

BMJ Open Selective citation in the literature on the hygiene hypothesis: a citation analysis on the association between infections and rhinitis

Bram Duyx,¹ Miriam J E Urlings,¹ Gerard M H Swaen,¹ Lex M Bouter,^{2,3} Maurice P Zeegers^{1,4}

To cite: Duyx B, Urlings MJE, Swaen GMH, *et al.* Selective citation in the literature on the hygiene hypothesis: a citation analysis on the association between infections and rhinitis. *BMJ Open* 2019;**9**:e026518. doi:10.1136/bmjopen-2018-026518

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-026518>).

Received 7 September 2018
Revised 26 November 2018
Accepted 28 November 2018

ABSTRACT

Objective Our objective was to assess the occurrence and determinants of selective citation in scientific publications on Strachan's original hygiene hypothesis. His hypothesis states that lack of exposure to infections in early childhood increases the risk of rhinitis.

Setting Web of Science Core Collection.

Participants We identified 110 publications in this network, consisting of 5551 potential citations.

Primary and secondary outcome measures Whether a citation occurs or not, measured and analysed according to the preregistered protocol.

Results We found evidence for citation bias in this field: publications supportive of the hypothesis were cited more often than non-supportive publications (OR adjusted for study design [adjOR] 2.2, 95% CI 1.6 to 3.1), and the same was the case for publications with mixed findings (adjOR 3.1, 95% CI 2.2 to 4.5). Other relevant determinants for citation were type of exposure, specificity, journal impact factor, authority and self-citation. Surprisingly, prospective cohort studies were cited less often than other empirical studies.

Conclusions There is clear evidence for selective citation in this research field, and particularly for citation bias.

BACKGROUND

The hygiene hypothesis postulates that a high degree of hygiene in early life will increase the risk of developing allergies later in life.^{1,2} The underlying mechanism has been the topic of scientific debate. Over time, this debate led to several adaptations and extensions of the hygiene hypothesis, which, as such, provides a good example of how science progresses. Ideally, this progress should be based on all existing evidence, but this is not always the case.³ A citation analysis can help to reveal which part of the available evidence is taken into account, and which evidence is ignored. The current study does not concern the validity of the hygiene hypothesis per se, but rather the citation relations within the scientific literature on this hypothesis.

Strengths and limitations of this study

- The study assesses how evidence regarding the hygiene hypothesis propagates over time by analysing the likelihood of citation.
- It investigates which characteristics of a publication—such as study outcome, journal impact factor, author gender and affiliation, and authority within the field—have an impact on citation.
- We check whether supportive studies are cited more often by other studies within the field.
- Only articles related to the original hygiene hypothesis are included in this analysis.

The hygiene hypothesis was originally proposed to explain the rising prevalence of allergies, with up to 20%–40% of the population in developed countries being affected.⁴ Modern, urbanised life in developed countries generally shows higher levels of hygiene than in previous times or in low/middle-income countries. Hygiene limits exposure to infections. Exposure to infections, especially early in life, helps to develop and adapt the immune system to the environment in which we happen to live, in such a way that it learns to discriminate between harmless and harmful intruders. According to the hygiene hypothesis, it is this lack of exposure to relatively harmless intruders early in life that causes the immune system to malfunction later in life, hence the rise in allergies.

The hygiene hypothesis has been amended several times since its early days to give rise to newer theories such as the '*old friends hypothesis*'.^{5,6} According to this theory, it is not hygiene per se that is causing the rise in allergy prevalence, but the lack of exposure to *some specific* infections, and also to the gut microbiome and to non-viable intruders from the natural environment, such as endotoxins. Humans have been exposed to these



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Mr Bram Duyx;
b.duyx@maastrichtuniversity.nl

'old friends' for many centuries and our immune system has co-evolved in their presence. As a result, our immune system has become dependent on the presence of these old friends in order to develop and function properly. This adapted hygiene hypothesis states that lack of exposure to these old friends may give rise not only to allergies, but to autoimmune diseases as well.

The original hygiene hypothesis and its later adaptations have a lot in common, and much of the evidence that is supportive for one hypothesis is equally supportive for the others. However, this is not always the case. In our project, publications are classified as either supportive or non-supportive with regard to a hypothesis. For that reason it is important to precisely define the investigated hypothesis. In our citation network, we focus on the hygiene hypothesis as it was originally stated by Strachan, and not on later modifications.^{1 2} This allows us to investigate the development of this hypothesis from the start. Concretely, this means that we focus on the impact of infections and the number of siblings on the development of rhinitis, like in Strachan's original study.²

The number of publications in the research on the hygiene hypothesis is large. It is therefore not feasible for authors to cite every relevant publication in the network and some kind of selection needs to take place. If this selection is based on study outcome, we speak of citation bias.^{3 7} The consequences of citation bias can be similar to those of publication bias and reporting bias: disregard of counterevidence leading to unfounded consensus⁸ or polarisation,⁹ ill-advised research programmes and research waste,^{8 10} distorted information in the media¹¹ and misguided medical decisions.¹² Citation bias has been studied in many disciplines. Our systematic review gives an overview of these studies.¹³ Many of these studies showed evidence for citation bias in their field, with supportive publications being cited about twice as often as non-supportive ones.

