcome was scored in three ways; based on statistical significance, direction of the effect and authors' conclusion. As the general hypothesis we take that MEHP leads to a harmful health effect. A publication was considered statistically significant when, for the primary study outcome, a p-value lower than 0.05 was reported. In case multiple health outcomes were reported with p-values both below and above 0.05, a publication was considered as having a mixed outcome. Odds ratios and relative risks higher than one were considered in line with the hypothesis that MEHP is harmful for human health, whereas point estimates between zero and one indicated absence of a harmful health effect. Finally, the study outcome was measured by studying the authors' conclusion of the publication. This can be either in line with the hypothesis that a harmful health effect exists, which is mentioned as a positive study; or not in line with the hypothesis that a harmful health effect exists, which is mentioned as a negative study. The study designs presented in this network were observational studies (cohort, cross-sectional and case-control studies), systematic reviews and narrative reviews. The journal impact factor at the moment of publication was read from Web of Science.

The determinant 'authority of the corresponding author' was measured on the publication's level and varied over time. All co-authors of all publications received an 'authority score', which was the number of citations received within this MEHP network, during each year that the network was active. The authority of each publication was determined by the co-author with the highest authority score. We hypothesized that authors with a high authority would elevate the credibility of a publication and therefore would lead to a higher likelihood of being cited by future publications. Self-citation was defined as the situation in which at least one author was listed on both the cited and the citing publication.

Statistical analysis

Each publication in the network could take the role of citing and cited publication. First, we were interested in the effect of only the characteristics of the *cited* publication on the likelihood of being cited. Therefore, the unit of analysis was the potential citation path. A potential citation path existed between one publication and every other publication in the network that was available online or on paper at the moment of submission of the citing publication. In the data set, each row represented a potential citation

path followed by an indication whether the potential citation path had actually been realized or not and the characteristics of the cited publication of that citation path.

A single publication cites multiple other publications, meaning that multiple citation pathways are leading to the same publication. These citation pathways are therefore not entirely independent. A multilevel approach was therefore required, in which the citation paths were nested under the citing publication. Random effect logistic regression was modelled to assess the effect of characteristics of the cited article on the likelihood of being cited.

Univariate analyses were performed to examine all potential determinants of citation described in table 1. Second, all analyses were repeated while adjusting for study design, which was considered to be a proxy for study quality.

Additionally, we assessed whether concordance between the characteristics of the cited and citing publication was a determinant of citation. Via fixed effect logistic regression analysis we tested whether concordance between the cited and citing publication determined the likelihood of citation. All statistical analyses were performed in Stata 13.

The outcomes of the logistic regression are reported as odds ratios. The odds ratio may overestimate the true relative risk in studies where the outcome is a common (22). In our network, the overall chance of being cited is 9.6% (551 actual citations of 5684 potential citations). With this incidence, we consider 'being cited' relatively common, but we expect the overestimation of the true relative risk to be still acceptable (22). Ultimately, the odds ratio gives an accurate estimation of the direction of the effect; only the exact magnitude of the effect should be interpreted with some caution. For the sake of readability of the publication, we interpret these values as if they are relative risks and therefore for instance speak about 'the likelihood of being cited for negative studies compared to positive studies'.

Results

The network consisted of 112 publications on human health effects of the metabolite MEHP, published between 2000 and 2018. The network contained 5684 potential citations of which 551 are actually performed citations, making the citation prevalence 9.6%. 37 publications in the network did not receive any citations. Six publications received more than twenty citations each, with a maximum of 33 citations. Four of these six highly cited publications were cross-sectional studies, all reporting non-significant findings with an effect size in line with the hypothesis that a harmful effect on health exists (23-26). The most cited publication was a cohort study, also reporting non-significant results with a point estimate in line with the hypothesis (27). Finally, a narrative review on human health and with an unclear conclusion on the health effect of MEHP received 27 citations within the network (28).

The number of substances that were studied additional to MEHP in one publication varied from one to 61, with a median of 8 substances. These substances were other phthalate metabolites, and other endocrine disruptors such as bisphenol A and various heavy metals. The total number of health outcomes per publication ranged from one to 22, with a median of three. The majority of health outcomes were related reproductive outcomes for men and women, such as semen quality, time to pregnancy and a number of hormones. Furthermore, MEHP was studied in relation to metabolic outcomes (e.g. diabetes type 2, cancer, cardiovascular disease) and general health.

Table 1. Distribution of article characteristics and determinants of citation of 112 human MEHP publications

Determinants	Categories	N publications	N citation paths (% citations performed)
Statistical significance	Non-significant health effect	53	2705 (11%)
	Significant health effect	11	533 (10%)
	Both significant and non significant health effects	29	1350 (8%)
	No p-values reported	19	1096 (7%)
Direction of the effect	No harmful health effect of MEHP	11	607 (4%)
	Harmful health effect of MEHP	42	2421 (12%)
	Mixed results	42	1731 (10%)
	No point estimate reported	17	925 (7%)
Authors' conclusion	No harmful health effect of MEHP	30	1801 (12%)
	Harmful health effect of MEHP	35	1786 (10%)
	Mixed conclusion	5	156 (8%)
	Unclear conclusion	42	1941 (8%)
Study design	Cohort study	34	1168 (10%)
	Cross-sectional study	41	2596 (11%)
	Case-control study	10	946 (6%)
	Narrative review	10	657 (10%)
	Systematic review	6	273 (7%)
	Systematic review including meta-analysis	2	44 (5%)
Sample size	0 - 146 participants	26	1568 (11%)
	146 - 315 participants	31	1549 (10%)
	> 315 participants	37	1593 (8%)
Title of publication	Neutral title	93	4564 (9%)
	Title indicating conclusion	19	1120 (13%)
Number of affiliations	1 - 3 affiliations	48	2537 (9%)
	4 - 5 affiliations	33	1539 (9%)
	> 5 affiliations	30	1608 (11%)
Journal Impact Factor	< 3.2	38	1945 (7%)
	3.2 - 5.0	37	1895 (11%)
	> 5.0	37	1843 (11%)

Determinants	Categories	N publications	N citation paths (% citations performed)
Funding source	Not for profit funding	75	3588 (10%)
	For profit funding	1	48 (6%)
	Both for profit and not for profit funding	12	583 (9%)
	Funding not reported	22	1357 (8%)
	No funding applicable	2	108 (11%)
Number of references	0 - 40 references	37	1920 (9%)
	41 - 56 references	38	2015 (12%)
	> 56 references	37	1749 (8%)
Gender	Male	50	2112 (12%)
	Female	55	3063 (9%)
	Unknown	7	509 (7%)
Affiliation	University	88	4503 (10%)
	Government	18	996 (6%)
	Industry	1	48 (6%)
	Other	5	137 (15%)
Continent	North America	46	2609 (11%)
	Europe	30	1545 (9%)
	Asia	36	1530 (7%)

Table 1 gives an overview of the distribution of all tested determinants of citation over the 112 publications. The majority of publications reported a non-significant harmful health effect of MEHP. According to the authors' conclusion, there seems to be no clear consensus on the topic. A large number of publications, 42 out of 112, did not draw a clear conclusion with regard to the health effect of MEHP. The number of publications drawing a conclusion either in line or not in line with the hypothesis that a harmful effect of MEHP on human health exists was almost similar. The majority of the publications describe empirical studies, with a great variation in sample size, ranging from 30 participants to 76 million participants, with a median of 240. The journal impact factor ranged from 0 to 9.8, with a median of 3.8.

Table 2. Crude and adjusted odds ratios (95%CI) for the likelihood of being cited

Determinant	Categories	Crude OR	Adjusted OR*
Significance	Non-signficant results	1.00 (ref)	1.00 (ref)
	Significant results	0.96 (0.72-1.28)	0.98 (0.74-1.29)
	Mixed results	**0.77 (0.67-0.88)**	**0.75 (0.65-0.87)**
Direction of effect	No harmful effect	1.00 (ref)	1.00 (ref)
	Harmful effect	**2.79 (2.04-3.80)**	**2.46 (1.83-3.31)**
	Mixed results	**2.32 (1.59-3.38)**	**1.96 (1.41-2.74)**
Authors' conclusion	No harmful health effect of MEHP	1.00 (ref)	1.00 (ref)
	Harmful health effect of MEHP	0.91 (0.78-1.06)	0.79 (0.67-0.94)
	Mixed conclusion	0.87 (0.59-1.27)	0.84 (0.58-1.23)
	Unclear conclusion	**0.69 (0.60-0.81)**	**0.62 (0.52-0.75)**
Study design	Narrative review	1.00 (ref)	1.00 (ref)
	Cross-sectional study	1.15 (0.89-1.50)	1.15 (0.89-1.50)
	Case-control study	0.70 (0.49-1.00)	0.70 (0.49-1.00)
	Cohort study	1.14 (0.85-1.54)	1.14 (0.85-1.54)
	Systematic review (incl. Meta-analyses)	0.76 (0.52-1.10)	0.76 (0.52-1.10)
Sample size**	0 - 146 participants	1.00 (ref)	1.00 (ref)
	146 - 315 participants	0.94 (0.78-1.14)	0.89 (0.74-1.07)
	> 315 participants	**0.77 (0.63-0.94)**	**0.69 (0.57-0.83)**
Title of publication	Neutral title	1.00 (ref)	1.00 (ref)
	Title indicating conclusion	**1.34 (1.13-1.58)**	**1.26 (1.07-1.49)**
Number of affiliations**	1 - 3 affiliations	1.00 (ref)	1.00 (ref)
	4 - 5 affiliations	1.19 (1.00-1.41)	1.11 (0.92-1.34)
	> 5 affiliations	1.17 (0.99-1.38)	1.06 (0.89-1.26)
Journal Impact Factor**	< 3.2	1.00 (ref)	1.00 (ref)
	3.2 - 5.0	**1.49 (1.28-1.74)**	**1.49 (1.27-1.75)**
	> 5.0	**1.51 (1.28-1.78)**	**1.51 (1.25-1.82)**
Funding source	Not for profit funding	1.00 (ref)	1.00 (ref)
	Both for profit and not for profit funding	0.90 (0.73-1.12)	0.93 (0.74-1.15)
	Funding not reported	**0.80 (0.69-0.93)**	**0.80 (0.67-0.95)**
	No funding applicable	1.22 (0.76-1.96)	**3.73 (1.43-9.72)**
Number of references**	0 - 40 references	1.00 (ref)	1.00 (ref)
	41 - 56 references	**1.31 (1.09-1.56)**	**1.28 (1.07-1.52)**
	> 56 references	0.92 (0.77-1.11)	**0.79 (0.64-0.96)**
Gender	Male	1.00 (ref)	1.00 (ref)
	Female	0.78 (0.66-0.92)	0.71 (0.60-0.83)
Affiliation	University	1.00 (ref)	1.00 (ref)
	Government	**0.64 (0.53-0.76)**	**0.66 (0.54-0.81)**
	Industry	0.74 (0.33-1.67)	1.07 (0.40-2.82)
	Other	1.34 (0.89-2.03)	1.25 (0.81-1.93)
Continent	North America	1.00 (ref)	1.00 (ref)
	Europe	**0.83 (0.71-0.97)**	**0.83 (0.70-0.98)**
	Asia	**0.66 (0.55-0.79)**	**0.66 (0.56-0.78)**
Authority**	0 - 6	1.00 (ref)	1.00 (ref)
	7-39	**1.49 (1.24-1.79)**	**1.43 (1.17-1.75)**
	> 40	**1.66 (1.36-2.03)**	**1.64 (1.33-2.02)**

*Model adjusted for study design, as proxy for study quality

** Variables are categorised based on tertiles

Bold figures are statistically significant findings ($p<0.05$)

As displayed in table 2, significant findings and a positive authors' conclusion, in line with the hypothesis, were not found to be associated with a higher chance of being cited. The direction of the reported point estimate was associate with the chance of being cited: an effect size reflecting a harmful effect of MEHP was cited approximately 2.5 times more often compared to effect sizes reflecting no harmful health effect. The relationship between study outcome, either defined as statistical significance, direction of the effect or authors' conclusion, could not be explained by the variation in the number of health outcomes studied in each publication.

Other determinants significantly associated with a higher chance of being cited, although with limited effect sizes, were: journal impact factor, authority of the author, male gender, being located in North America and stating the conclusion in the title of the publication. Content related variables, such as study design and sample size, did not show any association with the chance of being cited.

Table 3. Concordance analsyses of the charactistics of the cited and citi publications

Determinant	Crude OR	Adjusted OR*
Significance	0.78 (0.54-1.13)	0.76 (0.52-1.11)
Direction of the effect	0.94 (0.77-1.14)	0.88 (0.72-1.07)
Authors' conclusion	0.89 (0.73-1.09)	0.88 (0.72-1.07)
Study design	**1.75 (1.41-2.17)**	**1.75 (1.41-2.17)**
Sample size	**1.53 (1.21-1.93)**	**1.53 (1.21-1.93)**
Title of publication	0.92 (0.76-1.10)	0.96 (0.79-1.16)
Number of affiliation	1.18 (0.98-1.41)	1.16 (0.97-1.40)
Journal impact factor	0.80 (0.64-1.01)	0.82 (0.65-1.03)
Funding source	1.02 (0.85-1.21)	1.01 (0.84-1.21)
Number of references	1.08 (0.90-1.30)	1.08 (0.90-1.30)
Gender	**1.25 (1.04-1.50)**	**1.24 (1.03-1.50)**
Affiliation	1.06 (0.88-1.28)	1.03 (0.85-1.24)
Continent	**1.40 (1.17-1.68)**	**1.40 (1.17-1.68)**
Self-citation	**3.22 (2.52-4.11)**	**3.16 (2.46-4.06)**

* Adjusted for study design

Bold figures are statistically significant findings ($p<0.05$)

As shown in table 3, where we assessed concordance between the cited and citing publication, the impact of the characteristics of the *citing* publication on the chance of citation appeared limited. With regard to study design, sample size, gender and continent, a small effect was found that authors were more likely to cite publications that were similar to their own characteristics. The only factor that was of great influence is self-citation; authors were three times more likely to cite their own work compared to that of others.

Discussion

Our study shows that citation bias is only partly present in the literature on MEHP. Reporting of significant results and the authors' conclusion do not influence the likelihood of being cited. However, a point estimate indicating a harmful effect of MEHP does increase the chance of being cited by 2.5 times. This effect cannot be explained by the study design or the number of health outcomes reported in each publication. Since the aim of the study was also to create a broader view on determinants of citation, we did identify other variables that significantly influence the chance of being cited in MEHP literature. These were mostly factors that related to the author of the cited publication, such as the authority and gender of the corresponding author as well as self-citation. Additionally, article characteristics that do not relate to the content of the publication, e.g. journal impact factor and the title of the publication, were significantly associated with the chance of being cited. Factors that relate to the content of the publication, such as the study design and the sample size, were not associated with the chance of being cited.

Recently, the literature on harmful health effects of phthalates has been the subject of a study on outcome reporting bias (29). Outcome reporting bias refers to the situation where only a selection of the originally measured outcomes, most likely the positive findings, are reported in the publication (30). To prevent outcome reporting bias, research should be based on a pre-registered research protocol. This study protocol is needed to distinguish between confirmatory testing and performing exploratory analyses, which is crucial in interpreting statistical results (31). To test the occurrence of outcome reporting bias, the methods and results of the phthalate publications were compared to the planned analyses as laid down in a research protocol (29). It was found that only a very small proportion of the phthalate publications were based on a research protocol. The research protocols that were used, lacked the level of detail to reproduce the study, making it very likely that selective reporting of outcomes was present (29). Just like citation bias, outcome reporting bias is a questionable research practice. Both situations relate to selective communication of information and thereby negatively influence knowledge development (32). In the literature on phthalate this is especially relevant, because of the large number of health outcomes it is often associated with, which makes it a field that is prone to false positive findings.

Our study had several limitations. Because MEHP is often studied together with other substances and in relation to a large variation of health outcomes, it is difficult to demarcate a network of publications that should refer to each other. Since we did not score the nature of each citation (e.g. supportive or not), it is possible that a part of the citations was done not because of MEHP. Additionally, we are aware of the fact that we have tested a high number of associations, without adjusting for multiple testing. Since most tested hypotheses were independent from each other, the risk of a false positive finding for each test was not interlinked. Adjusting for multiple testing would in that

case not increase the accuracy of the research. Only for study outcome, which was operationalized in three ways, multiple testing could lead to misinterpretation of the results. However, these results did not change when we reanalysed the data with a Bonferroni correction.

In conclusion, we can state that the amount of citation bias is limited in the human literature on the harmful health effects of MEHP. However, the chance of citation was significantly associated with the journal impact factor, male gender, authority of the author, graphic location, the title of the publication and self-citation.

References

1. Jannot A-S, Agoritsas T, Gayet-Ageron A, Perneger TV. Citation bias favoring statistically significant studies was present in medical research. Journal of Clinical Epidemiology. 2013;66(3):296-301.
2. Bornmann L, Daniel H-D. What do citation counts measure? A review of studies on citing behavior. Journal of documentation. 2008;64(1):45-80.
3. Ioannidis JP, Boyack K, Wouters PF. Citation metrics: a primer on how (not) to normalize. PLoS biology. 2016;14(9):e1002542.
4. Duyx B, Urlings MJ, Swaen GM, Bouter LM, Zeegers MP. Scientific citations favor positive results: a systematic review and meta-analysis. Journal of clinical epidemiology. 2017;88:92-101.
5. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. BMJ. 2009;339:b2680.
6. Nieminen P, Rucker G, Miettunen J, Carpenter J, Schumacher M. Statistically significant papers in psychiatry were cited more often than others. Journal of Clinical Epidemiology. 2007;60(9):939-46.
7. Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. Journal of clinical epidemiology. 2002;55(4):407-10.
8. Sawin VI, Robinson KA. Biased and inadequate citation of prior research in reports of cardiovascular trials is a continuing source of waste in research. Journal of clinical epidemiology. 2016;69:174-8.
9. Ioannidis JP, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. Jama. 2011;305(21):2200-10.
10. Moed HF, Colledge L, Reedijk J, Moya-Anegon F, Guerrero-Bote V, Plume A, et al. Citation-based metrics are appropriate tools in journal assessment provided that they are accurate and used in an informed way. Scientometrics. 2012;92(2):367-76.
11. Garfield E. The history and meaning of the journal impact factor. JAMA. 2006;295(1):90-3.
12. Patsopoulos NA, Analatos AA, Ioannidis JA. Relative citation impact of various study designs in the health sciences. JAMA. 2005;293(19):2362-6.
13. Onodera N, Yoshikane F. Factors affecting citation rates of research articles. Journal of the Association for Information Science and Technology. 2015;66(4):739-64.
14. Conen D, Torres J, Ridker PM. Differential citation rates of major cardiovascular clinical trials according to source of funding: a survey from 2000 to 2005. Circulation. 2008;118(13):1321-7.
15. National Research Council. Phthalates and cumulative risk assessment: the tasks ahead: National Academies Press; 2009.
16. Duty SM, Calafat AM, Silva MJ, Ryan L, Hauser R. Phthalate exposure and reproductive hormones in adult men. Human reproduction. 2005;20(3):604-10.
17. Kay VR, Chambers C, Foster WG. Reproductive and developmental effects of phthalate diesters in females. Critical reviews in toxicology. 2013;43(3):200-19.
18. Goodman M, LaKind JS, Mattison DR. Do phthalates act as obesogens in humans? A systematic review of the epidemiological literature. Critical reviews in toxicology. 2014;44(2):151-75.
19. Kim B-N, Cho S-C, Kim Y, Shin M-S, Yoo H-J, Kim J-W, et al. Phthalates exposure and attention-deficit/hyperactivity disorder in school-age children. Biological psychiatry. 2009;66(10):958-63.
20. Duty SM, Calafat AM, Silva MJ, Brock JW, Ryan L, Chen ZY, et al. The relationship between environmental exposure to phthalates and computer-aided sperm analysis motion parameters. Journal of Andrology. 2004;25(2):293-302.
21. Mariana M, Feiteiro J, Verde I, Cairrao E. The effects of phthalates in the cardiovascular and reproductive systems: A review. Environment International. 2016;94:758-76.
22. Cummings P. The relative merits of risk ratios and odds ratios. Archives of pediatrics & adolescent medicine. 2009;163(5):438-45.
23. Duty SM, Silva MJ, Barr DB, Brock JW, Ryan L, Chen ZY, et al. Phthalate exposure and human semen parameters. Epidemiology. 2003;14(3):269-77.