Factors other than study outcome may also have an impact on citation, as was recently shown by Onodera and Yoshikane.¹⁴ Measures for journal status (impact factor), author status (number of citations, country of affiliation) and collaboration (number of authors, number of affiliations) were often found to be related to citation count. The same was consistently found for the number of references of the cited publication. Furthermore, the reporting¹⁵ and source^{16 17} of funding were shown to be related to citation, but the impact of author's affiliation¹⁸ and gender¹⁹⁻²¹ is less clear. On the other hand, sample size and study design—both markers of study quality, and as such legitimate reasons to base a citation on—often seem unrelated to citation.^{17 18 22-24} In our previous citation networks, we also found associations with self-citation and the specificity of a publication, but not with the title of a publication.^{25 26}

In our study, we aimed (1) to assess the occurrence of citation bias in the scientific literature on the original hygiene hypothesis and (2) test for other signs of selective citation by assessing the impact of the other factors

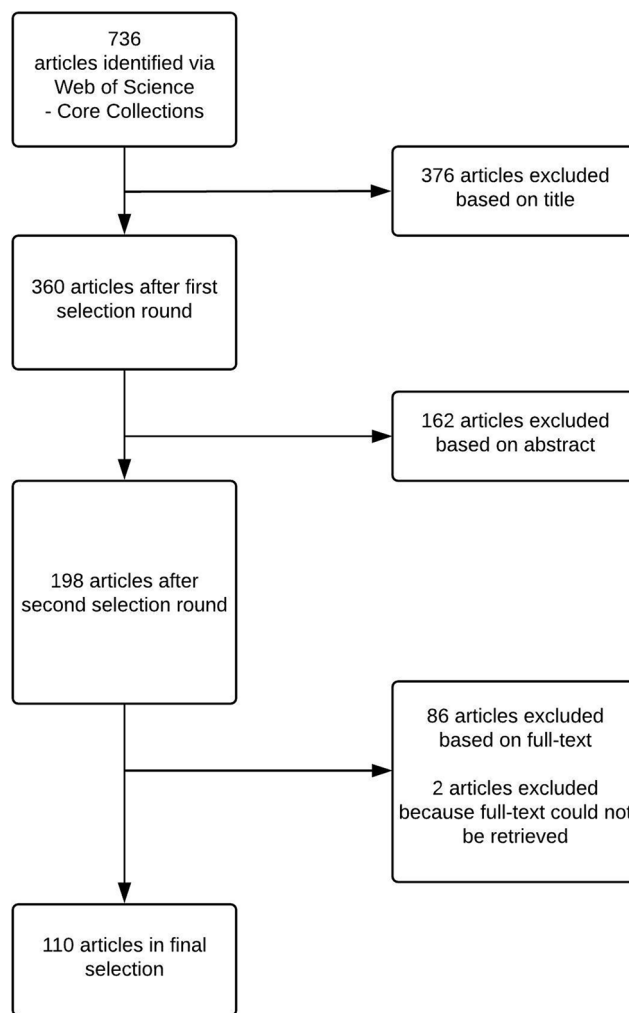


Figure 1 Flow diagram of article selection process.

described above. We will make use of the claim-specific methodology developed by Greenberg,⁸ but with a modification of the statistical analysis that allows us to test the impact of multiple factors, adjust for study design and take into account the variation in publication time.

METHOD

Prior to performing the citation network analysis, we described our methods in a study protocol and stored it at an online repository.²⁷ (Protocol deviations are described in the online supplementary text S1.) In brief, we applied a search strategy to the Web of Science Core Collection (WoSCC), identified relevant literature, downloaded these records with their reference lists, extracted data for each publication, built a dataset with potential citation paths and used specialised software to determine which citations had occurred. These steps will be explained in more detail below. Article selection (first based on title, then on abstract and finally full text; [figure 1](#)) and data extraction were performed independently by MJEU and BD. Results were compared after each step, and disagreements were resolved in consensus meetings.

For clarification: a publication in our network can both cite and be cited by other publications in the network, leading to a multitude of citation paths. Not all citation paths are possible as one can only cite articles that were published before. In our study, a citation is considered possible if the cited publication is published before the citing publication is submitted. If such *potential* citation occurred, we call it an *actual* citation (see also online supplementary text S2.)

Search strategy

First, we took Strachan's seminal article in which the hygiene hypothesis was launched as point of departure.² Next, we identified all literature within WoSCC referring to this article. Finally, we limited the output to publications that mentioned hay fever in their title, keywords or abstract ('*hay fever*' OR '*hayfever*' OR '*hay-fever*' OR '*rhinitis*' OR '*rhino**'). The search was performed by BD and updated until 16 August 2017. Only English language publications were included.

The search output was then limited to publications that investigated exposures related to the original hygiene hypothesis. This means that only publications investigating the effect of *number of siblings* and *infection history* were included. Publications on helminth infections were excluded, as different versions of the hygiene hypothesis would make contradictory predictions regarding their impact on allergies. Both empirical and non-empirical publications were included.

Data extraction

A range of characteristics were extracted or derived from each included publication. These characteristics are described below and were all tested as determinant of citation in the statistical analysis.

Publication characteristics: content related

The following variables were in this subcategory: type of exposure, publication type, sample size, specificity and study outcome.

Type of exposure refers to the type of exposure that is being studied or reviewed: only number of siblings, only infection history or both.

Publication type was classified into empirical and non-empirical publications. Empirical publications were further classified into the following *study designs*: cross-sectional, case-control, retrospective cohort and prospective cohort. Non-empirical publications were further classified into: narrative reviews, systematic reviews and other (editorials, leading articles, commentaries).

Sample size concerned the number of participants in the publications. Non-empirical publications had no sample size. The sample size of the empirical publications was classified into three equal categories based on tertiles.

The *specificity* of the publications varied. Some publications only deal with Strachan's hygiene hypothesis, others are broader. Specificity ranges from 1 (very broad) to 3 (very specific). For instance, an empirical publication

that only investigates the association between number of siblings and rhinitis would be classified as '3'; if it also investigates the impact of helminth infections and growing up on a farm, and if it also includes other health outcomes such as asthma or autoimmune diseases, it would be classified as '1'.

Study outcome was scored as follows: (1) supportive of the hygiene hypothesis; (2) mixed or unclear results; (3) non-supportive of the hygiene hypothesis. An inverse relationship between past exposure and rhinitis is considered to be supportive for the hygiene hypothesis, while a neutral or positive relationship was scored as non-supportive. The scoring was based on the authors' interpretation of the results, as it was stated in the text of the publication (see also online supplementary text S2 for more details).

Publication characteristics: not content related

The following variables were in this category: conclusiveness of the title, funding source, number of authors, number of affiliations and number of references. *Title conclusiveness* was coded as yes if in the title a conclusion was stated that included the direction of the relationship (eg, 'inverse relationship between infections and allergies'), otherwise as no (eg, 'infections, rhinitis and their relationship'). *Funding source* was coded as non-profit (eg, government or university), for-profit, both or not reported.

Journal characteristics

The following variables were in this category: publisher and journal impact factor. *Journal impact factor*, in the publication year of the cited publication, was retrieved from the Journal Citation Reports (JCR) database. *Journal publisher* was also retrieved from JCR.

Author characteristics

The following variables were in this category: *gender* of the corresponding author (see also online supplementary text S2), *continent* of the corresponding author and affiliation of the corresponding author. *Affiliation* was classified as government, university, industry or other.

Citation characteristics

There were some variables that depend on the cited publication as well as the citing publication: self-citation and within-network authority. A *self-citation* was defined as a citation between two publications that have at least one author in common.

Authority was a measure for the authority of the authors within the network. It was calculated for each author and each year separately, by counting the number of within-network citations to all publications in which the author had been involved. As the number of citations is likely to increase each year, so does the author's authority. Because we were interested in the authority at the moment of citation, the authority value of a cited publication also depends on the publication year of the citing publication. In case of multiple authors, we used the authority value of the author with the highest authority in that year.

Statistical analysis

The dataset consisted of all potential citation paths between cited and citing publications. A potential citation path means that the cited publication is published before submission of the citing publication. The underlying assumption is that publications can only cite other publications up to the date of submission of the citing publication, and that publications can only be cited from their publication date onwards. All analyses were preregistered in the study protocol unless mentioned otherwise.

Impact of characteristics in the cited publication

Our binary dependent variable was citation within the network (or, more precisely, whether a potential citation had occurred or not). This was determined by the built-in algorithm of CitNetExplorer.²⁸ This algorithm makes use of reference lists that can be downloaded from the WoSCC. It links the reference lists of all publications in the network with the actual publications in the network. If possible, this linkage was done by digital object identifier (DOI), that is assigned to most present-day publications; otherwise it was based on a combination of first author's surname, first author's first initial, publication year, volume number and first page number. The determinants of citation in our analyses were the characteristics of the cited publication as described above.