24. Jonsson BAG, Richthoff J, Rylander L, Giwercman A, Hagmar L. Urinary phthalate metabolites and biomarkers of reproductive function in young men. Epidemiology. 2005;16(4):487-93.
25. Main KM, Mortensen GK, Kaleva MM, Boisen KA, Damgaard IN, Chellakooty M, et al. Human breast milk contamination with phthalates and alterations of endogenous reproductive hormones in infants three months of age. Environmental Health Perspectives. 2006;114(2):270-6.
26. Hauser R, Meeker JD, Duty S, Silva MJ, Calafat AM. Altered semen quality in relation to urinary concentrations of phthalate monoester and oxidative metabolites. Epidemiology. 2006;17(6):682-91.
27. Swan SH, Main KM, Liu F, Stewart SL, Kruse RL, Calafat AM, et al. Decrease in anogenital distance among male infants with prenatal phthalate exposure. Environmental Health Perspectives. 2005;113(8):1056-61.
28. Hauser R, Calafat AM. Phthalates and human health. Occupational and Environmental Medicine. 2005;62(11).
29. Swaen GM, Urlings MJ, Zeegers MP. Outcome reporting bias in observational epidemiology studies on phthalates. Annals of epidemiology. 2016;26(8):597-9. e4.
30. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS one. 2008;3(8):e3081.
31. Veldkamp CLS, Bakker M, Van Assen MALM, Crompvoets E.A.V., Ong HH, Nosek BA, et al. Ensuring the quality and specificity of preregistrations. 2018.
32. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technology Assessment. 2010;14(8):1-193.

6

Selective citation in scientific literature on the human health effects of bisphenol A

M.J.E. Urlings, B. Duyx, G.M.H. Swaen, L.M. Bouter, M.P. Zeegers

This chapter has been published in *Research Integrity and Peer Review*:

Urlings, M. J. E., Duyx, B., Swaen, G. M. H., Bouter, L. M., & Zeegers, M. P. (2019). Selective citation in scientific literature on the human health effects of bisphenol A. Research Integrity and Peer Review, 4(1), 6. https://doi.org/10.1186/s41073-019-0065-7

Abstract

Introduction: Bisphenol A is highly debated and studied in relation to a variety of health outcomes. This large variation in the literature makes BPA a topic that is prone to selective use of literature, in order to underpin one's own findings and opinion. Over time selective use of literature, by means of citations can lead to skewed knowledge development and a biased scientific consensus. In this study we assess which factors drive citation and whether this results in the overrepresentation of harmful health effects of BPA.

Methods: A citation network analysis was performed to test various determinants of citation. A systematic search identified all relevant publications on human health effect of BPA. Data were extracted on potential determinants of selective citation, such as study outcome, study design, sample size, journal impact factor, authority of the author, self-citation and funding source. We applied random effect logistic regression to assess whether these determinants influence the likelihood of citation.

Results: 169 Publications on BPA were identified, with 12,432 potential citation pathways of which 808 citations occurred. Positive studies have a 1.5 times greater chance of being cited compared to negative studies. Additionally, authority of the author and self-citation are consistently found to be positively associated with the likelihood of being cited. Overall, the network seems to be highly influenced by two highly cited publications, whereas 60 out of 169 publications received no citations.

Conclusion: In the literature on BPA, citation is mostly driven by positive study outcome and author-related factors, such as high authority within the network. Interpreting the impact of these factors and the big influence of a few highly cited publications, it can be questioned to which extent the knowledge development in human literature on BPA is actually evidence-based.

Introduction

Bisphenol A (BPA) is a chemical substance, which is used in plastics of, for example, food containers and can linings. It is considered a potential endocrine disruptor, as it might bind to estrogen receptors in the body and mimic estrogen's function (1). Most research of the potential harmful effects of BPA and its underlying mechanism has been conducted using *in vitro* studies or animal models (2). In the in vitro setting, it was found that BPA can directly bind to androgen receptors and thereby block endogenous androgen action (3). Because of its various uses, exposure to BPA in humans is widespread. Epidemiological studies have linked exposure to BPA to a large variety of health outcomes, such as reproductive outcomes, metabolic diseases, behavioural outcomes and intermediate health effects (e.g. DNA methylation and oxidative stress) (2, 4-6). In 2012, the WHO concluded that the epidemiological evidence with respect to human health effects of bisphenol A is limited and not coherent across the different health outcomes (7). Additionally, the European Food Safety Authority has concluded that there is no health concern for humans at the expected level of intake (8). BPA has not only been debated in the scientific community. It has also been a topic of extensive public debate, in which different stakeholders are involved such as industry and non-governmental organisations (9, 10). The public discussion on the health risks of BPA, combined with the variety of health effects BPA has been linked with makes it a topic that is vulnerable to distorted use of the evidence. Especially when scientific evidence is the basis for decision-making processes, such as setting maximum levels of exposure, a complete and balanced view is crucial. Therefore, it is important to understand the knowledge development in this field of research.

Scientific knowledge development to a large extent is driven by citations. Due to the large and growing number of scientific publications in the biomedical domain and the limitation of the maximum number of references in many journals, it is often not feasible to refer to all available relevant literature. In many cases it is unclear on which grounds researchers decide to select the articles they cite. Selecting references based on their study results, usually meaning that positive studies are cited more often than negative studies is called citation bias (11). Citation bias has been studied in a variety of research areas, by using different methodologies and showing different results (12-15). Looking at the selection of references in a broader sense, authors might have different motives to select their references; which can take the form of justified (e.g. methodological quality of a publication) or unjustified determinants (e.g. study outcome) of selective citation. Determinants that have shown to be related to citation rate in multiple studies are sample size, study design, journal impact factors and the number of references (16-18). With regard to funding, it is often suggested that for-profit funding is less credible because only results that are preferred by the funder would be published (19). A study by Kulkarni et al (2007) showed that industry-funded studies that reported industry-favorable results, were indeed associated with a higher annual citation count

(20). Factors that have been occasionally linked to citation count are gender of the author, number and type of affiliations included in a publication, the authors' reputation and whether the title of the publication includes its conclusion or not (16, 18). It should be recognised that the effect of most determinants of selective citation will be located somewhere on the sliding scale between justified and unjustified determinants of citation with regard to their effect on knowledge development.

The objective of this study is to assess the prevalence and determinants of selective citation in human studies on bisphenol A in a quantitative manner.

Methods

The design of this study was described in a study protocol, which was finalized and published online prior to the data collection (https://bit.ly/2kiDK4Z). The main steps of the citation network analysis will be described in the following paragraphs.

Search strategy and article selection

All relevant publications were identified via Web of Science – Core Collections. Identification of articles by checking the reference lists was not applied, since this would interfere with the research question. Checking reference lists would result in an overrepresentation of articles that are cited within the network, whereas articles that have been neglected by the network would still be missed. To prevent missing important publications a broad search strategy was applied, namely ("Bisphenol A" OR "BPA") AND ("Human*"). No limitations with regard to the health outcomes studied were applied.

The article selection was carried out in two phases. The first selection round was based on the publication title, to limit the number of publications. The second selection round included studying abstracts, figures and tables, to finalize the network of human BPA studies. The complete article selection was conducted individually by two researchers, MJEU and BD, followed by several consensus meetings. In case no consensus could be reached, a third researcher (GMHS) was asked to take a decision.

Data extraction

All publications in the network were scrutinized for a number of characteristics that may be potential determinants of citation (see Table 1). Data extraction was performed independently by MJEU and GMHS. In all cases consensus was reached. Study outcome was scored in two ways. First, the data presented in the article were scored according to the reported statistical significance (statistically significant, not significant or mixed). A publication was scored statistically significant when the primary study outcome reported a p-value lower than 0.05. When multiple health outcomes were reported and the data showed p-values both higher and lower than 0.05, a publication was consid-

ered mixed. Narrative reviews and systematic reviews without meta-analysis, which do not present new data, were not scored on their statistical significance. Secondly, the study outcome is scored via the authors' conclusion in the publication. This can be either in line or not in line with the general hypothesis that bisphenol A has an adverse effect on human health. The health outcomes studied in the network were grouped into eight categories: reproductive outcomes, metabolic diseases, intermediate health parameters, hormone production, birth outcomes, behavioral outcomes, cancer and other. The study designs presented in this network were observational studies (experimental, cohort, cross-sectional and case-control studies), systematic reviews and narrative reviews. The journal impact factor at the moment of publication was measured via Web of Science.

The determinant 'authority of the corresponding author' was measured on the publication's level and can vary over time. All co-authors of all publications received an 'authority score', which was the number of citations received within this BPA network, during each year that the network was active. The authority of each publication was determined by the co-author with the highest authority score. We hypothesized that authors with a high authority increased the credibility of a publication and therefore would lead to a higher likelihood of being cited. Self-citation was defined as the situation in which at least one author was listed on both the cited and the citing publication.

Statistical analysis

Each publication in the network could take the role of the citing and the cited publication. We were solely interested in the effect of the characteristics of the cited publication on the likelihood of being cited and therefore the unit of analysis was the potential citation path. A potential citation path existed between one publication and every other publication in the network that was available online at the moment of submission. In the data set, each row represented a potential citation path followed by an indication whether the potential citation path had actually been realized or not and the characteristics of the cited publication of that citation path.

A single publication normally references multiple other publications, meaning that multiple citation pathways are leading to the same publication and are therefore not entirely independent. A multilevel approach was therefore required, in which the citation paths were nested under the citing publications. Random effect logistic regression was modeled to assess the effect of characteristics of the cited article on the likelihood of being cited.

First, univariate analyses were performed to test all potential determinants of citation, described in the previous paragraph, in the cited publication as predictor for the likelihood of being cited. Second, all analyses were adjusted for study design, which was considered a proxy for study quality.

Additionally, we assessed whether concordance between the characteristics of the cited and citing publication was a determinant of citation. Via fixed effect logistic regression analysis we tested whether concordance between the cited and citing publication determined the likelihood of citation. All statistical analyses were performed in Stata 13.

The outcomes of the logistic regression are reported as odds ratios. The odds ratio may overestimate the true relative risk in studies where the outcome is common (21). In our network, the overall chance of being cited is 6.5% (808 actual citations of 12,432 potential citations). With this incidence, we consider 'being cited' not very common and consequently the overestimation of the true relative risk will be small (21). Ultimately, the odds ratio gives an accurate estimation of the direction of the effect; only the exact magnitude of the effect should be interpreted with some caution. For the sake of readability of the publication, we interpret these values as if they are relative risks and therefore, for instance, speak about 'the likelihood of being cited for negative studies compared to positive studies'.

Results

A network of 169 publications on human effects of bisphenol A was identified, published between 2002 and the beginning of 2017 (figure 1). The publications are connected by 12,432 potential citations, of which only 808 citations were actually realized, making the likelihood of being cited in this network 6.5%.

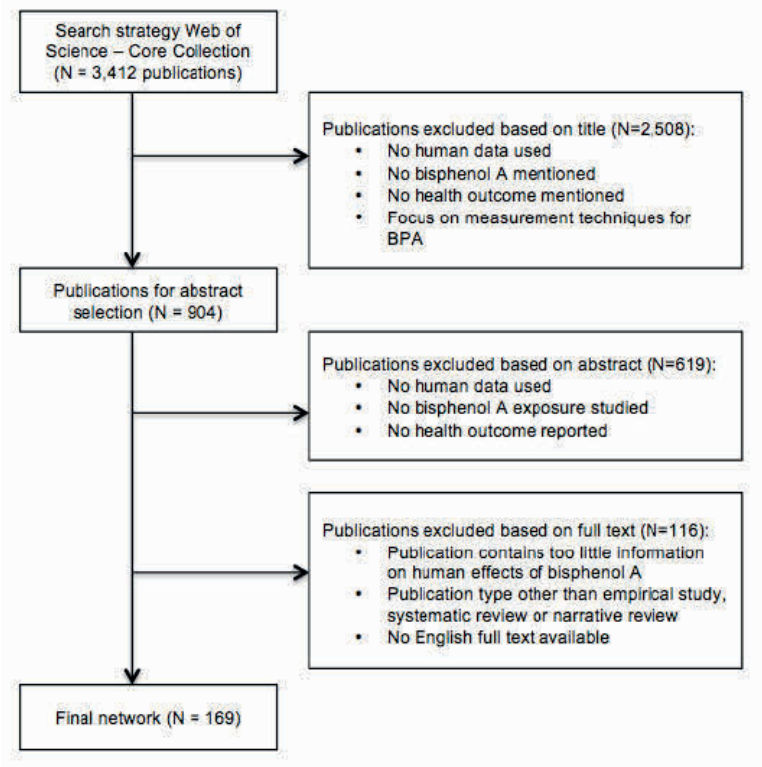


Figure 1. Flow diagram of the network selection process

Publication characteristics

Table 1 describes the distribution of the potential determinants of citation over the publications in the network. BPA was most frequently studied in relation to reproductive outcomes (N=49). The reproductive outcomes studied included, among others, polycystic ovary syndrome, miscarriage, sperm quality and in-vitro fertilisation implementation failure. In terms of study design, BPA is mostly studied in cross-sectional studies. The network contains 126 publications that report empirical data. These data were summarized in 43 publications, which were either narrative or systematic reviews. None of the systematic reviews included a meta-analysis. Since the network contained only one experimental study, this publication was classified as a cohort study. Looking at the evidence on adverse effects of BPA on human health, 40 publications reported statistically significant results, 36 publications reported non-significant results and 47 publications reported mixed results. The authors of 92 publications concluded that there was a harmful effect of BPA on human health. A mixed or unclear conclusion was drawn in 49 publications, against 28 publications that concluded there was no harmful health effect of

BPA. None of the studies were funded solely by for-profit organisations, which made it impossible to assess the effect of funding source as a determinant of citation. This underrepresentation of private parties in BPA research is also visible in the affiliation of the corresponding authors. Corresponding authors of 136 publications are affiliated with university whereas only one corresponding author is affiliated with industry.

Citation pattern

Although the first human BPA studies were published in 2002, the majority of the literature is published from 2010 onwards. Nevertheless, it seems that some of the early studies attract high number of citations. Two publications, a narrative review published in 2007 and a cross-sectional study published in 2008, received more than fifty citations (22, 23). On the other hand, sixty publications in the network received zero citations, which led to a very skewed distribution in the number of citations per publication. The median number of citations per publication was one.

Tabel 1: Characteristics of the bisphenol A network of 169 publications, 12,432 potential citation pathways and 808 realized citations

Variable	Categories	N publications	N potential citation pathways (% citations realized)
Study outcome			
Statistical significance	Yes	40	2800 (8%)
	No	36	2813 (5%)
	Mixed	47	3819 (9%)
	Not reported	46	2800 (8%)
Authors' conclusion	In line with hypothesis	92	6826 (7%)
	Not in line with hypothesis	28	2114 (4%)
	Mixed	32	2658 (5%)
	Unclear	17	834 (9%)
Content related determinants			
Health outcome	Reproductive outcomes	49	3853 (7%)
	Metabolic diseases	45	3218 (8%)
	Intermediate health factors	24	1594 (2%)
	Hormone production	18	1300 (8%)
	Birth outcomes	11	971 (6%)
	Behavioral outcomes	7	557 (3%)
	Cancer	4	286 (2%)
	Other	11	653 (6%)
Study design	Cohort study	34	2471 (7%)
	Cross-sectional study	63	5013 (9%)
	Case control study	29	2019 (4%)
	Narrative review	35	2374 (4%)

	Systematic review	8	555 (5%)
Sample size*	<168	42	3259 (6%)
	168 – 430	42	3154 (6%)
	> 430	43	3212 (10%)
Title of publication	Suggestive of conclusion	33	2327 (5%)
	Not suggestive of conclusion	136	10105 (7%)
Not content related determinants			
Number of affilations*	<3	51	4241 (5%)
	3 – 5	59	3445 (7%)
	> 5	59	4746 (7%)
Journal impact factor*	< 2.85	60	4188 (5%)
	2.85 – 4.6	53	4210 (7%)
	> 4.6	56	4034 (8%)
Funding source	Not for profit	135	10548 (7%)
	For profit	0	0 (0%)
	Both	4	369 (7%)
	Funding not reported	20	968 (3%)
	No funding applicable	10	547 (3%)
Number of references*	< 46	65	4152 (7%)
	46 – 58	47	4200 (8%)
	> 58	57	4080 (5%)
Corresponding author related determinants			
Gender	Male	86	7359 (7%)
	Female	74	4472 (6%)
	Unknown	9	601 (3%)
Affiliation	University	136	10403 (8%)
	Government	14	925 (6%)
	Industry	1	132 (1%)
	Other	18	1152 (7%)
Continent	America	73	6022 (8%)
	Asia	47	3612 (5%)
	Europe	42	2425 (4%)
	Australia	1	35 (3%)
	Africa	2	181 (2%)
	Middle East	4	157 (0%)

* For descriptive purposes, continuous variables have been transformed to categories based on tertiles

Univariate and multivariate analyses

Study outcome, measured both as statistical significance and as authors' conclusion in line with the hypothesis that BPA is harmful for health, shows a significant positive association with the likelihood of citation. Significant and positive studies are approxi-

mately 1.5 times more likely to be cited compared to negative and non-significant studies, an effect that remains after adjustment for study design. The concordance analysis showed that study outcome was not likely to be concordant between the cited and citing publication (OR 1.06 (0.79-1.42)).