Since each publication could refer to multiple other publications, the potential citation paths were related. Therefore, we used a multilevel approach in which the potential citations were nested under the citing publication. Specifically, we performed a univariate random-effects logistic regression for each determinant of citation. We repeated these analyses while adjusting for study design, as a proxy for study quality. Another proxy for study quality would be the study sample size. However, as reviews do not have a sample size, this adjusted analysis could only be performed on the subselection of cited empirical publications so we did not adjust for sample size in the main analysis.

In addition to the original analysis plan in the protocol, we also calculated the explained variance of the adjusted models, so that these models are easier to compare. For this purpose we calculated McFadden's R^2 .

Additional analyses were performed on subselections of the network: (1) only cited empirical publications were included (to investigate which empirical evidence is picked up by the rest of the field; explorative analysis); (2) only cited empirical publications and citing synthesis publications were included (to investigate which empirical evidence is picked up particularly by reviews and editorials). These analyses were adjusted for study design and for log-transformed sample size because in these subselections all cited publications had a sample size.

To check the robustness of our findings, we also ran some sensitivity analyses in which the following publications or citation paths were excluded: (3) the most cited publications (explorative analysis); (4) citation paths with less than 1 year between publication date of the cited

publication and submission date of the citing publication were excluded (to check if a lag time would make a difference as it takes some time before most publications are known and have an impact); (5) citing publications that have less than 10 potential citations.

Impact of concordance

Where applicable, we also calculated whether the cited and the citing publications had the same characteristics (*concordance*). This would, for instance, be the case if supportive publications would prefer to cite other supportive publications, and if non-supportive publications would prefer to cite other non-supportive publications. If citation would be based on the concordance of study outcome, it would be another measure of citation bias. To test if concordance on several characteristics has an impact on the likelihood of citation, univariate and adjusted (for study design) fixed-effects logistic regression analyses were applied.

Software

We used the built-in algorithm of CitNetExplorer v1.0.0 to extract the actual citations between publications.²⁸ We used R v3.2.4 to create a dataset with all potential citation paths, based on the data extraction sheet and the actual citations, and also to calculate the within-network authority and self-citation score for each potential citation path. Finally, we used Stata v13.1 to analyse the results.

Patient and public involvement

No patients were involved in this study.

RESULTS

A total number of 110 publications were identified that fit our criteria, published between 1995 and 2017 (figure 1, online supplementary text S3). Of these, 28 publications focused exclusively on the impact of household size on rhinitis, 48 on the impact of having had infections and 34 on the impact of both types of exposure. This network of 110 publications had a total of 5551 potential and 392 actual citation paths (7%) between these publications. Their main characteristics are depicted in table 1 (for more details see online supplementary table S1). About two-thirds of all publications in the network are empirical studies (39 cross-sectional, 4 case-control and 30 cohort studies), one-third are reviews (27 narrative reviews, 2 systematic reviews and 8 editorials or leading articles). The study outcome for 35 of the publications was mixed or unclear. Of the remaining publications with a clear study outcome, about 50% was supportive of the hygiene hypothesis (41 publications with an inverse association between siblings/infection and rhinitis) and about 50% was non-supportive (34 publications with no association or with a positive association). The number of citations ranged from 0 (45 publications) to 35, with a median of 1 citation per publication. A ranking of the most cited

Table 1 Main characteristics of all 110 publications in hygiene hypothesis network

Characteristic	Category	N publications	n potential citations	n actual citations (%)
Type of exposure	Only no of siblings	28	1512	100 (7)
	Only infection history	48	1946	144 (7)
	Both siblings and infections	34	2093	148 (7)
Study outcome exposure—rhinitis	Supportive	41	2322	198 (9)
	Mixed results	35	1913	129 (7)
	Non-supportive	34	1316	65 (5)
Publication type/ <i>study design</i>	Empirical	73	3517	337 (10)
	<i>cross-sectional</i>	39	1697	179 (11)
	<i>case-control</i>	4	249	36 (14)
	<i>retrospective cohort</i>	15	817	89 (11)
	<i>prospective cohort*</i>	15	754	33 (4)
	Synthesis	37	2034	55 (3)
	<i>narrative review</i>	27	1423	16 (1)
	<i>systematic review</i>	2	80	20 (25)
	<i>editorial</i>	8	531	19 (4)
Sample size (cat; for empirical publications)	Low (1–999)	24	909	56 (6)
	Medium (1000–7999)	25	1327	143 (11)
	High (≥8000)	24	1281	138 (11)
Journal impact factor (cat)	0–2	28	1275	27 (2)
	2–4	41	2087	145 (7)
	≥4	32	1671	176 (11)
Gender	Male	65	3368	265 (8)
	Female	42	2024	123 (6)
	Unclear	3	159	4 (3)
Affiliation	University	88	4402	258 (6)
	Government	9	410	22 (5)
	Industry/other	13	739	112 (15)
Continent	Europe	62	3885	324 (8)
	North America	19	688	38 (6)
	Asia	21	502	9 (2)
	Australia	8	476	21 (4)
Total		110	5551	392 (7)

*Including 1 experimental study.

publications and authors can be found in online supplementary table S2.

Impact of characteristics in the cited publication

The results of the regression analyses are presented in table 2. Empirical publications were cited more often than non-empirical publications. Compared with empirical studies with a cross-sectional design, prospective cohort studies, narrative reviews and editorials had a lower likelihood of citation, while the two systematic reviews had a higher likelihood of citation. Other determinants that increased the likelihood of citation were specificity, journal impact factor, sample size and within-network authority. Sample size had a modest impact on citation.

Publications on only one type of exposure were cited less often than publications on both types of exposure.

Supportive publications had a higher likelihood of being cited than non-supportive publications. This is in line with our hypothesis. However, publications with mixed results were cited even more often. This may be due to our scoring algorithm. After all, if a publication investigated both the number of siblings and the infection history, and it reported dissimilar outcomes for these two exposures, then this publication would have been scored as having mixed results. An explorative χ^2 test confirmed that type of exposure and study outcome were related ($\chi^2(4)=52$, $p<0.0005$), with 71% of all publications on both types of exposure reporting

Table 2 ORs (95% CIs) for the chance of being cited, all types of publications included, N=110, n=5551

	Crude OR	Adjusted OR*	R ² *
Publication characteristics, content related			
Type of exposure (ref: both siblings and infections)			0.10
Only no of siblings	0.8 (0.6 to 1.1)	0.5 (0.4 to 0.7)	
Only infection history	1.3 (1.0 to 1.6)	0.8 (0.6 to 1.1)	
Study outcome (ref: non-supportive results)			0.11
Mixed/unclear results	1.2 (0.9 to 1.7)	3.1 (2.2 to 4.5)	
Supportive results	1.5 (1.1 to 2.0)	2.2 (1.6 to 3.1)	
Publication type (empirical vs synthesis)	4.3 (3.2 to 5.7)		0.04 (crude)
Study design (ref: cross-sectional)			0.09 (crude)
Case-control	1.4 (0.9 to 2.2)		
Retrospective cohort	0.9 (0.7 to 1.2)		
Prospective cohort	0.3 (0.2 to 0.5)		
Narrative review	0.1 (0.0 to 0.1)		
Systematic review	3.3 (1.8 to 5.8)		
Editorial/other	0.3 (0.2 to 0.4)		
Sample size (ref: low, n=3517)			0.02
Medium	1.6 (1.2 to 2.3)	1.6 (1.0 to 2.6)	
High	1.6 (1.2 to 2.3)	1.9 (1.2 to 3.0)	
Specificity (ref: low)			0.11
Medium	2.5 (1.5 to 4.0)	2.7 (1.6 to 4.5)	
High	8.8 (5.8 to 13.5)	5.0 (3.1 to 7.9)	
Publication characteristics, not content related			
Conclusive title (yes vs no)	0.4 (0.3 to 0.7)	0.3 (0.2 to 0.4)	0.10
Funding source (ref: exclusively non-profit)			0.09
Profit or both profit/non-profit	0.9 (0.6 to 1.2)	0.8 (0.6 to 1.2)	
Not reported	0.4 (0.3 to 0.5)	0.8 (0.6 to 1.0)	
No of authors (ref: 1–2)			0.09
3–5	1.9 (1.4 to 2.5)	1.0 (0.7 to 1.5)	
≥6	3.6 (2.7 to 4.9)	1.5 (1.0 to 2.4)	
No of affiliations (ref: 1)			0.09
2	2.6 (2.0 to 3.5)	1.8 (1.3 to 2.5)	
≥3	2.1 (1.6 to 2.8)	1.6 (1.1 to 2.2)	
No of references (ref: <30)			0.09
30–50	1.1 (0.8 to 1.3)	1.0 (0.8 to 1.2)	
≥50	0.4 (0.3 to 0.6)	0.8 (0.5 to 1.3)	
Journal characteristics			
Journal impact factor (ref: 0–2, n=5033)			0.11
2–4	3.4 (2.2 to 5.3)	2.7 (1.7 to 4.2)	
≥4	6.0 (4.0 to 9.2)	4.9 (3.2 to 7.6)	
Author characteristics			
Gender (female vs male, n=5392)	0.8 (0.6 to 1.0)	0.7 (0.6 to 1.0)	0.09
Continent (ref: Europe)			0.11
North America	0.9 (0.6 to 1.3)	0.9 (0.5 to 1.4)	
Asia	0.3 (0.1 to 0.6)	0.2 (0.1 to 0.3)	
Australia	0.5 (0.3 to 0.8)	0.5 (0.3 to 0.8)	
Type of Affiliation (other vs university)	2.3 (1.8 to 2.9)	2.0 (1.5 to 2.5)	0.10