Table 2: Univariate and multivariate analyses on potential determinants of selective citation

Variable	Categories	Crude OR	Adjusted OR
Study outcome			
Significance	Yes vs No	1.57 (1.28-1.92)	1.48 (1.21-1.80)
Authors' conclusion	In line vs Not in line with hypothesis	1.57 (1.29-1.92)	1.65 (1.34-2.03)
Content related determinants			
Study design	Narrative review	1.00 (ref)	
	Cohort study	1.61 (1.26-2.07)	
	Cross sectional study	2.00 (1.64-2.44)	
	Case control study	1.08 (0.84-1.38)	
	Systematic review	1.36 (0.99-1.87)	
Sample size *	< 168	1 (ref)	1.00 (ref)
	168-430	1.04 (0.88-1.22)	1.00 (0.85-1.17)
	> 430	1.62 (1.27-2.05)	1.39 (1.12-1.74)
Title of publication	Suggestive title vs not suggestive title	1.25 (1.07-1.45)	1.16 (1.00-1.35)
Not content related determinants			
Number of affiliations*	< 3	1.00 (ref)	1.00 (ref)
	3-5	1.46 (1.22-1.75)	1.27 (1.04-1.56)
	> 5	1.50 (1.24-1.82)	1.32 (1.06-1.65)
Journal Impact Factor*	< 2.8	1.00 (ref)	1.00 (Ref)
	2.8-4.6	1.21 (1.07-1.36)	1.08 (0.96-1.22)
	> 4.6	1.41 (1.22-1.63)	1.22 (1.06-1.41)
Funding source	For-profit vs not-for-profit **	NA	NA
	Not reported vs reported	0.41 (0.31-0.55)	0.74 (0.43-1.28)
Number of references*	< 46	1.00(ref)	1.00 (ref)
	46 – 58	1.24 (1.08-1.42)	1.10 (0.95-1.26)
	> 58	0.75 (0.63-0.89)	0.78 (0.65-0.92)
Author related determinants			
Gender of corresponding author	Male vs Female	1.00 (0.89-1.11)	0.97 (0.86-1.09)
Affiliation of corresponding author	Private vs public sector	0.94 (0.75-1.17)	1.31 (0.57-3.01)
Authority of the authors*	< 3	1.00 (ref)	1.00 (ref)
	3-26	2.21 (1.84-2.66)	2.16 (1.78-2.63)
	> 26	3.20 (2.59-3.96)	3.32 (2.64-4.18)
Self-citation	Yes vs no	5.14 (3.88-6.81)	5.16 (3.81-6.99)

* Adjusted model is adjusted for study design ** Continuous variables were categorised based on tertiles ***None of the publications was funded solely by for-profit organisations, therefore this analysis was not possible

Contrary to our expectation, systematic reviews were not more likely to be cited than narrative reviews in the full network. Sample size, number of affiliations and journal impact factor showed a moderate positive association with the likelihood of being cited, with ORs between 1 and 2. These effects could partly be explained by study design. The type of affiliation of the corresponding author, gender of the corresponding author, reporting of funding and the number of references showed no association with citation. Authority of the author and self-citation were found to have the strongest association with the likelihood of being cited. High authority, which was measured by a combination of the number of publications and the number of earlier citations in this field, increased the likelihood of citation by approximately three times. Authors were five times more likely to cite their own work compared to that of others.

Sensitivity analysis

Knowing that the number of citations per publication is very skewely distributed, we tested to which extent the results are driven by the two highly cited studies. As a data-driven, posthoc analysis we excluded these two studies, which received more than 50 citations (table 3). The significant effects that were found for the sample size and the number of affiliations in the overall network disappeared. This can be explained by the fact that one of the highly cited studies was a cross-sectional study with a large sample size of 1455 participants (23). The other highly cited study was a narrative review, which means the study had no specified sample size (22). Both studied were performed by relatively large research groups of 5 and 6 affiliations, respectively. In this sensitivity analysis the study outcome, journal impact factor, authority and the author and self-citation remained significantly associated with the chance of citation.

Table 3: Sensitivity analysis: Crude and adjusted odds ratios for the chance of being cited, excluding two highly cited publications (>50 citations)

Variable	Categories	Crude OR	Adjusted OR*
Significance	Yes vs no	1.57 (1.28-1.92)	1.48 (1.21-1.80)
Authors' conclusion	In line vs not in line with hypothesis	1.45 (1.18-1.77)	1.53 (1.25-1.87)
Study design	Narrative review	1.00 (ref)	
	Cross-sectional study	4.44 (3.26-6.04)	
	Case control study	2.61 (1.88-3.62)	
	Cohort study	3.93 (2.82-5.46)	
	Systematic review	3.28 (2.22-4.85)	
Sample size**	<168	1.00 (ref)	1.00 (ref)
	168-430	1.03 (0.87-1.21)	0.98 (0.84-1.16)
	>430	1.40 (1.09-1.79)	1.22 (0.97-1.54)
Title of publication	Conclusive title vs non-conclusive title	1.07 (0.91-1.25)	0.96 (0.82-1.12)
Number of afilliations**	<3	1.00 (ref)	1.00 (ref)
	3-5	1.46 (1.21-1.77)	1.06 (0.86-1.30)
	>5	1.09 (0.86-1.38)	0.70 (0.55-0.90)
Journal impact factor**	<2.8	1.00 (ref)	1.00 (ref)
	2.8-4.6	1.62 (1.41-1.87)	1.38 (1.19-1.60)
	>4.6	1.59 (1.33-1.89)	1.43 (1.20-1.70)
Funding source	not reported vs reported	0.46 (0.34-0.63)	0.83 (0.62-1.13)
Number of references**	<46	1.00 (ref)	1.00 (ref)
	46-58	1.07 (0.92-1.24)	0.96 (0.82-1.12)
	>58	0.51 (0.42-0.61)	0.56 (0.46-0.69)
Gender of corresponding author	Male vs female	1.14 (1.01-1.28)	1.20 (1.05-1.37)
Affiliation of corresponding author	Private vs public sector	1.07 (0.86-1.35)	1.08 (0.85-1.38)
Authority of the authors**	<3	1.00 (ref)	1.00 (ref)
	3-26	2.26 (1.85-2.76)	2.09 (1.68-2.60)
	>26	2.75 (2.18-3.47)	2.69 (2.08-3.48)
Self-citation	Yes vs no	5.46 (4.09-7.28)	5.05 (3.75-6.81)

* Adjusted model is adjusted for study design ** Continuous variables were categorised based on tertiles

Discussion

With this citation network analysis, we aimed to quantify the occurrence of citation bias in the human bisphenol A literature and the determinants that influence citation behaviour in this field. Based on the finding that positive studies have an approximately 1.5 higher likelihood of being cited compared to negative studies, we conclude that citation bias is present in the bisphenol A literature, although its magnitude might be limited. This effect was not confounded by study design and remained after excluding the most highly cited studies. Also, based on the results from the concordance analysis, citation bias does not appear to be influenced by the study outcome of the citing publication.

These results are in line with the findings of a recent meta-analysis on citation bias in various scientific fields, of which most were biomedical (24). This systematic review and meta-analysis showed that citation bias is prevalent throughout multiple biomedical research fields and that significant findings lead to an approximately 1.5 times higher chance of citation compared to non-significant findings (24). This was a pooled effect over a variety of disciplines, such as Alzheimer's disease, coronary heart disease and psychiatry (25-27). Also the finding that the authors' conclusion has a stronger effect on citation than the significance level of the data was confirmed by previous research in this meta-analysis (24).

Second aim of this study was to assess the effect of other potential determinants of citation. In the complete network of 169 publications sample size, journal impact factor, the number of affiliations involved, the authority of the author and self-citation were found to affect the likelihood of being cited. This was in line with our expectations, based on previous research in different research areas (28, 29). However, after exclusion of the two publications with the highest number of citations only the journal impact factor, authority of the authors and occurrence of self-citation appeared to be stable determinants of citation in the BPA literature. Different than study outcome influencing the likelihood of citation, the occurrence of self-citation is not necessarily leading to biased knowledge development. To some extent self-citation is inevitable, since academics are working to expand on their previous work (30). Of course, it might lead to selective overrepresentation of certain results and their interpretation and thereby skew knowledge development (30). Additionally, self-citation might be a way for authors to promote their own vision, which might be lead to an authority-based instead of evidence-based knowledge development. Before drawing conclusions on the possible effect of self-citation on knowledge development, we should keep in mind that self-citations can be used in different ways, apart from promoting certain results and substantiating an argument, authors refer to their own work to introduce a method that was described earlier or to explain the relevance of their research topic (31). Based on the current research, we could not conclude whether the amount of self-citation leads to a biased knowledge development in the BPA literature, since we did not asses in which paragraph of the publication self-citations were used.

In addition to the citation bias found, we should be aware that a large proportion of the literature seems to be completely ignored. More than one third of the publications received zero citations and even though these are both positive and negative publications, it means that part of the evidence is being left out of the picture and researchers are not appreciated for their work. Looking at the distribution of number of citations per publication over time, it seems that the highly cited publications are early publications in the field. With the growing amount of literature, the chance of not being cited at all seems to increase. Although it is logical and acceptable that founding publications are often mentioned to describe the research field, we should be aware that the more recent evidence is less often referred to. Especially because BPA is a research field that

is highly debated in risk assessment and risk management procedures, it is important to have a complete overview of all available evidence. The finding that a big part of literature is not valued in terms of citations has also been found in other research fields (32-34). E.g. Robinson and Goodman (2011) showed that, in the field of clinical trials, only a quarter of available trials got cited in the development of a new trial. Also, the number of trials that were cited did not increased with a bigger number of available trials (32). This gives support to the idea that an abundance of literature leads to reduced visibility for individual publications, potentially leading to research waste and misinterpretation of the literature in decision-making processes.

If we look, on the other hand, at the highly cited publications, it is remarkable that these studies have a narrative review and a cross-sectional study design, both of which are study designs that are typically not very highly valued. Although we did not look into the content of the publications in this study, we should be aware that both studies do not give a complete overview of the literature, as a systematic review would do, and thereby improve the chance of skewed knowledge development.

One of the limitations of the current study is that the search strategy was only applied to Web of Science – Core Collection, making it quite possible that some relevant publications have been missed. The search was limited to this database because Web of Science is the only database that has the option to download the publications together with their reference lists. This information was necessary to set up the database and perform statistical analysis. Nevertheless, we have no reason to believe that the identified determinants of selective citation would be different, if literature from other sources would have been included in the network.

Conclusion

Concluding, we found proof that citation bias is present in the human literature on BPA. Publications concluded a harmful health effect of BPA are 1.5 times more likely to be cited compared to negative publications. The association between other determinants and the chance of being cited is found to be hard to quantify, since our analysis was highly influenced by a low number of highly cited publications. Nevertheless, journal impact factor and author related factors such as author's authority and self-citation show consistent positive association with the chance of being cited. With these findings, we could conclude that the BPA literature seems to be mostly authority-based, instead of evidence-based.

References

1. Rubin BS. Bisphenol A: an endocrine disruptor with widespread exposure and multiple effects. The Journal of steroid biochemistry and molecular biology. 2011;127(1):27-34.
2. LaKind JS, Goodman M, Mattison DR. Bisphenol A and indicators of obesity, glucose metabolism/type 2 diabetes and cardiovascular disease: a systematic review of epidemiologic research. Critical reviews in toxicology. 2014;44(2):121-50.
3. Wetherill YB, Akingbemi BT, Kanno J, McLachlan JA, Nadal A, Sonnenschein C, et al. In vitro molecular mechanisms of bisphenol A action. Reproductive toxicology. 2007;24(2):178-98.
4. Rochester JR. Bisphenol A and human health: a review of the literature. Reproductive toxicology. 2013;42:132-55.
5. Tsutsumi O. Assessment of human contamination of estrogenic endocrine-disrupting chemicals and their risk for human reproduction. Journal of Steroid Biochemistry and Molecular Biology. 2005;93(2-5):325-30.
6. Hong YC, Park EY, Park MS, Ko JA, Oh SY, Kim H, et al. Community level exposure to chemicals and oxidative stress in adult population. Toxicology Letters. 2009;184(2):139-44.
7. Bergman Å, Heindel JJ, Jobling S, Kidd KA, Zoeller RT. Endocrine Disrupting Chemicals-2012. 2012.
8. Efsa Panel on Food Contact Materials. A. Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. EFSA Journal. 2015;13(1):3978-n/a.
9. Vogel SA. The politics of plastics: the making and unmaking of bisphenol a "safety". American journal of public health. 2009;99(S3):S559-S66.
10. Brewer PR, Ley BL. Contested evidence: Exposure to competing scientific claims and public support for banning bisphenol A. Public Understanding of Science. 2014;23(4):395-410.
11. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010;14(8):1-193.
12. Fanelli D. Positive results receive more citations, but only in some disciplines. Scientometrics. 2013;94(2):701-9.
13. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. Bmj. 2009;339:b2680.
14. Valachis A, Mauri D, Neophytou C, Polyzos NP, Tsali L, Garras A, et al. Translational medicine and reliability of single-nucleotide polymorphism studies: can we believe in SNP reports or not? International journal of medical sciences. 2011;8(6):492.
15. Robins RW, Craik KH. Is there a citation bias in the judgment and decision literature? Organizational Behavior and Human Decision Processes. 1993;54(2):225-44.
16. Bornmann L, Daniel H-D. What do citation counts measure? A review of studies on citing behavior. Journal of documentation. 2008;64(1):45-80.
17. Kostoff R. The difference between highly and poorly cited medical articles in the journal Lancet. Scientometrics. 2007;72(3):513-20.
18. Onodera N, Yoshikane F. Factors affecting citation rates of research articles. Journal of the Association for Information Science and Technology. 2015;66(4):739-64.
19. Sismondo S. How pharmaceutical industry funding affects trial outcomes: causal structures and responses. Social science & medicine. 2008;66(9):1909-14.
20. Kulkarni AV, Busse JW, Shams I. Characteristics associated with citation rate of the medical literature. PloS one. 2007;2(5):e403.
21. Cummings P. The relative merits of risk ratios and odds ratios. Archives of pediatrics & adolescent medicine. 2009;163(5):438-45.
22. Vandenberg LN, Hauser R, Marcus M, Olea N, Welshons WV. Human exposure to bisphenol A (BPA). Reproductive Toxicology. 2007;24(2):139-77.

23. Lang IA, Galloway TS, Scarlett A, Henley WE, Depledge M, Wallace RB, et al. Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults. Jama-Journal of the American Medical Association. 2008;300(11):1303-10.
24. Duyx B, Urlings MJ, Swaen GM, Bouter LM, Zeegers MP. Scientific Citations Favor Positive Results: A Systematic Review and Meta-analysis. Journal of Clinical Epidemiology. 2017.
25. Schrag M, Mueller C, Oyoyo U, Smith MA, Kirsch WM. Iron, zinc and copper in the Alzheimer's disease brain: a quantitative meta-analysis. Some insight on the influence of citation bias on scientific opinion. Progress in neurobiology. 2011;94(3):296-306.
26. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. BMJ. 1992;305(6844):15-9.
27. Nieminen P, Rucker G, Miettunen J, Carpenter J, Schumacher M. Statistically significant papers in psychiatry were cited more often than others. Journal of clinical epidemiology. 2007;60(9):939-46.
28. Jannot AS, Agoritsas T, Gayet-Ageron A, Perneger TV. Citation bias favoring statistically significant studies was present in medical research. J Clin Epidemiol. 2013;66(3):296-301.
29. Kivimaki M, Batty GD, Kawachi I, Virtanen M, Singh-Manoux A, Brunner EJ. Don't let the truth get in the way of a good story: an illustration of citation bias in epidemiologic research. Am J Epidemiol. 2014; 15;180(4):446-8.
30. Gami AS, Montori VM, Wilczynski NL, Haynes RB. Author self-citation in the diabetes literature. Canadian Medical Association Journal. 2004;170(13):1925-7.
31. Hyland K. Self-citation and self-reference: Credibility and promotion in academic publication. Journal of the American Society for Information Science and Technology. 2003;54(3):251-9.
32. Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. Annals of Internal Medicine. 2011;154(1):50-5.
33. Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? Clinical Trials. 2005;2(3):218-32.
34. Frandsen TF. Citing the innovative work of the original inventors. An analysis of citations to prior clinical trials. Information Research. 2017;22(1).

7

Selection bias as a threat for objective and independent risk assessment: a case study of bisphenol A

MJE Urlings

This chapter has emerged from discussions with Prof. Ellen Vos, professor of European Union Law at Maastricht University.

The chapter will be submitted to the *European Journal of Risk Regulation*. The used format is in line with this journal.

Abstract

Independence is a central concept in the European Food Safety Authority (EFSA). In their role as risk assessor, they should provide an independent, objective and transparent assessment of the scientific literature. However, independence is studied only from the perspective of industrial or political conflicts of interest. In the current study, we will broaden that scope by looking at scientific independence. Since EFSA panel members are also working as academics, they might have strong convictions that lead to intellectual conflicts of interests. One of the ways this conflict of interest may bias the risk assessment, is by selective use of the available evidence. As a case study we have studied the use of literature in the risk assessment on bisphenol A (BPA). We focused on the epidemiological studies mentioned in five updated risk assessments since 2006, and compared this with all 36 available epidemiological publications in the field of BPA. Thirthy epidemiological publications were included in the risk assessment, while six publications were missed by EFSA. Of these six excluded studies, five report a harmful effect of BPA on human health. Of the thirty publications included in the risk assessment, nineteen report a harmful effect. Apart from this selectivity, we have observed a focus on toxicology in the composition of the responsible panel and on the use of literature. Therefore, we conclude that selection bias with respect to the risk assessment exists in the field of BPA, with a focus on toxicological evidence. This finding threatens the goal of EFSA's risk assessment to be independent, transparent and objective.

Introduction

Various food scandals in the 1990's, such as the Bovine spongiform (BSE) crisis, called for a drastic reform of the way food safety is regulated [1]. The consumers' confidence in the way food safety was regulated by the European Commission was lacking and needed to be restored. Main problems in the regulation of food safety were the lack of transparency and the blurred relationship between scientific and political decisions [2]. In the food law as we now know it, regulation of food safety is based on risk analysis and therefore on scientific evidence [3]. To disentangle science and politics, the European Food Safety Authority (EFSA) was established [4]. EFSA got assigned the task to provide independent scientific advice on the legislation and policy in all fields related to food and feed safety, formally known as risk assessment [5]. Because EFSA is divided into scientific panels consisting of specialists on different subjects, they answer the need for specialisation and expertise. As laid down in the General Food Law, risk assessment needs to be 'based on the available scientific evidence and undertaken in an independent, objective and transparent manner' [6]. The outcomes of EFSA's risk assessment functions as input for the policy decisions of risk management, which remained with the European Commission [7].

Independence is a very central term in the establishment and the functioning of EFSA. However, independence is a concept that is not clearly described and that can be interpreted in different ways. In the academic literature, independence within the European context is mostly studied in relation to political and national activities as well as industry involvement. In the selection of panel members of EFSA, independence is operationalized by screening members on their recent ties to industry or political functions. This is a challenging task, since EFSA aims to gets influential scientific experts involved in their panels. It is likely that researchers who are an authority in their field, also have been involved in industry or national advisory tasks [8]. In the current study we

[1] Holland, D. & Pope, H. (2004) EU food law and policy. Chapter 3: New approach – Green paper and white paper. The Hague: Kluwer Law International

[2] Lavrijssen, S. & Ottow, A. (2012) Independent supervisory authorities: a fragile concept. In: P.J. Kuijper, T. Eijsbouts, R. Smits, J. Mathis, A. Schrauwen, K. Cseres (2012) Legal issues of economic integration. Kluwer Law International. Alphen aan den Rijn, The Netherlands

[3] European Commission (1997) The general principles of food law in the European Union – Commission green paper. COM 176.

[4] European Commission (2000), White paper on food safety. COM 719.