Continued

Table 2 Continued

	Crude OR	Adjusted OR*	R ² *
Citation characteristics			
Authority (ref: low)			0.11
Medium	2.4 (1.8 to 3.3)	1.8 (1.3 to 2.5)	
High	3.7 (2.8 to 4.8)	2.7 (2.0 to 3.7)	

*Adjusted for study design. Bold odds ratios are statistically significant at $p < 0.05$.

Supportive: supportive for Strachan's original hygiene hypothesis, that is, inverse association between siblings/infections and allergy.

Non-supportive: no association or positive association between siblings/infections and allergy.

N, number of publications; n, number of potential citation paths.

mixed results, compared with 4% of the publications on only number of siblings and 21% of the publications on only infection history. As double exposure studies are also cited more often compared with the single exposure studies, type of exposure should be considered as a confounder of study outcome. To correct for this, we performed an explorative random-effects logistic regression of citation on study outcome, adjusted for both study design and type of exposure. It showed that supportive publications had the highest chance of being cited (adjusted OR (adjOR) 3.1, 95% CI 2.2 to 4.3), compared with publications with mixed results (adjOR 2.4, 95% CI 1.5 to 3.7) and with non-supportive results (reference category; model $R^2=0.12$).

Surprisingly, publications with a conclusive title were less likely to receive citations. The format of the title may be prescribed by the journal regulations. We ran some explorative analyses in which we additionally adjusted for the (log-transformed) journal impact factor or publisher on top of study design. The impact of title conclusiveness remained high when additionally adjusted for journal impact factor (adjOR 0.4, 95% CI 0.2 to 0.6) or publisher (adjOR 0.3, 95% CI 0.2 to 0.6).

The above results are related to the network as a whole. Of particular importance is how empirical, evidence-generating publications were cited by the rest of the network. We repeated the above analyses on a subset of the cited publications, namely the empirical publications and tested which of their characteristics were related to citation. The results (online supplementary table S3) are very similar to the analyses on the complete network that include the cited non-empirical publications.

Likewise, we tested how *empirical* publications were cited by *synthesis* publications (online supplementary table S4). Again, the direction and magnitudes of the effects were all very similar, except for study outcome. Adjusted for study design, (log-transformed) sample size and type of exposure, supportive empirical publications were much more likely to be cited (adjOR 7.3, 95% CI 3.5 to 15.5) by reviews and editorials, whereas empirical publications with mixed results seemed less likely to be cited (adjOR 0.4, 95% CI 0.2 to 0.9) compared with non-supportive empirical publications (reference category; model $R^2=0.12$). As a side note: these analyses are

based on a smaller number of cited and citing publications and should be interpreted with caution.

The sensitivity analyses without the four most cited publications showed some dissimilar results (online supplementary table S5). The impact of study outcome decreased, the impact of male authors and of North American authors disappeared and the impact of case-control studies reversed. The other two sensitivity analyses (with a 1-year lag time immediately after publication; without citing publications with less than 10 potential citation paths) all showed similar results as the main analyses (online supplementary tables S6 and S7).

Impact of concordance

In addition, we tested whether publications were more likely to be cited by publications with similar characteristics. The results are shown in table 3. It shows that publications tended to be cited mostly by publications with the same type of exposure, with a similar study outcome, with a corresponding author from the same continent and with one or more authors in common ('self-citation').

DISCUSSION

Our research aim was to evaluate the impact of study outcome and other factors on the likelihood of being cited in the scientific literature on the original hygiene hypothesis stated by Strachan.² We found that study outcome, type of exposure, study design, specificity, title conclusiveness, journal impact factor and the authors' continent, affiliation, authority and self-citation all have a substantial impact on the likelihood of citation.

With regard to study outcome, supportive publications are cited more than three times more often than non-supportive publications, while publications with mixed results are cited more than two times as often. This is a clear sign of citation bias, and corroborates previous findings.¹³ Similarly, publications are more likely to refer to other publications with the same study outcome rather than to those that provide counterevidence to their conclusion. This type of citation bias (based on concordance) has not been studied frequently. In our previous network analyses, on trans fatty acids—cholesterol, and on chlorinated water—asthma, we found no evidence for increased citations between publications with

Table 3 Concordance ORs (95% CIs) for the chance of being cited, all types of articles included, N=110, n=5551

	Crude OR	Adjusted OR*	R ² *
Content related			
Type of exposure (conc. vs not)	10 (5.6 to 18)	13 (7.1 to 23)	0.14
Study outcome (conc. vs not, n=1799)†	2.9 (1.4 to 6.0)	3.4 (1.6 to 7.1)	0.06
Publication type (conc. vs not)	1.1 (0.9 to 1.3)		0.00 (crude)
Not content related			
Funding source (conc. vs not, n=1475)‡	1.3 (0.8 to 1.9)	1.2 (0.8 to 1.8)	0.05
Publisher (conc. vs not, n=4971)§	0.9 (0.7 to 1.2)	0.9 (0.7 to 1.2)	0.08
Author affiliation (conc. vs not)	0.7 (0.6 to 0.9)	0.8 (0.7 to 1.0)	0.08
Author gender (conc. vs not, n=5254)¶	1.0 (0.8 to 1.3)	1.0 (0.8 to 1.3)	0.08
Author continent (conc. vs not)	2.0 (1.7 to 2.5)	2.1 (1.7 to 2.6)	0.09
Self-citation (yes vs no)	6.1 (3.8 to 9.7)	6.1 (3.7 to 9.9)	0.09

*Adjusted for study design of cited publication. Bold odds ratios are statistically significant at $p < 0.05$.

†Publications with mixed results excluded from analysis.

‡Publications without reported funding source excluded from analysis.

§Three main publishers are differentiated: Wiley-Blackwell, BMJ and Elsevier. Either the cited or the citing publication should be in one of these categories to be included in the analysis.

¶Publications with unclear author's gender excluded from analysis.

N, number of publications; n, number of potential citation paths.

the same study outcomes,^{25 26} but three other studies, all related to cardiovascular disease, did find evidence for this type of citation bias.^{9 29 30}

The magnitude of citation bias even increases if we focus on how empirical publications are cited by reviews and editorials. Reviews and editorials in our network are up to eight times more likely to cite supportive publications rather than non-supportive ones. As reviews are generally assumed to give an unbiased summary of the existing evidence, this is a worrying finding. It confirms the notion that people should be cautious to rely on narrative reviews.

Greenberg states that reviews play an important role in the development and acceptance of belief systems.⁸ According to him, reviews can amplify the impact of empirical studies because their evidence is propagated when these reviews are cited themselves. Trinquart *et al* showed that reviews (including systematic reviews) on the health impact of salt intake display signs of citation bias, and that the conclusions of these reviews were in the same direction as the evidence they include.⁹ A similar link between the selective citation of supportive evidence and supportive conclusions of reviews was found by Leng.²⁹ This mechanism might explain how reviews can amplify the effect of citation bias. If reviews draw supportive conclusions based on selective citation of supportive evidence, then support for a hypothesis will be propagated while counterevidence will fade from the literature.

In our analyses, we consider study design as a proxy for study quality. We believe systematic reviews to be of higher quality than narrative reviews and editorials, and thus to receive more citations. In our network, this is indeed the case. Similarly, we believe that cohort studies outrank cross-sectional and case-control studies but to our surprise they are less likely to be cited. Prospective

cohort studies, even though they provide the highest type of evidence in this network, receive the fewest citations of all empirical study designs. This may be due to the fact that these cohort studies tend to focus on multiple risk factors of which only one or two are relevant for the hygiene hypothesis. But the fact that multiple risk factors are investigated in these cohort studies does not imply that their findings on the impact of siblings or infections are of any lesser value or should be ignored.

This study has several limitations. First, it includes two overlapping subnetworks, as is shown by the high OR in the concordance analysis of type of exposure. This makes it difficult to infer for which result a certain publication was cited. Related to this issue, the preregistered operationalisation of study outcome could not be applied because of the hybrid nature of the network, so we developed a scoring system that fits better. Also, there are different versions of the hygiene hypothesis, and support for one version may not be supportive for another one. We dealt with this issue by limiting ourselves to Strachan's original hygiene hypothesis, and by excluding any determinants with conflicting predictions in different versions. Despite these limitations, sensitivity analyses show that the results seem robust against chance findings. Another limitation is our use of ORs to assess the likelihood of citation. The OR may overestimate the true relative risk in studies where the outcome is common (ie, occurs in more than 5% of all cases³¹). In our network, citation is not a common outcome (7%) and consequently the overestimation of the true relative risk will be relatively small.