[5] Regulation (EC) No 178/2002 of the European Parliament and the Coucil laying down the general principles and requirements of food law, establishing the European food safety authority and laying down procedures in matters of food safety. Chapter 3: European Food Safety Authority, article 22.2: Mission of the Authority

[6] Regulation (EC) No 178/2002 of the European Parliament and the Coucil laying down the general principles and requirements of food law, establishing the European food safety authority and laying down procedures in matters of food safety. Chapter 2: General Food Law, Article 5.2: General objectives

[7] Everson, M. & Vos, E. (2009) The scientification of politics and the politicization of science. In. M. Everson, E. Vos (2009) Uncertain risks regulated. London: Routledge-Cavendish

[8] Vos, E. (2016) EU Agencies and Independence. In: D. Ritleng (2016) Independence and Legitimacy in the Institutional System of the European Union. Oxford University Press

will study independence from a different angle, namely in relation to the scientific content of EFSA's risk assessment.

Scientific independence

Another way to look at independence, that has not been studied much before, is scientific independence. Since all members of EFSA panels are scientists, they do have a potential intellectual conflict of interest, based on their training and experience. This brings the risk that they might, intentionally or not, overstress their own viewpoint and findings, compared to the work of others. Therefore, it is important to balance different viewpoints and different scientific disciplines in the composition of the panels. Scientific knowledge develops through different stages; from laboratory and animal studies, to human studies and randomised clinical trials. Each of these levels might bring forward different types of evidence and require different expertise to interpret this evidence. Since it is the goal of the risk assessment to evaluate all available evidence, it is important that researchers from different scientific backgrounds are represented in the panel.

Another important factor to keep in mind in the process of risk assessment, is that only published research can be included in this risk assessment. Although this seems straightforward, in the last decades much criticism has come forward about the completeness of academic reporting. Different types of bias in the reporting of research can be distinguished, which all influence the interpretation of the available evidence and the development of knowledge [9]. First, there is publication bias, which means the chance of being cited is associated with the study outcome. This phenomenon has been demonstrated in many research fields and results in an overrepresentation of positive findings [10]. Also in the research that does get published, there is a chance that only a part of the obtained results are being reported, for example by exclusively reporting results of certain subgroups of the studied population. This selective reporting of findings is called outcome reporting bias [11]. A third form of reporting bias is citation bias. This means that the chance of being cited is associated with the study outcome [12]. Citations are used to describe what is already known in a field, and citing only sources with a specific result, might drive the reader in a direction that is not necessarily evidence-based. Due to the great amount of scientific publications in each field, it is difficult to cite every relevant literature source and selective use of citations is inevitable. The

[9] Song, F., Parekh, S., Hooper, L., Loke, Y.K, Ryder, J.S., Alex J., Harvey, I. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess, 14*(8), 1-193.
[10] Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A. W., Cronin, E., Decullier, E., Easterbrook, P.J., Von Elm, E., Gamble, C., Ghersi, D., Ioannidis, J.P.A., Simes, J. & Williamson, P.R. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PloS one*, *3*(8), e3081.
[11] Page, M.J., McKenzie, J.E., Forbes, A. (2013). Many scenarios exist for selective inclusion and reporting of results in randomized trails and systematic reviews. Journal of Clinical Epidemiology, 66, 524-537.
[12] Duyx, B., Urlings, M.J., Swaen, G.M., Bouter, L.M., & Zeegers, M.P. (2017). Scientific citations favor positive results: a systematic review and meta-analysis. Journal of clinical epidemiology, 88, 92-101.

impact of selective citation in knowledge development, depends on the way in which the selection has taken place, e.g. based on study design or funding source. However, when this selection is based on study outcome, this leads to citation bias. Multiple studies have presented evidence for citation bias in various research areas [13,14] even in systematic reviews and meta-analyses [15], showing it is a widespread problem within academia. Publication bias and outcome reporting bias will contribute to an overestimation of the effect in the literature, but this cannot be detected by EFSA's experts. Citation bias, on the other hand, can be introduced by EFSA in the way literature is cited in the risk assessment. If this happens, the independence and objectivity of the risk assessment are at stake, since the selection of citations drives the outcome of the risk assessment. This is especially harmful, because the outcome of the risk assessment directly impacts the policy making in risk management.

In the current study we will assess scientific independence of EFSA by means of a case study, in which we test the occurrence of citation bias in the risk assessment on bisphenol A.

EFSA's risk assessment procedure

In the past, the risk assessment procedure was largely expert driven. Because panel members were selected on their expertise, it was assumed that all relevant literature was known by them. Nowadays, the great number of publications in most areas related to nutrition research creates a challenge to have a complete overview of the literature when performing a risk assessment. Therefore, a more systematic, evidence-based, approach is needed compared to the expert-driven approach. At first, it is crucial to clearly define the substance and outcome under study. Although risk assessment is a purely scientific process that seems to be very non-judgemental, a normative decision needs to be made to determine when a health outcome is considered as undesirable and therefore will be defined as a risk [16]. This decision on how to define the exposure and the outcome can impact the conclusion of the risk assessment. Especially in topics that are part of the public debate or in case of scientific uncertainty, a clearly described systematic approach is required give a complete overview of the available knowledge and to pinpoint where additional research is needed. However, implementing a systematic approach has proven to be difficult. One of the reasons for this is the wide range of topics that fall within EFSA's responsibility, e.g. intervention assessments, disease incidence estimates and exposure assessments. Also to aim of the assessment differs, de-

[13] Greenberg, S.A. (2009). How citation distortions create unfounded authority: analysis of a citation network. BMJ, 339, b2680.
[14] Ioannidis, J.P.A., Panagiotou, O.A. (2011) Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA, 305, 2200-2210
[15] Robinson KA, Goodman SN. A Systematic Examination of the Citation of Prior Research in Reports of Randomized, Controlled Trials. Annals of Internal Medicine ;154:50–55.
[16] Jensen, K.K. & Sandoe, P. (2002) Food safety and ethics: the interplay between science and values. Joural of Agricultural and Environmental Ethics. 15: 245-253

pending on the regulation under discussion. E.g. in case of evaluating a health claim, the focus will not be directly on assuring food safety, but on preventing misleading of the consumer [17]. In a topic such as food additives, the focus will be on food safety and protection of public health [18]. Because of different aims and differences in the available literature, different topics require different approaches in their risk assessment.

Case study: scientific independence in epidemiological studies on bisphenol A

To test whether scientific independence is at risk in the risk assessment of EFSA, we have performed a case study on bisphenol A. Bisphenol A (BPA) is a substance often present in plastics that are used in food containers. Since these plastics are in touch with food or beverages, they are considered as food contact material and are evaluated under EU food law, namely Commission Regulation (EU) No 10/2011 of 14 January 2011 on plastic materials and articles intended to come into contact with food [19]. This regulation aims to set limits for the migration of substances from food packaging into the food product. Regulating BPA is a politically sensitive topic, since in the mainstream media, it has often been framed as a risk for human health, specifically in children [20]. Since bisphenol A is a topic of public debate, it is even more important to perform an independent, transparent and objective risk assessment in which all available evidence is evaluated. EFSA panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids (CEF) has issued a scientific opinions on BPA in 2006, 2008, 2010 and 2015. With each risk assessment, the tolerable daily intake (TDI) was evaluated based on new emerging evidence. In 2006, the TDI was set at 0.05 milligram per kilogram body weight. In 2008 and 2010 re-evaluations of this TDI were performed, based on new toxicological evidence. In 2010, more attention was given to scientific uncertainty based on animal studies. However, the level of 0.05 mg/kg body weight has not been changed.

Methods

As an operationalisation to study scientific independence, we will assess the occurrence of citation bias in this risk assessment. The occurrence of citation bias can be tested by comparing the citations within the risk assessment to all available publications that were available. For practical reasons, we focus on the epidemiological studies on bisphenol A in the current study.

[17] Regulation (EC) No 1924/2006 of the European Parliament and the Council on nutrition and health claims made on foods

[18] Regulation (EC) No 1333/2008 of the European Parliament and of the Council on food additives

[19] Commission Regulation (EU) No 10/2011 of 14 January 2011 on plastic materials and articles intended to come into contact with food

[20] Fox, T., Versluis, E., Van Asselt, M.B.A. (2011) Regulating the use of bisphenol A in baby and children's products in the European Union: Current developments and scenarios for the regulatory future. European Journal of Risk Regulation

Recently, we have performed a citation analysis on the epidemiological studies on BPA in which we have performed an extensive literature search to identify the literature in this field [21]. We have identified 169 publications on human studies on BPA, which were published between 2002 and 2017. Within this scientific literature, citation bias was demonstrated and positive studies, which report a harmful health effect of BPA, were 1.5 times more likely to be cited compared to negative studies. In the current case study, we will test whether this identified literature is indeed included in one of the five EFSA's risk assessments. In case not all available literature was included in the risk assessment, we will assess which factors might explain this.

Results

In total, 30 epidemiological publications were mentioned in all the risk assessments upto 2015 (table 1). These publications studied BPA in association with a variety of health outcomes, which mostly relate to reproductive health, hormone production and metabolic diseases. Studies were scored as positive, in case the publications concludes that BPA has a harmful effect on health. Nineteen publications drew such a conclusion, whereas eleven publications were inconclusive or did not find a harmful health effect of BPA. When we compared the literature that was mentioned in the risk assessments with the total of epidemiological BPA publications, it appeared that six empirical publications were not mentioned in EFSA's work (table 2). This suggests that selective citation is actually present in the risk assessment, however we should look for explanations for why these publications are excluded.

Table 1. Epidemiological studies included in risk assessment by EFSA

First author	Title	Journal	Year	Study outcome*	Health outcome
Takeuchi	Serum bisphenol A concentrations showed gender differences, possibly linked to androgen levels	Biochemical and biophysical research	2002	Positive	Reproductive function
Takeuchi	Positive relationship between androgen and the endocrine disruptor, bisphenol A, in normal women and women with ovarian dysfunction	Endocrine Journal	2004	Positive	Hormone production
Lang	Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults	JAMA	2008	Positive	Metabolic disease
Padmanabhan	Maternal bisphenol-A levels at delivery: a looming problem?	Journal of Perinatology	2008	Negative	Birth outcomes

[21] https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/BZD0DM

First author	Title	Journal	Year	Study outcome*	Health outcome
Yang	Effects of bisphenol A on breast cancer and its risk factors	Archives of Toxicology	2009	Positive	Cancer
Cantonwine	Bisphenol a exposure in Mexico City and risk of prematurity: a pilot nested case control study	Environmental health	2010	Positive	Reproductive function
Galloway	Daily Bisphenol A Excretion and Associations with Sex Hormone Concentrations: Results from the InCHIANTI Adult Population Study	Environmental Health Perspectives	2010	Mixed results	Hormone production
Li	Occupational exposure to bisphenol-A (BPA) and the risk of Self-Reported Male Sexual Dysfunction	Human Reproduction	2010	Positive	Reproductive function
Li	Relationship Between Urine Bisphenol-A Level and Declining Male Sexual Function	Journal of Andrology	2010	Positive	Reproductive function
Meeker	Urinary Bisphenol A Concentrations in Relation to Serum Thyroid and Reproductive Hormone Levels in Men from an Infertility Clinic	Environmental science and technology	2010	Positive	Hormone production
Mendiola	Are Environmental Levels of Bisphenol A Associated with Reproductive Function in Fertile Men?	Environmental health perspectives	2010	Negative	Reproductive function
Mok-Lin	Urinary bisphenol A concentrations and ovarian response among women undergoing IVF	International journal of Andrology	2010	Positive	Reproductive function
Bloom	Serum unconjugated bisphenol A concentrations in men may influence embryo quality indicators during in vitro fertilization	Environmental toxicology and pharmacology	2011	Positive	Birth outcomes
Carwile	Urinary bisphenol A and obesity: NHANES 2003-2006	Environmental research	2011	Positive	Metabolic disease
Clayton	The Impact of Bisphenol A and Triclosan on Immune Parameters in the U.S. Population, NHANES 2003-2006	Environmental Health Perspectives	2011	Positive	Immune parameters
Miao	In Utero Exposure to Bisphenol-A and Anogenital Distance of Male Offspring	Birth defects research	2011	Positive	Birth outcomes
Shankar	Relationship between Urinary Bisphenol A Levels and Diabetes Mellitus	Endocrine Research	2011	Positive	Metabolic disease
Silver	Urinary Bisphenol A and Type-2 Diabetes in US Adults: Data from NHANES 2003-2008	Plos One	2011	Mixed results	Metabolic disease

First author	Title	Journal	Year	Study outcome*	Health outcome
You	Renal Function, Bisphenol A, and Alkylphenols: Results from the National Health and Nutrition Examination Survey (NHANES 2003-2006)	Environmental health perspectives	2011	Mixed results	Metabolic disease
Ehrlich	Urinary Bisphenol A Concentrations and Implantation Failure among Women Undergoing in Vitro Fertilization	Environmental Health Perspectives	2012	Positive	Reproductive function
Ehrlich	Urinary bisphenol A concentrations and early reproductive health outcomes among women undergoing IVF	Human reproduction	2012	Positive	Reproductive function
Fenichel	Unconjugated bisphenol A cord blood levels in boys with descended or undescended testes	Human Reproduction	2012	Negative	Birth outcome
Hanna	DNA methylation changes in whole blood is associated with exposure to the environmental contaminants, mercury, lead, cadmium and bisphenol A, in women undergoing ovarian	Human Reproduction	2012	Positive	DNA methylation
Li	Exposure to bisphenol A is associated with low-grade albuminuria in Chinese adults	Kidney International	2012	Positive	Albuminuria
Olsen	Associations between circulating levels of bisphenol A and phthalate metabolites and coronary risk in the elderly	Ecotoxicology and environmental safety	2012	Mixed results	Metabolic disease
Philippat	Exposure to Phthalates and Phenols during Pregnancy and Offspring Size at Birth	Environmental Health Perspectives	2012	Negative	Birth outcome
Shankar	Bisphenol A and Peripheral Arterial Disease: Results from the NHANES	Environmental Health Perspectives	2012	Negative	Metabolic disease
Spanier	Prenatal Exposure to Bisphenol A and Child Wheeze from Birth to 3 Years of Age	Environmental Health Perspectives	2012	Mixed results	wheeze
Wang	Urinary Bisphenol A (BPA) Concentration Associates with Obesity and Insulin Resistance	Journal of Clinical Endocrinology and Metabolism	2012	Positive	Metabolic disease
Zhao	The effects of bisphenol A (BPA) exposure on fat mass and serum leptin concentrations have no impact on bone mineral densities in non-obese premenopausal women	Clinical Biochemistry	2012	Mixed results	Bone health

* A positive study outcome refers to a statistically significant harmful health effect of BPA, a negative outcome refers to a non-significant finding or no harmful effect of BPA, mixed results can be found in case multiple outcomes are scored which are both positive and negative.

Table 2. Epidemiological studies not included in risk assessment by EFSA

Author	Title	Journal	Year	Study outcome	Health outcome
Yang	Urinary concentrations of bisphenol a in relation to biomarkers of sensitivity and effect and endocrine-related health effects	Environmental and molecular mutagenesis	2006	Negative	DNA damage
Hong	Community level exposure to chemicals and oxidative stress in adult population	Toxicology Letters	2009	Positive	Oxidative stress
Meeker	Semen quality and sperm DNA damage in relation to urinary bisphenol A among men from an infertility clinic	Reproductive toxicology	2010	Positive	Reproductive function
Melzer	Association of Urinary Bisphenol A Concentration with Heart Disease: Evidence from NHANES 2003/06	Plos One	2010	Positive	Metabolic disease
Meeker	Relationship between Urinary Phthalate and Bisphenol A Concentrations and Serum Thyroid Measures in US Adults and Adolescents from the National Health and Nutrition Examination	Environmental health perspectives	2011	Positive	Hormone production
Meeker	Urinary Concentrations of Parabens and Serum Hormone Levels, Semen Quality Parameters, and Sperm DNA Damage	Environmental health perspectives	2011	Positive	Reproductive function

* A positive study outcome refers to a statistically significant harmful health effect of BPA, a negative outcome refers to a non-significant finding or no harmful effect of BPA, mixed results can be found in case multiple outcomes are scored which are both positive and negative.

The empirical studies reported in table 2, which were not included in the EFSA's risk assessment, are all cross-sectional studies. This means the exposure to BPA and the health outcome under study are measured at the same time, making it difficult to draw conclusions with regard to causality. Additionally, these studies looked at BPA exposure in the general population, where the exposure level might be too low and with too little variation among the study participants to make a claim about the risk caused by the BPA. These arguments would have been acceptable reasons for EFSA to exclude them from the risk assessment. However, not all cross-sectional studies are excluded, which makes it unclear why the chosen selection was made.

Given the fact that EFSA has applied an extensive and reproducible search strategy in multiple databases, it seems unlikely that these studies were missed completely. It is possible that they were deliberately excluded for methodological reasons. However, no clear in- and exclusion criteria were formulated, which makes it impossible to reproduce the article selection. By mentioning clear reasons for excluding studies, the risk assessment would have been made more transparent and reproducible.

To test if the selection of literature might lead to bias, we have compared the study outcomes of the in- and excluded studies. In the excluded studies, all but one of the empirical studies is positive, indicating a harmful health effect of BPA in humans. This is a striking difference with the studies that are included in the risk assessment, which are much more balanced in terms of study outcome. Nineteen publications conclude that there is a harmful health effect of BPA, versus five publications reporting no harmful effect and six publications with a mixed conclusion.

Additional analysis

As a side note, it became apparent during this case study that only a very small proportion of the publications referred to in the risk assessment is epidemiological research. Therefore, we wondered which disciplines were present in the CEF panel that performed the risk assessment. Since the first BPA risk assessment of 2006, the CEF panel consisted of 74 unique members. With respect to the scientific background this included toxicology (n=26), chemistry (n=22), epidemiology (n=10), food technology (n=6), pharmacy (n=5) and engineering (n=3). Although a variety of disciplines are represented, this composition indicates a focus on mechanistic studies, rather than on population level evidence. This is also reflected in the use of evidence, since more that 500 references are made towards toxicological or mechanistic studies, opposing 36 epidemiological studies.

In light of independence, we checked whether the CEF panel members were personally active in the field of BPA. A first exploration showed that fourteen of these panel members were cited at least once in one of the risk assessments, all in the field of toxicology. Although this is not necessarily problematic, it could be questioned how these studies were identified and evaluated in the whole process of risk assessment.

This focus on one scientific discipline is a form of selection that impacts the independence and objectivity of the risk assessment and that might drive the decision in the risk management phase in a certain direction. Further studies are needed to study how the conclusion of the risk assessment might be impacted by this selection.

Discussion

Independence is an important concept in the functioning of EFSA. So far, much attention has been given to financial and political independence. Since EFSA is set up as a scientific body and has a great responsibility in assuring safety on the European food market, scientific independence, as well as objectivity and transparency, should be assured as well. One of the ways how independence can be threatened, is by intellectual conflict of interests and selective use of citations to stress ones own view. By selectively citing only a part of the available evidence, the outcome of risk assessment is not

scientifically sound and leads to unfounded conclusions in the risk management phase. In the current study, selective citation was identified in the epidemiological studies in the risk assessment of bisphenol A. Six cross-sectional studies were not included in the analysis, of which five showed a harmful effect of bisphenol A on health. Although there might be valid reasons to exclude these studies, this was not mentioned in the risk assessment. The risk assessment did not describe clear in- and exclusion criteria in the selection of publications taken into account in the risk assessment, reducing its reproducibility and transparency.

Apart from selective citation, we have identified another form of selection in the risk assessment procedure, namely the focus on toxicological evidence. This focus is visible in the composition of the CEF panel and in the use of literature in setting the tolerable daily intake for BPA. Although we recognize the importance of toxicology in assessing safety at the individual and micro level, it is only one aspect of the relevant knowledge. Where toxicological data is useful to understand the mechanism of action, epidemiology is providing evidence on a population level [22]. Both of these approaches should be addressed in a balanced way, in weighing the risk for human health.

Recommendations to deal with intellectual conflict of interests

Although it is challenging to assess whether a scientist has an intellectual conflict of interest, it is an understudied topic that can greatly impact the outcome of the risk assessment. The awareness should grow that also scientists are not completely objective, especially when it comes to the research field they are personally active in. Due to their working experience, they might have developed a certain conviction on the topic, which influences their review of the available evidence. Since the panel consists of many members, it might be expected that the panel corrects itself in case of biased decision making. However, we have seen the variation in the scientific disciplines represented in the CEF panel is limited, which increases the likelihood that these disciplines have a disproportionate input in the risk assessment. To avoid this problem, we would encourage EFSA to provide more variation with regard to the scientific disciplines represented in the panels. Also assuring sufficient variation *within* each discipline, to represent potential opposing views and to reduce the impact of personal convictions. To get a better insight in the intellectual conflict of interest among panel members, we would suggest that each panel member gives a written statement of their academic work in the topic under discussion, including their main findings. From this information, the personal view on the topic can be estimated. Depending on the topic, it might be decided that these panel members are excluded from the risk assessment. This approach, which is similar to how financial conflict of interests are dealt with, will improve transparency, independence and objectivity of EFSA's risk assessment.

[22] Pearce, N. (1999) Epidemiology as a population science. International journal of epidemiology. 28: 1015-1018

Limitations of this study

A limitation of the current study is the fact that we have only looked at the epidemiological research on BPA in the assessment of citation bias and we excluded toxicological studies. Although this is only a part of the scope of EFSA's work, it is a central part of evidence in assessing the risk of a product on human health. However, based on the current study, we cannot draw any conclusion on the occurrence of citation bias in the non-epidemiological literature. Secondly we only looked at whether or not selective citations occurred in risk assessment reports in a quantitative way. To get more detailed insight in scientific independence, we should also look at the content of the citations, to understand the reason why certain sources were cited in a specific context. Previous researchers have studied the content of citations in fields related to nutrition and found that citations in academic literature are often not giving a complete and reliable impression of the cited publication [23,24].

Conclusion

Concluding, in this case study on bisphenol A we have shown that selective citation has taken place in epidemiological studies in EFSA's risk assessment. There might have been valid reasons to exclude these studies, such as low study quality. In terms of transparency, we have learned that the procedure for article selection is not clearly reported and consequently cannot be reproduced completely. By reporting clear criteria for excluding publications, a better interpretation of the conclusion of the risk assessment can be made. Additionally, the CEF panel might have some selection bias in the scientific disciplines of the panel members, with a focus on toxicology and chemistry. This also is a potential threat for scientific independence of the panel, since scientists predominantly interpret evidence in light of their own discipline. A one-sided image might be created by focussing on only a part of the relevant research fields. From the current case study, we conclude that the goal of risk assessment to evaluate all available evidence in an objective, independent and transparent way, was probably not reached with the current approach.

[23] Ravnskov, U. (1995) Quotation bias in reviews of the diet-heart idea. Journal of Clinical Epidemiology, 48: 713-719

[24] Greenberg, Steven A. (2009). How citation distortions create unfounded authority: analysis of a citationnetwork. *BMJ, 339*, b2680.

8

General discussion

Note: The studies reported in this dissertation are part of a bigger project, the Sound Science project, including three other citation analyses. The current discussion summarizes all findings from the complete project. Therefore, part of this discussion is also reported in the dissertation 'Standing on one schoulder: citation bias in the epidemiological literature' written by Bram Duyx, which contains extensive reports of the three additional citation analyses.

Concretely, the overlapping parts concern the paragraphs 'Main findings of Sound Science project' and 'Pattern recognition on determinants of selective citation'.

This dissertation studied the occurrence of some types of bias in the reporting of scientific research and addressed its potential impact on knowledge development and science-based decision making. **Chapter 2** described a study on outcome reporting bias, studying protocol adherence and selective reporting in scientific publications. Unfortunately, we found that many research projects were performed without the use of a study protocol. Furthermore, the willingness to share available study protocols was surprisingly low. As a result, it was impossible to compare the initial analysis plan with the reported findings and the occurrence of outcome reporting bias could not be tested (1). Additional to outcome reporting bias, this dissertation studied the occurrence of citation bias. In **chapter 3**, we started with a systematic review and meta-analysis to map what is already known (2). We learned that evidence for the existence of citation bias has been found in multiple research fields. Additionally, we learned that many different methods were used to study citation bias. We aimed to develop a new citation analysis methods, learning from earlier methods. With this citation analyses method we did not only study citation bias, but also looked at other potential determinants of selective citation. In **chapters 4, 5 and 6** our citation analysis was applied to different research fields, namely the relationship between trans fatty acid intake and serum LDL- and HDL-cholesterol, epidemiological studies on phthalates and epidemiological studies on bisphenol A. The three citation analyses presented in this dissertation were part of a bigger project, named the Sound Science project. Within the Sound Science project three additional citation analyses were performed. These involved the following subjects: the relationship between swimming in chlorinated water and childhood asthma (3), the hygiene hypothesis (4) and the relationship between diesel emission and lung cancer in humans (5). Although each citation network analysis is merely a case study, by performing a total of six of these case studies we aimed to look for patterns and to distinguish between general and field-specific determinants of citation. In **chapter 7** we stepped outside of the scientific community and assessed selection bias in risk assessment as performed by the European Food Safety Authority (EFSA). This risk assessment functions as the basis for policy decisions by the European Commission. Selective use of evidence in this report can jeopardize EFSA's core values of independence and objectivity and lead to decisions that are not evidence-based.

This last chapter will discuss the main findings of my dissertation and place them into a wider context. Potential consequences of reporting bias and selective citation will be discussed, together with suggestions to reduce the problematic issues in the future. Finally, I will discuss the methodological insights gained while working on this thesis, including limitations of the study and suggestions for future research.

Main findings Sound Science project

The following paragraph will discuss the findings of all six citation analyses performed within the Sound Science project. The three citation analyses not published in this dissertation, on swimming in chlorinated water in relation to childhood asthma, the hygiene hypothesis and diesel emission in relation to lung cancer, are described in detail in the dissertation 'Standing on one schoulder: citation bias in the epidemiological literature' written by Mr. B. Duyx. Study protocols and additional information on all studies within the Sound Science project are shared publically or upon request in the online repository Dataverse (https://dataverse.nl/dataverse/SoundScience).

Overview of citation networks

The topics for the six networks were carefully chosen, based on a number of criteria. First, we did not study research topics in which one of the project members was personally involved. Nevertheless, we did focus on clinical and epidemiological studies on biomedical topics to make sure we had sufficient knowledge to understand and interpret the content of the publications. During the process of conducting the citation analyses, we learned that this is crucial to understand dynamics within a field. If needed, we contacted content experts in the phase of setting up the search strategy, to make sure we did not miss important search terms. Additionally, we aimed to define the subject of each network in a clearly delineated manner, to make sure all relevant publications could be identified and could be expected to cite to each other as a network. In five out of six networks this was done by defining one determinant and one or more associated health outcomes. In the network on the hygiene hypothesis we took a different approach and defined the network on the basis of the original hygiene hypothesis posted by Strachan in 1989 (6). For feasibility reasons we aimed for the networks to be between 80 and 200 publications, since the selection of the publications and the scoring of the publication characteristics were done manually by two assessors. This aim was met for all but one of the networks, namely the network on swimming in chlorinated water and childhood asthma. This network contained only 36 publications and therefore it has limited power and we should interpret its findings with caution.

Table 1 displays general information in the six citation networks. Five out of six networks contain around 100 publications or more, with a citation prevalence between 6% and 15%. This quite low citation frequency might be explained by the amount of literature in each field. The citation prevalence is much higher in the network on swimming in chlorinated water and its effect on childhood asthma, which contains only 36 publications. In a bigger network, researchers might not be aware of all published work, and even if they are, they might not have the possibility to cite all relevant publications.

The main aim of the Sound Science project was to study the association between study outcome and the chance of citation. Therefore, we found it relevant to look at the distribution of supportive and non-supportive publications in each field. Although study outcome could be scored in different ways, for example based on the results of the statistical analysis or the conclusion drawn by the authors, we chose to use the conclusion of the authors as the ultimate study outcome. A supportive publication means that the authors conclude that a harmful health effect has been identified. In case of a non-supportive publication, no support for that harmful effect, or even a reverse effect, is found. It is remarkable to see that, in each of the networks, the supportive studies outnumber the non-supportive studies in absolute terms. Although this might mean that a harmful effect is actually present, it might also be a sign of publication or outcome reporting bias, where non-supportive studies are less likely to be published or mainly supportive findings of the study are reported while the non-supportive findings are not.

Looking at the number of citations per publication, this is clearly a very scewed distribution. In all six networks the majority of publications receives only a small number of citations, whereas few publications attract a lot of attention and thereby heavily impacting the knowledge development. The question 'What makes a publication highly cited?' is difficult to answer with the approach followed in the Sound Science project, since this is probably also influenced by more content-related determinants that are not included in this project.

Table 1. Overview of citation networks

	Trans fatty acid and cholesterol	Chlorinated water and asthma	Bisphenol A and human health	Hygiene hypothesis	Phthalates and human health	Diesel exposure and lung cancer
Number of publications	108	36	169	110	112	96
Number of potential citation pathways	5041	570	12432	5551	5684	4317
Percentage occupied citation pathways	13%	34%	6%	7%	10%	16%
Median number of citations per publication	2	4	1	1	2	5
Maximum number of citations per publication	73	26	64	35	33	34
Time period of publication years	1990-2015	2002-2015	2002-2017	1995-2017	2000-2018	1988-2017
Number of supportive vs non-supportive publications	86 vs 16	16 vs 10	92 vs 28	41 vs 35	35 vs 30	51 vs 34

Content related network characteristics

Although there are many similarities among the six studied networks, also each network shows divergent characteristics. During the project, it became clear that each network shows specific characteristics, that must be understood to be able to understand and interpret the citation patterns in that field. The following paragraph will discuss some examples of these network-specific characteristics.

The network on **trans fatty acids and cholesterol** was selected as a topic since it is known that there is consensus on its core issue for already a long time. It is considered textbook knowledge nowadays that trans fatty acids increase LDL-cholesterol and decrease HDL-cholesterol which leads to increased likelihood of cardiovascular disease (7). It was clearly the first publication published on the topic (8) that became an authority, attracting much more citations compared to the rest of the network. In the network of 108 publications, 57 publications were narrative reviews. Together with nine systematic reviews, the summarizing review publications outnumber the empirical publications, which report new data. Although a sensitivity analyses, in which all review articles were excluded, did not show substantial differences with regard to the determinants of citation, it is questionable what the value of these reviews is and how they impact the research field. The network on **chlorinated swimming water and childhood asthma** was characterized by two research groups, of which one reports mostly supportive findings, whereas the other groups reports mostly non-supportive findings. No citation bias was found in this network. Potentially, these two opposing research groups do cite each other with the goal to refute each other's findings. Furthermore, it is difficult to score the health outcome in this field, since different operationalisations for asthma were used throughout the literature. This varied from use of biomarkers to coughing and self-reported asthma symptoms.

The network on **bisphenol A** differed from the previous networks, in the sense that multiple health outcomes were included. This makes that actually multiple subnetworks exists within the studied pool of publications, and it would be desirable to ignore potential citations of publications that discuss different health outcomes. However, we did choose it as a topic for a citation analysis, since it is a highly debated subject with also great interest of the general public. In this public debate, BPA is labeled as an endocrine disruptor which is expected to be bad for human health, without making reference to specific health outcomes (9). This general hypothesis that BPA is an endocrine disruptor makes it plausible that publications would cite to literature on BPA in relation to a variety of health outcomes. During the analyses of the network, it became apparent that a large proportion of publications, 60 out of 169 publications, did not receive any citations. On the other hand, two publications received more than 50 citations each, which makes them very influential. As mentioned before, the network on the **hygiene hypothesis** was defined in a different way, compared to the other networks. Two selection criteria were used in setting up this network. First, all publications had to cite the origi-

nal hygiene hypothesis as defined by Strachan in 1989. Second, they had to report on the relationship between having siblings or infections and rhinitis. This turned out to be a difficult approach for a citation analyses, since the theory under study develops over time. Therefore, the original theory gets outdated at some point and, due to new insights, develops into a new theory. If this happens, it is no longer needed to cite the previous theory, without leading to citation bias. Additionally, because the hypothesis is still under development, many publications might have an exploratory nature. This might be a justifiable reason for selective citation. This is different from the other networks we studied, where most research is hypothesis testing instead of hypothesis generating.

In both the bisphenol A network and the **phthalate** network numerous publications reported more than one health outcome. Because multiple associations were reported, it was difficult to score a publication as being supportive or non-supportive. Reporting multiple associations also increases the potential to be cited. The majority of the empirical studies in this network are cross-sectional studies.

The network on **diesel emission** and lung cancer distinguished itself by the fact that industry involvement played a larger role compared to the previous networks. The diesel industry involvement was associated with methodological concerns, such as studies that did not include smoking as a confounder. This might explain that harmful effects were almost exclusively reported in studies not funded by the diesel industry. Another difficulty in the assessment of health effects of diesel is the development of the product. Over the last decades, the quality of diesel fuel and motor technology has improved by removing potential harmful substances. This development of the product, made it difficult to compare studies and to judge the quality of long term studies, since they might already be outdated at the moment of publication. Finally, there is substantial heterogeneity in the way diesel emission exposure is measured. This can be done via job exposure matrices, but often only job occupation is used as proxy for job exposure, which is a less accurate method to measure diesel exposure.

Pattern recognition on determinants of selective citation

One of the aims of the Sound Science project was to gain insight in the question to which extent determinants of selective citation were field specific, or if certain factors would apply to multiple research fields. Table 2 reports an overview of all the examined potential determinants of citation over the six network analyses.

Table 2: Overview of odds ratios, adjusted for study design, of the association between various determinants and the likelihood of being cited in six different cases

Determinant	Categories	Trans fatty acids and cholesterol	Swimming in chlorinated water and childhood asthma	Bisphenol A and human health	Hygiene hypothesis	Phthalates and human health	Diesel emmission and lung cancer
Significance	Significant vs non-significant	**3.2 (2.5-4.2)**	1.3 (0.8-2.3)	**1.5 (1.2-1.8)**	-	1.0 (0.7-1.3)	-
Authors' conclusion	Positive vs. negative	**2.4 (1.9-3.1)**	1.4 (0.9-2.3)	**1.7 (1.3-2.0)**	**3.1 (2.2-4.3)**	0.8 (0.7-0.9)	**1.4 (1.1-1.7)**
Study design	Empirical vs review	**3.9 (3.2-4.8)**	**3.6 (2.3-5.7)**	**1.6 (1.3-1.9)**	**4.3 (3.2-5.7)**	1.1 (0.9-1.4)	1.1 (0.9-1.4)
Sample size	High vs low	**7.7 (4.6-12.8)**	**6.4 (3.1-13)**	**1.4 (1.1-1.7)**	**1.9 (1.2-3.0)**	0.7 (0.6-0.8)	**3.3 (2.4-4.4)**
Title of publication	Conclusive vs non-conclusive	0.9 (0.7-1.2)	1.1 (0.7-1.7)	**1.2 (1.0-1.4)**	0.3 (0.2-0.4)	**1.3 (1.1-1.5)**	0.8 (0.5-1.1)
Number of affiliations	High vs low	1.1 (0.9-1.4)	1.4 (0.9-2.3)	**1.3 (1.1-1.7)**	**1.6 (1.1-2.2)**	1.1 (0.9-1.3)	**3.4 (2.6 – 4.3)**
Journal impact factor	High vs low	**5.4 (3.7-7.8)**	**1.7 (1.1-2.8)**	**1.2 (1.1-1.4)**	**4.9 (3.2-7.6)**	**1.5 (1.3-1.8)**	**4.0 (3.0-5.5)**
Funding source	Profit vs not-for-profit	1.2 (0.9-1.6)	**2.0 (1.2-3.5)**	**NA**	0.8 (0.6-1.2)	0.9 (0.7-1.2)	**1.4 (1.0-1.9)**
Number of references	High vs low	0.8 (0.6-1.0)	**6.2 (3.1-12)**	0.8 (0.7-0.9)	0.8 (0.5-1.3)	0.8 (0.6-1.0)	**1.7 (1.3 – 2.2)**
Gender	Male vs female	**1.5 (1.3-1.9)**	**1.8 (1.1-2.8)**	1.0 (0.9-1.1)	0.7 (0.6-1.0)	**1.4 (1.2-1.7)**	1.0 (0.8-1.2)
Affilation	all other vs university	1.2 (0.8-2.0)	0.8 (0.5-1.3)	1.3 (0.6-3.0)	**2.0 (1.5-2.5)**	0.7 (0.6-0.8)	1.3 (1.0-1.5)
Continent	N.-America vs Europe	-	1.2 (0.7-2.2)	**1.6 (1.3-2.0)**	0.9 (0.5-1.4)	1.2 (1.0-1.4)	1.8 (1.5 – 2.2)
Authority	High vs low	**6.4 (4.7-8.9)**	**3.6 (1.9-7.1)**	**3.3 (2.6-4.2)**	**2.7 (2.0-3.7)**	**1.6 (1.3-2.1)**	**4.9 (3.7-6.5)**

Bold figures indicate statistically significant findings in the expected direction ($p<0.05$)

The main question of the Sound Science project concerned the occurrence of citation bias, which refers to the association between study outcome and the chance of citation. Study outcome was operationalised in two ways: by looking at **statistical significance** of the results and by scoring the **authors' conclusion**. In four out of six networks there is indeed evidence of a significant, positive association between study outcome and citation. One network, on epidemiological studies on phthalates showed a significant negative association, meaning that non-supportive studies are more likely to be cited compared to supportive studies (OR: 0.8). The magnitude of citation bias in case of a supportive association varies from and odds ratio of 3.1 in the literature on the hygiene hypothesis, to an odds ratio of 1.4 in the literature on diesel emision and lung cancer.

Contrary to general expectations, in each of the networks, empirical studies show a higher chance to be cited compared to review articles. In four out of six networks this effect of **study design** is statistically significant. This is opposing the general idea that empirical studies stop to be cited, once they have been taken up in a systematic review. Since many journal allow a maximum number of citations in each publication, it would be convenient to cite reviews in order to include as much information as possible. According to our findings, this does not seem to be the case. To explain our findings, this most likely relates to the fact that the majority of review publications are narrative reviews. These publications are not based on a systematic search strategy and therefore they are not providing an overview of all literature, but merely the highlights deemed important by the review author. We might question what the role and added value of narrative reviews in knowledge development is, since they tent to report selectively report on the available evidence. This selection might be influenced by the authors' view on the topic (10). Apart from knowledge development, narrative reviews might have a role in research agenda-setting and allocating reseach funds. Because of their less complex methodology, they might be easier to understand by policy makers and other non-scientists, and thereby influence knowledge development via this route. On the other hand, considering the great amount of narrative reviews, especially seen in the literature on trans fatty acids and cholesterol, we might argue that this is a form of research waste. Another potential reason why empirical studies are more likely to be cited compared to reviews, might be found in the way the topics of the reviews are defined. Especially in the case studies used in the Sound Science project, reviews tend to have a broader scope compared to rather specific empirical studies. Therefore, empirical studies might be more central in the network, whereas reviews are also cited by other, related, topics which are not included in these network analyses.

Together with study design, we considered **sample size** as an other factor that is closely related to study quality. In five out of six networks, a strong positive correlation was found between sample size and chance of citation. A note should be made here that sample size could only be scored for empirical studies and therefore the reviews were left out of this analysis. The reported figures in table 2 are adjusted for study design, so this will not have impacted the results.

Journal impact factor was found to be a strong and consistent determinant of citation. To make sure there was no interference between the journal impact factor and the chance of citation, we scored the JIF at the moment that the cited article was published. Journal impact factor is a measure of attention at the journal level, so this does not necessarily needs to be translated to a single publication (11). It is known that almost 80% of the impact factor of a journal is attributed to 20% of the publications (11, 12).

In the public opinion, research **funded** by for-profit organisations is often perceived as lower quality, because of the risk of potential conflict of interests. This was not reflected in the citation analyses. However, we should note that only a very small number of the publications were funded solely by for-profit organisations. In most situations, a consortium of both for-profit and not-for-profit partners was formed.

With regard to **gender** of the researchers, previous research has suggested that women publish less (13), but they might make up for that in terms of citations (14). In the current studies we only reported the gender of the corresponding author. Since this is likely to be the senior researcher, this might partly explain why indeed the number of men was higher than the number of women. In terms of citations, no large difference is found. Although a significant difference in favour of the men was found in three networks, the effect sizes were not very large. The other three networks did not show a significant effect of gender on citation.

Concerning the **geographical location**, most studies are performed in Europe and North America. No big differences were observed in their likelihood of being cited. Publications from other continents on the other hand are subordinated in several ways. First, they are limited in number, but also their likelihood of being cited is significantly lower than that of European and American publications. Unfortunatly, this analysis cannot be adjusted for quality related factors, due to the low power.

Authority of the author has been found as a stable determinant of citation. Although it is logical that expertise of a researcher leads to recognition by his or her peers, we should be careful in the way we operationalize authority. In the current research we have operationalized authority as the total number of citations received before publishing a new article. However, this number can be influenced by the researchers themselves. First, there is the problem of self-citation. Second, via anonymous peer review researchers can recommend their work to others, whose work they are reviewing. Another often seen activity is the so-called salami slicing of publications. By increasing the number of publications, the researcher gets more exposure and more opportunities to be cited. However, in this process the amount of valuable information in each publication decreases.

At the beginning of the project, we raised the idea that the chance of being cited might not only be influenced by the characteristics of the cited publication per se, but also by the concordance between the cited and citing publication. Our hypothesis was, that researchers will be most likely to cite publications that are in line with the citing publication. In table 3, the results of these concordance analyses are reported.

Table 3. Concordance analyses for all six citation network

	Trans fatty acids and serum cholesterol	Swimming in chlorinated water and childhood asthma	Bisphenol A and human health	Hygiene hypothesis	Phthalates and human health	Diesel exposure and lung cancer
Author conclusion	**1.9 (1.6-2.4)**	1.6 (1.0-2.6)	1.1 (0.8-1.4)	**3.4 (1.6-7.1)**	0.9 (0.7-1.1)	1.2 (1.0-1.4)
Study design	**1.3 (1.0-1.5)**	1.4 (1.0-1.9)	**1.4 (1.2-1.6)**	1.1 (0.9-1.3)	**1.8 (1.4-2.2)**	0.4 (0.3-0.5)
Journal impact factor	0.5 (0.1-2.2)	0.9 (0.6-1.3)	**1.3 (1.1-1.6)**	1.0 (0.8-1.3)	0.8 (0.7-1.0)	1.2 (0.9-1.5)
Funding source	0.8 (0.6-1.0)	1.1 (0.6-2.0)	**1.5 (1.2-1.9)**	1.2 (0.8-1.8)	1.0 (0.8-1.2)	0.7 (0.5-0.9)
Gender	1.1 (0.9-1.3)	1.1 (0.8-1.6)	1.1 (0.9-1.3)	1.0 (0.8-1.3)	1.2 (1.0-1.5)	1.1 (0.9-1.3)
Affiliation of corresponding author	1.1 (0.9-1.3)	**2.1 (1.0-4.4)**	0.9 (0.7-1.1)	0.8 (0.7-1.0)	1.0 (0.9-1.2)	1.0 (0.8-1.2)
Continent	-	-	-	**2.0 (1.6-2.5)**	**1.4 (1.2-1.7)**	**1.8 (1.5-2.1)**
Self-citation	-	**5.4 (3.2-9.2)**	**5.2 (3.8-7.0)**	**6.1 (3.7-9.9)**	**3.2 (2.5-4.1)**	**4.1 (2.9-5.7)**

Bold figures are significant in the expected direction ($p<0.05$)

The idea that authors are more likely to cite publications with the same study conclusion was not found as a general pattern. Combined with the citation bias outcomes from table 2, means that both supportive and non-supportive studies are likely to cite more supportive studies compared to non-supportive studies. With respect to funding source and affiliation, a division between industry-funded research and publicly funded research was expected, because of the earlier mentioned public distrust in studies funded and performed by industry. However, this effect was only seen in the network on bisphenol A for funding source and in the network on swimming in chlorinated water for affiliation.

Authors are more likely to cite studies that have been performed on the same continent. This might relate to the informal network, e.g. by being a member of national scientific associations, so that they are more familiar with each others work.

Self-citation has been found to be strongly associated with the chance of citation. Self-citation has been defined as at least one common author on the citing and the cited publication. Self-citation appears to be present in all six cases we studied. It is difficult to determine the effect of self-citation on knowledge development. As a researcher, you build a career in a certain field, which makes it logical that the publications follow each other and thereby refer to each other's findings. Additionally, once you have established some authority in a field, it becomes difficult to not cite your own work. Another reason for self-citations, is when one refers to a method that was previously developed by the scientist at issue. On the other hand, self-citations are an easy method to stress ones own conviction and to strenghten the evidence base for a certain statement.

Citation bias in empirical studies could be considered to be less problematic, since these publications present new evidence instead of only depending on what is already known. For reviews, this is different. Their function is to summarize the available evidence, to inform scientists and practitioners on the available knowledge in their field (15). To gain better understanding of the potential consequences of citation bias, we have looked at citation bias in review articles. For this analyses we have only included reviews (both narrative and systematic) as *citing* publications in combination with only empirical studies as *cited* publications. The results of these analysis are reported in table 4. We did not look at citation bias only in systematic reviews as citing publications, since the power was too limited for that.

Table 4. Odds ratios for the association between study outcome of empirical articles and the chance to be cited by all publications, and specifically by review publications

Cited publication	Citing publication	Trans fatty acids and serum cholesterol		Chlorinated water and childhood asthma		Bisphenol A and human health		Hygiene hypothesis		Phthalates and human health		Diesel exposure and lung cancer	
		N	OR	N	OR	N	OR	N	OR	N	OR	N	OR
Empirical studies	**All**	2160	**3.7 (2.8-4.9)**	351	**1.8 (1.1-3.0)**	7645	**1.6 (1.3-2.0)**	3517	**4.4 (2.9-6.5)**	4710	0.8 (0.6-0.9)	2948	**1.7 (1.4-2.2)**
Empirical studies	**All reviews**	1201	**3.8 (2.5-5.6)**	162	1.2 (0.6-2.6)	1758	**2.2 (1.5-3.2)**	1097	**5.6 (2.8-11)**	505	0.5 (0.3-0.9)	767	**1.5 (1.1-2.1)**

Bold figures are significant in the expected direction ($p<0.05$)

Methodological considerations

One of the main activities in my PhD project was the development of new methods to study the determinants of citation. In this paragraph we will describe the learning points, limitations and ideas for future research encountered during the project.

Learning points

Since citation bias and selective citation do not have a long research tradition, there are still many things to learn. Both in the developing of the citation analysis methods, the construction of the citation networks and in conducting the citation analyses, we have encountered several learning points which will be discussed in the following paragraph.

During the Sound Science project, six citation analysis were performed. In each of these analyses, all identified publications were scored on all potential determinants of citation independently by two assessors. Although scoring determinants such as 'study outcome' seem very straightforward, from experience we have learned it can evoke many discussions. Most epidemiological publications report more than one association. Starting off with a crude analysis to assess the primary research question, often several adjusted analyses and subgroup or dose-response analyses are reported. Although each of the network analyses were performed on the basis of a prespecified study protocol, there was sufficient room for discussion while executing the protocol. This room for interpretation of prespecified study protocol, can be described as undesirable researcher degrees of freedom (16). However, this problem is not completely avoidable, because before conducting the study it is difficult to predict which decision points you will encounter. The researcher degrees of freedom should be reduced by specifying the protocol as much as possible. All the discussions and decisions that were made during the research ought to be reported as an addendum to the study protocol. This is important because the later decisions might be data-driven. And also for improving the reproducibility of the study.

Second, we have learned that each research field has very specific characteristics, that might impact the citation behaviour. For example, in the network on swimming in chlorinated water and childhood asthma, there were two research teams that had opposing views. They did cite each other, but only to explain why they did not agree with one another. Therefore it is important to understand the dynamics in a research field in order to interpret the outcomes of the citation analysis. Additionally, these field specific differences make it difficult to apply the exact same citation analysis method to all fields. The operationalisation of determinants of citation might be different for various networks. For example in the network on BPA, none of the studies were funded solely by industry and therefore funding sources needed to be operationalised in a different way.

Thirdly, we observed that each network has a few publications that attract many citation while others are not cited at all. Although we already included quite an extensive list of potential determinants of citation, we did not find an explanation for these great differences between publications. In the network on BPA we have performed a sensitivity analysis excluding two highly cited publications. Only small differences were observed in comparison to the overall analysis, in the effect of sample size, title of the publication and number of affiliations. Most likely there are reasons in the content of publication that really explain why some publications attract many citations. To explore this idea, a citation analysis with a more qualitative approach should be performed. Additionally, during our citation analysis on the hygiene hypothesis, we learned that it is important to have clear-cut associations that are being studied. As the hygiene hypothesis per se developed over time, so did its predicted associations. This implied that an association could be supportive for one version of the hypothesis, but not for the other, which makes it very difficult to study citation bias. Therefore, clear associations between an exposure and health outcome need to be defined in order to be able to apply our methods.

Finally, citations can be made in different parts of a publication. Citations in the introduction have the function to underpin the relevance of the research question and is likely to describe contradicting views in the field. Citations in the methods section can be used to refer to previously described methods and will therefore be more neutral in nature. In the discussion, authors aim to answer the research question by interpreting their own findings and therefore are expected to use citations to underpin their point of view. These different locations of citations will have different effects on knowledge development and the risk of bias.

Limitations

One of the limitations of the presented study was in the selection of publications for the networks. Due to practical limitations, the search strategy could only be performed in Web of Science – Core Collection. This was needed, because this is the only database that has the possibility to download all reference lists. This information is needed to construct the citation network. By leaving out other medical databases, such as Medline or Pubmed, it is likely that an unknown part of the literature has been missed. However, we have no reason to believe that the selection of citation would show different patterns in these databases.

A second limitation relates to the statistical analysis. We did not include more than one confounder, namely study design, so we do not know if the effects reported can be explained by other confounding factors. For example, the association between study outcomes and citation might be explained by funding source, if industry funded studies are more likely to find supportive results compared to publically funded studies. In the

initial research protocol, we planned to perform a multivariate analysis that included all statistically significant determinants. However, when doing this wide confidence intervals were found, making it impossible to interpret these findings. Most likely this is caused by the creation of small strata, leading to limited power and high likelihood of chance findings. Additionally we did not manage to find a valid universal measure for study quality. Available quality checklists were tested on samples in two cases. However, each study design requires another checklist and not much quality variation was found within the study designs. Furthermore, no checklist was available for narrative reviews, which comprised half of the network in the case of trans fatty acids and cholesterol. Therefore, study design was used as a proxy for study quality. In the network on diesel exposure this was expanded by a number of field specific factors. This included the type of diesel, exposure assessment and whether or not a study adjusted for smoking. These factors were important for this specific network, since the type of diesel developed over time. The level of detail in the exposure assessment varied. Sometimes job occupation was used as proxy for exposure, whereas other publications used a more detailed exposure measurement. Adjustment for smoking was important since this was likely a very important confounder for the association between diesel and lung cancer. Thirdly, some publications reported a multitude of results, and not all of them were relevant for the association under investigation. This implies that we cannot exclude that a publication was cited because of the other association instead of the one under investigation.

A final study limitation lies in the fact that a number of continuous variables have been reduced to three categories. This concerns the determinants sample size, number of affiliations involved, journal impact factor, number of references and authority of the author. We have experimented with logistically transforming the data, instead of making categories, but the results were more difficult to interpreted in a meaningful way. We recognize that with the use of categories, the power is reduced. On the other hand, literature suggests that the number of type I and type II errors are not impacted, in case no multicollinearity is expected (17), which is not the case in our analyses.

Further research opportunities

Based on the described learning points and study limitations, there is still much work to do. Regarding the still growing number of scientific publications published each year (18), it is of great importance to better understand the biases which influence knowledge development. First, as an extension of the current approach, we could score citation not on the level of the complete publication, but score the citations in each paragraph of the publication. This would help to interpret the effect of selective citation in the different parts of a publication and estimate its effect on knowledge development. Second, a more qualitative approach can be followed to also score the nature of

the citation. In the current analysis, we do not know if citations are correct and justified in terms of content. Previous research showed that the interpretation of the citing authors is not always in line with the information provided by the cited authors and this interpretation varies between different citing publications (19).

Furthermore, it would be worthwhile to search for ways to automate the detection of different forms of reporting bias, including publication bias, outcome reporting bias and ctation bias, and calculate the additional effect of these biases. For example for publishers, it would be very relevant to have an insight in the occurrence of outcome reporting bias in empirical articles and citation bias in review articles. Also for research funders and people in decision-making positions, it is relevant to know to which extent the development of a research field has been driven by selective reporting. Recently, a study has been published where the additional effect of publication bias, outcome reporting bias and citation bias are shown in anti-depression drug randomized controlled trials (20). Including such an overview in applications for research grants and in systematic reviews, would reduce research waste and improve the evidence base in decision-making. However, one important condition in this approach is that all planned studies are recorded in a standardized way and made publically available, in order to be able to measure publication bias and outcome reporting bias.

Potential solutions for reporting biases

When we study research misconduct (RM) and questionable research practices (QRP), we often refer to it as being caused by human behaviour. This implies it is an individual's responsibility to behave ethically. Consequently RM and QRP could be treated by punishing the individual researcher. For behaviours such as fraud and fabrication, this seems indeed be the suitable approach. In case of questionable research practices, however, it will not be sufficient to address only the individual's responsibility. To tackle this problem, more attention should be given to removing the perverse incentives, such as granting promotions only on the basis of number of publications, and changing the research culture. Even more so because the individual researcher might not be aware of engaging in QRPs. For example in citation bias, this selection of citations might not always be deliberate.

Also we should keep in mind that one publication including selected citations, will not bias a complete field. The problem lies in the fact that every publication cites only a selection of available evidence, and this selection will be most often in favour of large studies reporting significant findings. In order to decrease problems related to publishing, the solutions should be looked for not on the individual level, but in the research culture. Publishing is a central activity in the way academics communicate and how knowledge develops. Since factors like the number of publications in high impact factor journals and the number of received citations are of big influence on receiving promo-

tions and research grants, this leads to publication pressure and competition among researchers. Consequently, researchers are tempted to cut corners and to violate basic methodological principles in order to drive the research towards the most favourable results (21, 22). By using other metrics, that are more related to quality of research or collaborations with other researchers, the publication pressure can be reduced. Also by teaching about the occurrence and consequences of QRPs in an early stage and introducing role models, QRPs can be avoided.

Registries of study protocols can serve multiple purposes in reducing reporting biases (23). First, by clearly defining the research aim, study design and analysis plan, the study is replicable (24). Second, when performing a systematic review and meta-analysis to summarize the available evidence, this is no longer limited to published study results. By extracting study results from registries, the information that is missed due to publication and outcome reporting bias can be taken into account and a more accurate estimate of the effect can be reported. This would reduce the problem that effect sizes reported in meta-analyses overestimate the true effect due to publication bias, as was shown by multiple studies (25, 26). In case of trials, which need to be assessed by medical ethics committee, the value of study registries has been acknowledged and it is starting to become common practice to register research protocols (27, 28). A similar movement is encouraged in observational research, especially since these studies are known for testing a high number of associations, increasing the risk of coincidence findings. The first chapter of this dissertation showed that use of study protocols and sharing of protocols is not yet common practice in observational epidemiology. Additionally, we learned that the level of detail in the obtained protocols was generally very low and varied greatly between studies. For example, many protocols did not provide a power calculation or a concrete statistical analysis plan. In order to efficiently make use of study protocols in reducing publication and outcome reporting bias, it would be useful to set standards with regard to the content and the level of detail a protocol should provide. A similar initiative, which was started in social psychology a few years ago, to reduce reporting biases and improve reproducibility of science is the use of Registered Reports (29). This means manuscripts are peer-reviewed and accepted by a journal prior to data collection. In this manner, Registered Reports focus on the relevance of the research question and the proposed methods, instead of the results. Thereby, the incentive for researchers to not submit replication studies and non-significant findings to journals because the chance of publication is low, is changed (29, 30).

Conclusion

As a conclusion, we can state that reporting biases play a big role in the development of knowledge, especially because they are difficult to identify. Publications which are not

based on a study protocol create unclarity, since it is unclear which initial hypothesis was tested and part of the observed results are likely to remain unreported.

As an addition to the information lost in publication and outcome reporting bias, citation bias seems to be a problem in several research fields. This overrepresentation of supportive findings will drive knowledge development, research agenda setting and the public opinion, and thereby putting evidence based decision making at risk. With regard to other determinants of citation, many differences between research areas were identified. However, some general determinants have been found. Factors that seem to be consistently positively associated with citation are journal impact factor, authority of the author and self-citation. The effect of quality indicators such as study design and sample size differ between research fields, both in the direction of the association and its magnitude.

Ultimately, we have learned that citation analyses are difficult to automate and cannot be done without understanding the context of the research topic at issue. At first, this is needed to make sure all available publications are identified. Second, each network has specific characteristics that need to be taken into account, such as relevant confounders and influential publications. Nevertheless, more effort should be put into promoting complete reporting of findings and balanced citations of previous work in order to make sure knowledge development takes place in an evidence-based manner. Reducing the problem of selective reporting will increase the credibility of science-based decisions, such as in policy making and clinical therapies. Further studies are needed to require more detailed knowledge on selective citation, by testing the correctness of the content of citations and studying the content-related effect of selective citation in decision-making processes.

References

1. Swaen GM, Urlings MJ, Zeegers MP. Outcome reporting bias in observational epidemiology studies on phthalates. Annals of epidemiology. 2016;26(8):597-9. e4.
2. Duyx B, Urlings MJ, Swaen GM, Bouter LM, Zeegers MP. Scientific citations favor positive results: a systematic review and meta-analysis. Journal of clinical epidemiology. 2017;88:92-101.
3. Duyx B, Urlings MJ, Swaen GM, Bouter LM, Zeegers MP. Selective citation in the literature on swimming in chlorinated water and childhood asthma: a network analysis. Research integrity and peer review. 2017;2(1):17.
4. Duyx B, Urlings MJ, Swaen GM, Bouter LM, Zeegers MP. Selective citation in the literature on the hygiene hypothesis: a citation analysis on the association between infections and rhinitis. BMJ Open. 2019;9:e026518
5. Duyx B, Urlings MJ, Swaen GM, Bouter LM, Zeegers MP. Selective citation in the literature on diesel exposure and lung cancer: a citation analysis.
6. Strachan DP. HAY-FEVER, HYGIENE, AND HOUSEHOLD SIZE. Br Med J. 1989;299(6710):1259-60.
7. Hall JE. Guyton and Hall textbook of medical physiology e-Book: Elsevier Health Sciences; 2015.
8. Mensink RP, Katan MB. Effect of dietary trans fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. New England Journal of Medicine. 1990;323(7):439-45.
9. Rochester JR. Bisphenol A and human health: a review of the literature. Reproductive toxicology. 2013;42:132-55.
10. Joyce J, Rabe-Hesketh S, Wessely S. Reviewing the reviews: The example of chronic fatigue syndrome. JAMA. 1998;280(3):264-6.
11. Neylon C, Wu S. level metrics and the evolution of scientific impact. PLoS biology. 2009;7(11):e1000242.
12. Campbell P. Escape from the impact factor. Ethics in science and environmental politics. 2008;8(1):5-7.
13. West JD, Jacquet J, King MM, Correll SJ, Bergstrom CT. The role of gender in scholarly authorship. PloS one. 2013;8(7):e66212.
14. Eisenberg N, Martin CL, Fabes RA. Gender development and gender effects. 1996.
15. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Annals of internal medicine. 1997;126(5):376-80.
16. Wicherts JM, Veldkamp CL, Augusteijn HE, Bakker M, Van Aert R, Van Assen MA. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. Frontiers in Psychology. 2016;7:1832.
17. Iacobucci D, Posavac SS, Kardes FR, Schneider MJ, Popovich DL. The median split: Robust, refined, and revived. Journal of Consumer Psychology. 2015;25(4):690-704.
18. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology. 2015;66(11):2215-22.
19. Elkiss A, Shen S, Fader A, Erkan G, States D, Radev D. Blind men and elephants: What do citation summaries tell us about a research article? Journal of the American Society for Information Science and Technology. 2008;59(1):51-62.
20. De Vries Y, Roest A, de Jonge P, Cuijpers P, Munafò M, Bastiaansen J. The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. Psychological medicine. 2018;48(15):2453-5.
21. Bouter LM. Commentary: Perverse incentives or rotten apples? Accountability in research. 2015;22(3):148-61.
22. Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. Science and engineering ethics. 2007;13(4):437-61.
23. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. Journal of clinical epidemiology. 2000;53(2):207-16.

24. Bouter LM. Unsuccessful replication is not a sign of research misconduct. Journal of clinical epidemiology. 2018;104:140-1.
25. Dechartres A, Atal I, Riveros C, Meerpohl J, Ravaud P. Association Between Publication Characteristics and Treatment Effect Estimates: A Meta-epidemiologic Study. Annals of internal medicine. 2018;169(6):385-93.
26. Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. PLoS medicine. 2008;5(11):e217.
27. Dickersin K, Rennie D. The evolution of trial registries and their use to assess the clinical trial enterprise. JAMA. 2012;307(17):1861-4.
28. Rennie D. Trial registration: A great idea switches from ignored to irresistible. JAMA. 2004;292(11):1359-62.
29. Nosek BA, Lakens D. Registered reports. Hogrefe Publishing; 2014.
30. Chambers CD, Dienes Z, McIntosh RD, Rotshtein P, Willmes K. Registered reports: realigning incentives in scientific publishing. Cortex. 2015;66:A1-A2.

Summary

This dissertation studied the occurrence of biases in the reporting of scientific research and addressed its potential impact on knowledge development and science-based decision making. Scientific publications are still the main form of communication among scientists in the development of knowledge. Biases in this process of publishing can be rather subtle, but of high impact on the development of knowledge. Chapter 2 described a case study on outcome reporting studying protocol adherence and selective reporting in scientific publications on phthalates. Use of study protocols is important in interpreting research findings and to prevent reporting of false positive outcomes. Unfortunately, we found that many research projects were performed without the use of a study protocol. Furthermore the willingness to share available study protocols was surprisingly low. As a result, it was impossible to compare the initial analysis plan with the reported findings and the occurrence of outcome reporting bias could not be tested. Additional to outcome reporting bias, this dissertation studied the occurrence of citation bias. Specifically, citation bias means the chance of being cited depends on the study outcome. Mostly this means that positive studies are cited more than negative studies. In chapter 3, we started with a systematic review and meta-analysis to map what is already known. We learned that evidence for the existence of citation bias has been found in multiple research fields. Additionally, we learned that many different methods were used to study citation bias. We have aimed to develop a new citation analysis methods, learning from earlier methods. With this citation analyses method we did not only study citation bias, but also looked at other potential determinants of selective citation. In chapters 4, 5 and 6 our citation analysis was applied to various research fields, namely the relationship between trans fatty acid intake and serum LDL- and HDL-cholesterol, epidemiological studies on phthalates and epidemiological studies on bisphenol A. The networks included all relevant scientific publications in each field, identified via a systematic search. The citation network analyses answered two questions: “Does citation bias occur?” and “Which other determinants influence the likelihood of citation?”. Apart from study outcome, the following determinants were tested: study design, sample size, number of affiliations, funding source, gender and affiliation of the corresponding author, journal impact factor, number of references, authority of the author and self-citation.

The three networks showed clearly distinguishing charactistics, which made it challenging to use the exact same method and compare the results. For example, the literature on trans fatty acids and serum cholesterol was largely influenced by one empirical publication, while the majority of the literature consisted of review articles. This network showed the strongest evidence for citation bias, with positive studies being three times more likely to be cited compared to negative studies.

In the networks on phthalates and bisphenol A many health outcomes were discussed. A strong public opinion consists on these two topics, where they are considered endocrine disruptors. Nevertheless, this opinion was not directly reflected in the scientific literature. For example, in the network on phthalates, a large proportion of the

studies did not come to a purely positive or negative conclusion. These studies, showing mixed results, can be cited both as support for a harmful and for a safe effect on human health. Unfortunately, in the current analysis, these different interpretations could not be detected. No clear evidence for the existence of citation bias was found in this network.

In the network on bisphenol A, it was remarkable that the reported results and the authors' conclusion were not always coherent. Because many publications described multiple health outcomes, they often reported both significant and non-significant findings. Nevertheless, the vast majority of these publications ultimately concluded that a harmful effect on human health exists. Evidence for citation bias was found in the network on BPA, but with a smaller magnitude of 1.5 times more chance of citation for positive studies compared to negative studies.

The three citation analyses presented in this dissertation were part of a bigger project, named the Sound Science project. Within the Sound Science project three additional citation analyses were performed. These involved the following subjects: the relationship between swimming in chlorinated water and childhood asthma, the hygiene hypothesis and the relationship between diesel emission and lung cancer in humans. Although each citation network analysis is merely a case study, by performing a total of six of these case studies we aimed to look for patterns and to distinguish between general and field-specific determinants of citation. Evidence for citation bias was found in four out of six networks, with varying magnitudes. With regard to other factors that might impact the chance of citation, we found journal impact factor, authority of the author and self-citation as general determinants of citation. Also the quality-related determinants study design and sample size showed a consistent association with citation. However, for study design the observed effect was different than expected, with empirical studies being more likely to be cited compared to review articles.

In chapter 7 we stepped outside of the scientific community and assessed selection bias in risk assessment as performed by the European Food Safety Authority (EFSA). This risk assessment functions as the basis for policy decisions by the European Commission. Basic principles of EFSA are to conduct a risk assessment in an objective, independent and transparent manner. In this context, much attention is going to financial and political conflicts of interest. In this dissertation, we address the problem of scientific independence and corresponding intellectual conflict of interest. Scientists can have such conflict of interest, because of experience in a certain field and because they interpret evidence in light of their own discipline. Selective use of evidence in risk assessment can jeopardize EFSA's core values of independence and objectivity and directly lead to decisions that are not evidence-based. To assess the objectivity within the risk assessment, we performed a case study on the risk assessment on bisphenol A. The literature on BPA identified in the earlier performed citation analysis was compared to the literature that was used in the risk assessment on BPA. Out of 36 available publications, 30 were

included in the risk assessment. No reason was given for the exclusion of the other publications, although this might have been because of their cross-sectional nature.

Overall, we can conclude from this dissertation that scientific reporting can be distorted in different ways, with more and less serious consequences. Selective use of citations has shown to be associated with a number of determinants, among which is study outcome, but also more acceptable determinants such as sample size and authority of the author. Increasing transparency, for example by publishing study protocols, and reducing citation bias will increase the validity and credibility of science-based decisions, both within and outside of academia, such as in policy-making and development of clinical therapies.

Nederlandse samenvatting

Selectieve citatie en diens gevolgen

Deze dissertatie bestudeerde het vóórkomen van vertekening in het rapporteren van wetenschappelijk onderzoek en beschreef de mogelijke invloed hiervan op kennisontwikkeling en empirisch onderbouwde besluitvorming. Wetenschappelijke publicaties zijn nogsteeds de voornaamste vorm van communicatie tussen wetenschappers in de ontwikkeling van kennis. Op zeer subtiele wijze kan er vertekening in deze communicatie optreden, welke veel invloed heeft op de ontwikkeling van kennis. Hoofdstuk 2 beschrijft een case studie over zogenaamde outcome reporting bias in de literatuur naar phthalaten. Outcome report bias refereert aan het selectief rapporteren van onderzoeksresultaten in een publicatie, waarbij ongewenste resultaten worden weggelaten. In de case studie onderzoeken we de naleving van studie protocollen en vergelijken dit met de achteraf gerapporteerde studie uitkomsten. Het gebruik van studie protocollen is van belang bij het interpreteren van studie resultaten en om het rapporteren van vals-positieve resultaten te voorkomen. Helaas hebben we gevonden dat veel onderzoeksprojecten zijn uitgevoerd zonder gebruik te maken van een vooraf opgesteld onderzoeksprotocol. Daarnaast was de bereidwilligheid van onderzoekers om hun studie protocol te delen verrassend laag. Als gevolg daarvan was het onmogelijk om de vooraf geplande analyse te vergelijken met de gerapporteerde studie resultaten en het vóórkomen van outcome reporting bias kon niet worden vastgesteld.

Naast outcome reporting bias bestudeerde deze dissertatie het vóórkomen van citatiebias. Specifiek betekent citatiebias dat de kans om geciteerd te worden samenhangt met de uitkomst van de studie. Meestal betekent dit dat positieve studies een grotere kans hebben om geciteerd te worden in vergelijking met negatieve studies. In hoofdstuk 3 zijn we gestart met een systematische review en meta-analyse om een overzicht te geven van wat er al bekend is over dit onderwerp. Hierbij hebben we geconstateerd dat citatiebias is onderzocht en aangetoond in verschillende wetenschappelijke vakgebieden. Daarnaast hebben we gevonden dat er veel verschillende methoden werden gebruikt om citatiebias te onderzoeken. Wij hebben ons tot doel gesteld een nieuwe methode te ontwikkelen, waarbij we geleerd hebben van eerdere studies. Met onze nieuwe methode hadden we niet alleen ten doel citatiebias te onderzoeken, maar ook inzicht te krijgen in andere determinanten van citatie. In hoofdstukken 4, 5 en 6 hebben we onze citatie netwerk analyse toegepast op drie verschillende wetenschappelijke velden, namelijk de relatie tussen transvetten en LDL- en HDL-cholesterol, epidemiologische studies over phthalaten en epidemiologische studies over bisphenol A. Ieder netwerk omvatte alle relevante wetenschappelijke publicaties in het betreffende veld, die geidentificeerd waren via een systematische zoekstrategie. De citatie netwerk analyse werd gebruikt om twee vraagstellingen te beantwoorden: “Is er sprake van citatiebias?” en “Welke andere determinanten beinvloeden de kans op citatie?”. Naast studie uitkomst werden de volgende determinanten onderzocht: studie design, aantal deelnemers in een onderzoek, aantal affiliaties, onderzoeksfinancier, geslacht en affiliatie

van de auteur, impact factor van het tijdschrift, aantal referenties in een publicatie, autoriteit van de auteur en zelf-citatie.

De drie onderzochte netwerken lieten duidelijke verschillen zien, waardoor het lastig was één methode toe te passen op ieder onderwerp en om de resultaten te vergelijken. In de literatuur over transvetten en cholesterol was bijvoorbeeld sterk beinvloed door één publicatie, terwijl de meerderheid van de publicaties bestond uit samenvattende publicaties. Dit netwerk liet het sterkte bewijs voor citatiebias zien, waarbij positieve studies een drie keer zo grote kans hadden om geciteerd te worden in vergelijking met negatieve studies.

In de netwerken over phthalaten en bisphenol A werden er meerdere gezondheidsuitkomsten onderzocht. Beide onderwerpen kennen een sterk publiek debat, waarbij wordt gesteld dat beide stoffen een hormoonverstorende werking hebben. Deze visie is echter niet zonder meer ondersteund in de wetenschappelijke literatuur. In de literatuur naar phthalaten, bijvoorbeeld, kwam een groot deel van de publicaties niet tot een duidelijke positieve of negatieve conclusie. Deze studies, met gemixte resultaten, kunnen zowel als schadelijk en onschadelijk voor de menselijke gezondheid aangehaald worden in een citatie. In de huidige onderzoeksmethode hebben we helaas geen rekening kunnen houden met deze verschillen in interpretatie. Er was geen duidelijk bewijs voor citatiebias in dit netwerk gevonden.

In het netwerk over bisphenol A was het vooral opvallend dat de gerapporteerde resultaten en bijbehorende auteurs' conclusie niet altijd overeen kwamen. Doordat veel publicaties meerdere gezondheidsuitkomsten beschreven bevatten de meeste studies zowel significante als niet-significante resultaten. De meerderheid van studies komt echter wel tot de conclusie dat bisphenol A schadelijk is voor de gezondheid. In het netwerk van bisphenol A werd er bewijs gevonden voor de aanwezigheid van citatiebias, waarbij positieve studies een anderhalf keer grotere kans hebben om geciteerd te worden dan negatieve studies.

De drie citatie analyses die gepresenteerd werden in dit proefschrift zijn onderdeel van een groter project, genaamd het Sound Science project. Binnen het Sound Science project zijn er nog drie andere netwerk analyses uitgevoerd. De onderwerpen hiervan zijn als volgt: de relatie tussen zwemmen in gechloreerd water en de ontwikkeling van astma bij kinderen, de hygiene hypothese en de relatie tussen dieseluitstoot en longkanker bij mensen. Hoewel ieder netwerk slecht een casus beschrijft, hebben we geprobeerd een patroon te herkennen en onderscheid te maken tussen algemene en casus-specifieke kenmerken van citatie door zes casussen te onderzoeken. In vier van de zes netwerken werd er, in verschillende mate, bewijs gevonden voor het bestaan van citatiebias. Wat betreft andere mogelijke determinanten van citatie, werden de impact factor van het tijdschrift, de autoriteit van de auteurs en zelf-citatie als algemeen geldende determinanten van citatie gevonden. Daarnaast werden kwaliteit-gerelateerde determinanten zoals het aantal deelnemers per studie consistent als determinant van citatie gevonden. Wat betreft studie design was de relatie met de kans op citatie anders

dan verwacht, hierbij was de kans op citatie groter voor empirische studies in vergelijking met samenvattende studies.

In hoofdstuk 7 zijn we uit de wetenschappelijke context gestapt en keken we naar citatiebias in de risicobeoordeling door de Europese Voedselveiligheid Authoriteit (EFSA). Deze risicobeoordeling functioneert als de basis van beleidsbeslissingen door de Europese Commissie. Basisprincipes in deze wetenschappelijke risicobeoordeling zijn dat het op een objectief, onafhankelijk en transparante manier wordt uitgevoerd. In deze context is er tot nu toe veel aandacht voor financiele en politieke belangenverstrengeling. Dit proefschrift benoemt echter het probleem van wetenschappelijke onafhankelijkheid en bijbehorende intellectuele belangenverstrengeling. Wetenschappers kunnen last hebben van deze belangenverstrengeling, vanwege hun ervaring in één bepaald vakgebied. Daardoor kunnen ze bevindingen interpreteren in het licht van hun eerdere ervaringen, in plaats van er geheel objectief naar te kijken. Selectief citeren van bewijs in de risicobeoordeling van EFSA brengt echter de basis waarden van onafhankelijkheid en objectiviteit in gevaar en leidt daarmee tot beslissingen die niet gebasseerd zijn op al het beschikbare empirische bewijs. Om de objectiviteit in de risicobeoordeling te onderzoeken hebben we een case studie gedaan naar de risicobeoordeling van EFSA naar bisphenol A. De wetenschappelijke literatuur over bisphenol A, die was verzameld in het kader van de eerder uitgevoerde citatie analyse, werd vergeleken met de literatuur die gebruikt was in de risicobeoordeling door EFSA. Van de 36 beschikbare empirische studies, waren er 30 meegenomen in de risicobeoordeling. Er werd geen reden gegeven voor het excluderen van de overige zes publicaties, hoewel dit wellicht is toe te schrijven aan het dwarsdoorsnede onderzoeksdesign.

Concluderend uit dit proefschrift kunnen we stellen dat rapportage van wetenschappelijk onderzoek op verschillende manieren vertekend kan worden, waarvan de consequenties in meer of mindere mate problematisch zijn. Selectief gebruik van citaties is gerelateerd gebleken aan een aantal verschillende factoren, waaronder studie uitkomst, maar ook meer acceptabele determinanten zoals studie design en de authoriteit van de auteurs. Vergroten van transparantie, bijvoorbeeld door het publiceren van studie protocollen, en reduceren van citatiebias zal de validiteit en betrouwbaarheid van wetenschappelijke besluitvorming vergroten, zowel binnen als buiten de wetenschap, zoals in beleid maken en ontwikkeling van klinische therapieën.

Valorisation

After finalizing the academic work, it is time to reflect on the potential valorisation of the obtained knowledge presented in this dissertation. When speaking about valorisation, what often comes to mind is how to knowledge can be translated into a competitive product or other commercial activity. For research funders especially, this is important to see that the research investment can be made into use. However, for the presented research this financial valorisation is not directly possible. The academic field studying research integrity, and more specifically studying scientific reporting, is a relatively young research field. In this early stage, the focus of the research is mostly on getting an in-depth understanding the nature and magnitude of the problem. This also makes that the presented work is a rather fundamental type of research. This does not mean it is without societeal value. It is important to get a basic understanding of how scientific knowledge develops in order to assure its value and trustworthiness. Luckily, the public trust in science is still high, especially compared to trust in mainstream media (1). However, with the growing notion of research misconduct and questionable research practices, the scientific community needs to stay active to deserve that public trust and make sure valuable knowledge is created.

This valorisation paragraph will focus on societal impact of research integrity research on various stakeholders and look at opportunities for long-term developments. The relevance of the obtained knowledge in multiple non-academic activities will be discussed, namely for policy making, development of medical treatment and product innovation. Finally, we will look at the opportunities following from this dissertation for research publishers and funders.

Policy making

Policy is created on the basis of both scientific knowledge and political vision. The process of policy making is therefore divided into the scientific process of risk assessment and the political process in the risk management phase (2). Although the ultimate decision is made in the risk management phase, this highly depends on the outcome of the scientific risk assessment. In the European Union, risk assessment and risk management are strickly divided, to enhance the legitimacy of policy decisions and to assure the independence of the scientific risk assessment (2). The latter is an interesting objective in light of this dissertation. Risk assessment is considered independent and objective because it is performed by a panel of academics, who are experts on the topic under discussion. Attention is paid to the composition of the panel, to make sure all relevant disciplines are represented and panel members have sufficient knowledge about the subject. Additionally, panel members are screened for ties with industry and political involvement, on which basis they will be excluded. However, no attention is paid to the limitations of the scientific evidence that is being used. In chapter 7 of this dissertation, the concept of intellectual conflict of interest was discussed. The carreer of individual

scientists might impact the evidence that is being put forward or how this is weighted in the risk assessment. But also in a wider context, the discussion on questionable research practice and research misconduct that is currently taking place in the academic arena, is also applicable to the way science is being evaluated in risk assessments. A wide range of questionable research practice can impact the validity of the risk assessment. This includes problems with regard to the reporting of research, but also the use of inappropriate research designs or errors in the statistical analysis, which often occur in scientific publications (3). When scientists are not aware of the existence and magnitude of these questionable research practices, they will not be taken into account in the weighing of the evidence. Creating awareness for problems relating to scientific reporting, such as publication bias, reporting bias and citation bias is therefore an important first step to assure the validity of evidence-based policy making. Apart from creating awareness with the scientists performing the risk assessment, in the longer-term concrete actions should be implemented to improve the credibility and quality of evidence-based policy. Examples of these concrete actions might be the use of systematic search strategy as the basis for their risk assessment. In this way, the risk of citation bias can be reduced. When evaluating the quality of the presented evidence, a checklist might be used to check for the most common questionable research practices.

Medical treatment

Much scientific research revolves around the development of medical treatments. A strong evidence base needs to be build when developing new medical treatments and to get them accepted as the standard treatment. This includes a wide range of study designs starting with mechanistic studies, animal studies, human observational studies and potentially even randomised controlled trails. In each of these levels, knowledge might get to waste because of selective reporting of results and selective citation.

A very illustrative example of how clinical practice can be impacted by selective citations is the work of Andrade et al (2013) (4). They performed a citation analysis on the literature on treatment options for chronic nonspecific low back pain. This literature base consisted of two types of randomised controlled trials: RCTs that compared surgical treatment with non-surgical treatment and RCTs that compared two surgical treatments with each other. The RCTs comparing two surgical treatments far outnumbered the RCTs involving non-surgical treatment, showing that the research agenda was focused on finding the optimal surgical treatment. However, studying the content of all the RCTs, it appeared that no convincing evidence exists for chosing surgery over non-surgical treatment. This is a clear example of how selective citation, by not citing the RCTs including non-surgical treatments, can drive the research agenda into a certain direction that is not evidence-based. Consequently, much research money and time have been invested in unnecessary RCTs that compared two surgical treatments. Even

more important, patients have unnecessarily undergone surgery where other treatments would have been sufficient and actually better. Therefore, also in the development of medical drugs and devices, it is important to have a complete overview of all available publications. In this way, it can be determined if a research question is still relevant and research waste can be reduced.

Innovation

Next to drug development, scientific research functions as the basis for all kinds of innovations. This could include innovations in light of medical equipment, but also innovations to generate sustainable energy or innovations in the financial market. Although these innovations do not depend necessarily on *scientific* evidence, a lot of research is required before a successfull innovation can go to the market. This involves high financial investments as well as investments in terms of time and effort. Most likely, the process of product innovation is one of trial and error. Similar to academia, also product developers tend selectively focus on the success findings, while not reporting the failures (5). Also similar to academia, much can be learned from these failures and future failures and associated investments could be prevented. Additionally, by selectively reporting only successful innovations the unjustified image might occur that all innovations are successful and all investments in innovations are worthwile. By being more transparant in reporting both successes and failures the process of innovation can be made more efficient.

Research publishing and funding

Also within the academic arena, the obtained results are relevant for research funders and publishers. Publishing books and articles remain the core activity in the scientific enterprise and is the foundation for development of knowledge. Because publications are the main communication form among scientists, a great responsibility lies with the academic publishers, in facilitating this in an integer way. This is even more important, given that a high number of academics are actively competing for limited research grants. This high competition might lead to cutting corners when it comes to doing high quality research and a perverse incentive is created to publish research including questionable research practices. The chance of receiving a grant is still highly depending on traditional metrics, such as the number of publications, publishing in high impact factor journals and getting high numbers of citations. Unintentionally, this promotes salami-slicing of publications and self-citation. In this dissertation, we have empirically shown that publishing significant findings increases the chance of being cited in most fields, which brings a competitive advantage when it comes to obtaining research grants. As

an even bigger problem, research quality is not taken into account when evaluating a researchers performance. Potentially this explains also the finding in chapter 2 of the dissertation, saying that much research is carried out without a study protocol. Writing a protocol is a time consuming activity, without any guarantee that the work will be published or other form of reward. Here, we could see a role for publishers and editors. By requesting authors to upload a study protocol together with the manuschript and taking this into account in the peer-review process, authors will be encouraged to set a priori hypothesis and work according to a protocol. Subsequently, studies that were not performed on the basis of a study protocol might be notified as hypothesis generating studies instead of hypothesis testing studies. This is an important distinction to make in order to correctly interpret the research findings.

On a positive note, there is already a growing protest with regard to the way scientific output is measured and evaluated in the current system. A number of editors and publishers, but also governments, are looking for more quality-related measures of research output. Their motivitation to do so is shown for example by signing the DORA initiative, which was signed by the Dutch NWO and several editors of biomedical journals (6). Also in editorials, editors express their concern with regard to the current research climate and the occurrence of questionable research practice (7-9). Concluding, we could say that the intention of research publishers and funders to exclude research misconduct and questionable research practives seems to be positive. However, more research is needed to show the magnitude of the problems in the current system and to find suitable replacements.

References

1. Verma N, Fleischmann KR, Koltai KS. Human values and trust in scientific journals, the mainstream media and fake news. Proceedings of the Association for Information Science and Technology. 2017;54(1):426-35.
2. König A, Kuiper HA, Marvin HJ, Boon PE, Busk L, Cnudde F, et al. The SAFE FOODS framework for improved risk analysis of foods. Food Control. 2010;21(12):1566-87.
3. Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. Biochemia medica: Biochemia medica. 2015;25(1):5-11.
4. Andrade NS, Flynn JP, Bartanusz V. Twenty-year perspective of randomized controlled trials for surgery of chronic nonspecific low back pain: citation bias and tangential knowledge. The Spine Journal. 2013;13(11):1698-704.
5. Cannon MD, Edmondson AC. Failing to learn and learning to fail (intelligently): How great organizations put failure to work to innovate and improve. Long range planning. 2005;38(3):299-319.
6. Cagan R. The San Francisco declaration on research assessment. The Company of Biologists Ltd; 2013.
7. Knottnerus JA, Tugwell P. Research without good questions is a waste. Journal of clinical epidemiology. 2019;108:vi-viii.
8. Kleinert S, Horton R. How should medical science change? The Lancet. 2014;383(9913):197-8.
9. Glasziou P. The role of open access in reducing waste in medical research. Public Library of Science; 2014.

Acknowledgments

Nu het meeste werk gedaan is, is het tijd om terug te blikken op de afgelopen vier jaar. Hoewel de tijd voorbij is gevlogen, zijn er veel dingen gebeurd en hebben er veel mensen een bijdrage geleverd aan het tot stand komen van het proefschrift. Een aantal daarvan zou ik graag expliciet bedanken in dit dankwoord.

Allereerst wil ik graag mijn begeleiders bedanken voor de mogelijkheid om dit proefschrift tot stand te laten komen, hun uitvoerige hulp en vertrouwen tijdens het hele project. Het citatie onderzoek was voor ons allemaal een beetje nieuw terrein, ik ben blij dat we het samen hebben kunnen ontdekken.

Maurice, dank voor je prettige begeleiding, en het vertrouwen in de goede afloop. Daarnaast ook speciaal bedankt voor je vertrouwen in mij als academicus en de bijbehorende carriere adviezen. Ik kijk ernaar uit om onze samenwerking de komende jaren te kunnen voortzetten! Lex, hartelijk dank voor je grote betrokkenheid bij het project. Je was altijd bereikbaar voor snel en uitvoerig commentaar, dat heb ik zeer gewaardeerd. Naast leuke inhoudelijke discussies heb ik ook veel van je mogen leren wat betreft je visie op de wetenschap in het algemeen. Hopelijk komen we elkaar nog vaak tegen in de wereld van de wetenschappelijke integriteit! Gerard, dank voor je inzet in mijn begeleiding. Ik kon altijd bij je terecht met vragen en denk met veel plezier terug aan alle uitvoerige gesprekken die we gevoerd hebben, over de wetenschap en andere zaken. Veel geluk in de toekomst samen met Ada in Frankrijk!

I would like to thank the members of the assessment committee, Prof. David Townend, Prof. Andre Knottnerus, Prof. Aalt Bast, Dr. Gerben ter Riet and Dr. Joeri Tijdink, for their critical assessment of the thesis.

Thanks also to CEFIC and the academic board for funding the Sound Science project and giving us all the academic freedom to perform the research. I have enjoyed our annual discussions and your interest in the subject of research integrity.

Bram, bedankt voor de prettige samenwerking de afgelopen jaren. Hoewel we even aan elkaar moesten wennen in het begin, ben ik blij dat we samen het hele promotietraject hebben kunnen doorlopen en tot een goed einde hebben gebracht. Veel succes voor de toekomst!

Natuurlijk wil ik ook graag de andere comgen (oud)collega's bedanken, in het speciaal Kelly, Nikos, Frits, Sylvia, Anke, Marij en Hanneke. Toen ik in 2014 begon, zaten we nog samen in één grote kamer en hoewel dit niet altijd even productief was, denk ik met veel plezier terug aan deze leuke tijd. Dank ook voor al het advies dat ik van jullie heb gekregen, zowel over onderzoek als onderwijs. Also the more new colleagues I would like to thank. Nadia, Elena, Anna-Roos, Evan, Magda, Putri thank you very much for all the good times we had and good luck with your PhD's!

Erika, voor jou een bijzonder woord van dank. Ik ben blij dat ik altijd met alle vragen bij je terecht kan, maar ook voor een gezellig praatje sta je altijd klaar.

Naast mijn collega's bij complexe genetica wil ik ook graag de collega's van de **afdeling Epidemiologie** bedanken voor de mogelijkheid om na het afronden van mijn proefschrift, deel uit te maken van de afdeling. Bedankt voor al het advies over onderwijs, maar ook voor de gezellige praatjes op de gang en de vrijdagmiddagborrels.

Prof. John Ioannidis and **Prof. Steven Goodman**, thank you very much for the opportunity to visit METRICS at the beginning of 2018. It has been a great learning experience to be able to participate in your discussions on meta-research. Thank you very much for all your personal advice, it has been very helpful in finding my way in the field of meta-research.

Ellen, ik ben blij dat we iedere keer weer een nieuwe manier vinden om onze samenwerking voort te kunnen zetten. Ik wil je graag bedanken voor alle begeleiding die ik van je heb gekregen, zowel vakinhoudelijk als advies over carriere planning. Ik hoop dat ik nog lang van je mag leren!

Alie, ook jou wil ik graag bedanken. Allereerst voor de kans dat ik als masterstudent mocht meeschrijven aan de artikelen van jouw PhD. Ik kijk met plezier terug op onze samenwerking en de daaruit volgende gesprekken over een (vrouwelijke) toekomst in de wetenschap. Hopelijk kunnen we dit nog lang blijven doen!

Ook mijn nieuwe collega's bij **Studio Europa** wil ik graag bedanken voor jullie interesse in mijn onderzoek, de mogelijkheid om het voort te zetten en de prettige werkomgeving. Ik kijk uit naar onze samenwerking de komende jaren!

Ook al kost het werken aan een promotie veel tijd, gelukkig was er ook regelmatig tijd voor ontspanning. Daarvoor wil ik graag mijn vrienden en vriendinnen bedanken. **Vriendinnen van Stella Maris, de eetclub, Susan en Madelon, oud huisgenoten, vriendinnen uit Venlo**: in de afgelopen jaren zijn er behoorlijk wat vriendengroepjes gevormd. Ik kijk met heel veel plezier terug op alle leuke feestjes, etentjes, weekendjes weg en kopjes thee die we samen hebben meegemaakt! Hoewel we elkaar niet meer allemaal even vaak zien, door soms wel hele grote afstand, ben ik heel blij dat we nogsteeds contact hebben en de moeite nemen om elkaar op te zoeken!

En als laatste natuurlijk dank aan de familie. **Judith**, leuk dat we na bijna vijf jaar nu ongeveer samen gaan promoveren. Bedankt voor je interesse in mijn werk en voor alle leuke dingen die we samen doen. Hopelijk wordt dat nog meer nu de proefschriften in de kast staan. **Papa en mama**, hoewel het na vier jaar nog steeds niet goed kan uitleggen wat ik eigenlijk doe, luisteren jullie toch altijd geinteresseerd. Bedankt dat ik altijd terecht kan met mijn verhaal en voor jullie praktische hulp, waardoor ik me volledig op het werk kon concentreren.

Curriculum vitae

Miriam Urlings was born on 21 September 1991 in Heerlen, the Netherlands. She started her bachelor in Health Sciences at Maastricht University in 2009. Within this bachelor she graduated in the tracks on bioregulation and health education and promotion. During this bachelor, she developed a great interest in nutrition, and therefore continued her education in the master programme Health Food Innovation Management, also at Maastricht University, with the idea to go into industry. This education led to broad knowledge on the interaction between food and health, including human physiology, consumer behavior and policy development. Unexpectedly, she developed a special interest in food law. She graduated with a thesis on 'the role of science in the authorization procedure of genetically modified foods', under supervision of Prof. Ellen Vos. Additional to writing her thesis, she was active as an intern for the FP7-project INPROFOOD. In this consortium 17 European research institutes worked to map activities in the field of food innovation around Europe, to reduce the obesity problem. Miriam performed a stakeholder analysis on food innovation and research in the Netherlands and was responsible for the communication of project results. Her interest in the validity and complexity of scientific research, as well as the idea of working in academia, emerged during this period.

In November 2014, Miriam started her PhD where she focused on research integrity and the occurrence of questionable research practices. Specifically, she studied determinants of selective citation in various biomedical research fields, by means of case studies. At the World Conference on Research Integrity in Amsterdam in 2017, she was awarded the Excellence in Doctoral Research Award for her presentation on the citation bias methodology. In 2018 she paid a three-month visit to METRICS, the Meta-Research Innovation Centre at Stanford University in the United States. During this period she discussed her research with prof. Ioannidis and prof. Goodman, who are authorities on the field of meta-research. Additional to her experience as a researcher, she has gained ample experience as a teacher, mostly in the field of epidemiology, research methodology and critical assessment of scientific literature. During the end of her PhD, Miriam got a position as a teacher at the department of epidemiology and as a project coordinator at the Horizon 2020 project RECIPES. The RECIPES project discusses the relationship and interaction between the precautionary principle and innovation. Miriam was responsible for coordinating tasks, such as getting ethical approval, writing a data management plan and consortium agreement and maintain contact with project partners and the European Commission.

Currently she is appointed as postdoc researcher in the project Maastricht Working on Europe, at Maastricht University. Within this project, she will continue to work on the functioning of the scientific enterprise and specifically the translation of scientific evidence into policy advice by European agencies.

List of publications

Publications following from this thesis

Swaen, GMH, **Urlings, MJE**, & Zeegers, MP. (2016). Outcome reporting bias in observational epidemiology studies on phthalates. *Annals of epidemiology. 26*(8), 597-599. e594.

Duyx, B, **Urlings, MJE**, Swaen, GMH, Bouter, LM, & Zeegers, MP. (2017). Scientific Citations Favor Positive Results: A Systematic Review and Meta-analysis. *Journal of Clinical Epidemiology.* 88, 92-101.

Urlings, MJE, Duyx, B, Swaen, GMH, Bouter, LM, & Zeegers, MP. (2019) Citation bias in the literature on dietary trans fatty acids and serum cholesterol. *Journal of clinical epidemiology.* 106, 88-97

Urlings, MJE, Duyx, B, Swaen, GMH, Bouter, LM, & Zeegers, M.P. (2019) Selective citation in scientific literature on the human health effects of bisphenol A. *Research Integrity and Peer Review*, *4*(1), 6.

Duyx, B, **Urlings, MJE**, Swaen, GMH, Bouter, LM, & Zeegers, MP. (2017). Selective citation in the literature on swimming in chlorinated water and childhood asthma: a network analysis. *Research Integrity and Peer Review*. *2*(1), 17.

Other publications

De Boer, A, **Urlings, M,** Vos, E, & Bast, A. (2015). Enforcement of the Nutrition and Health Claim regulation. *European Food & Feed Law Review*, 334.

Clevers, E, & **Urlings, M**. (2015). From Dietary Guidelines to Dietary Guidance. *European Journal of Nutrition and Food Safety*, 5, 250-259

De Boer, A, **Urlings, MJE**, & Bast, A. (2016). Active ingredients leading in health claims on functional foods. *Journal of Functional Foods, 20*, 587-593.

Oral presentations conducted during the PhD

Plastics Europe (invited speaker) Research integrity and selective citation. October 2015. Brussels, Europe

European Risk Forum (invited speaker and discussion panel member). Selective citation in science based decision making. December 2015. Brussels, Belgium

Department of Toxicology, Maastricht University (invited speaker). Selective citation in science-based decision making. December 2015. Maastricht, The Netherlands

Netherlands Research Integrity Network. Which factors drive citation? A citation network analysis on trans fatty acid literature. May 2016. Amsterdam, The Netherlands

Eatwell project annual meeting (Maastricht University). Selective Citation in trans fatty acid literature. December 2016. Maastricht, The Netherlands

Doctoral Forum 5th World Conference on Research Integrity. Sound Science: Selective citation in science based decision-making. May 2017. Amsterdam, The Netherlands

5th World Conference on Research Integrity. Selective citation in trans fatty acid literature. May 2017. Amsterdam, The Netherlands

6th World Conference on Research Integrity. Selective citation in biomedical sciences: an overview of six research fields. June 2019. Hong Kong, China