To conclude, there is evidence for selective citation in this network. Several characteristics of a publication can make it more likely to be cited such as the authority and the continent of the author, the impact factor of the journal, the way

in which the title was stated, and also study design and study outcome. The fact that supportive publications are cited more often than non-supportive ones, particularly if we look at how empirical publications are being picked up by the rest of the network, is a clear sign of citation bias. Finally, this study also shows that particularly narrative reviews may have a preference to refer to supportive evidence.

Author affiliations

¹Nutrition and Translational Research in Metabolism (School NUTRIM), Maastricht University, Maastricht, Limburg, The Netherlands

²Department of Epidemiology and Biostatistics, Amsterdam University Medical Center, Amsterdam, The Netherlands

³Department of Philosophy, Faculty of Humanities, Vrije Universiteit, Amsterdam, Noord-Holland, The Netherlands

⁴Care and Public Health Research Institute (School CAPHRI), Maastricht University, Maastricht, Limburg, The Netherlands

Acknowledgements The authors would like to thank Jean Muris for his advice on the search strategy, Hans Verheij en Michel Duyx for their methodological advice, and the reviewers for their very helpful comments.

Contributors BD was involved in the research design and data analysis plan, prepared the research protocol, developed and performed the search strategy, wrote the R scripts for data transformation and calculation, performed article selection and data extraction, conducted the analyses and wrote the manuscript. MJEU was involved in the research design and data analysis plan, performed the article selection and data extraction, and read, commented on and approved the final manuscript. GMHS obtained funding, was involved in the research design and data analysis plan, and read, commented on and approved the final manuscript. MPZ obtained funding, was involved in the research design and data analysis plan, and read, commented on and approved the final manuscript. LMB was involved in the research design and data analysis plan, and read, commented on and approved the final manuscript.

Funding This project has received funding from the Long-Range Research Initiative (LRI) from the European Chemical Industry Council (CEFIC), grant no. LRI-Q3-UM.

Disclaimer LRI had no role in study design, data collection and analysis, preparation of the manuscript or decision to publish.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The protocol and the data of this study are available on request in the DataVerse repository, <http://hdl.handle.net/10411/ZKGGOG>, or by sending an email to b.duyx@maastrichtuniversity.nl.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Strachan D. Family size, infection and atopy: the first decade of the 'hygiene hypothesis'. *Thorax* 2000;55:2S–10.
2. Strachan DP. Hay fever, hygiene, and household size. *BMJ* 1989;299:1259–60.
3. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:1–193.
4. Platts-Mills TA. The allergy epidemics: 1870–2010. *J Allergy Clin Immunol* 2015;136:3–13.
5. Rook GA. Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. *Proceedings of the National Academy of Sciences* 2013;110:18360–7.
6. Rook GA. *Evolution, human-microbe interactions and the "epidemic" of inflammation-related disorders (alias: the hygiene hypothesis)*. London: The Biomedical and Life Sciences Collection: Henry Stewart Talks Ltd, 2016. <https://hstalks.com/bs/3317/>
7. Fanelli D. Positive results receive more citations, but only in some disciplines. *Scientometrics* 2013;94:701–9.
8. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. *BMJ* 2009;339:b2680.
9. 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;45:251–60.
10. 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;13:1698–704.
11. Chapman S, Ragg M, McGeechan K. Citation bias in reported smoking prevalence in people with schizophrenia. *Aust N Z J Psychiatry* 2009;43:277–82.
12. 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:e28–e32.
13. Duyx B, Urlings MJE, Swaen GMH, et al. Scientific citations favor positive results: a systematic review and meta-analysis. *J Clin Epidemiol* 2017;88:92–101.
14. Onodera N, Yoshikane F. Factors affecting citation rates of research articles. *J Assoc Inf Sci Technol* 2015;66:739–64.
15. Rigby J. Looking for the impact of peer review: does count of funding acknowledgements really predict research impact? *Scientometrics* 2013;94:57–73.
16. 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;118:1321–7.
17. Kulkarni AV, Busse JW, Shams I. Characteristics associated with citation rate of the medical literature. *PLoS One* 2007;2:e403.
18. Kostoff RN. The difference between highly and poorly cited medical articles in the journal Lancet. *Scientometrics* 2007;72:513–20.
19. Dion ML, Sumner JL, Mitchell SM. Gendered citation patterns across political science and social science methodology fields. *Political Analysis* 2018;26:312–27.
20. Esarey J, Bryant K. Are papers written by women authors cited less frequently? *Political Analysis* 2018;26:331–4.
21. Østby G, Strand H, Nordås R, et al. Gender gap or gender bias in peace research? publication patterns and citation rates for journal of peace research, 1983–2008. *International Studies Perspectives* 2013;14:493–506.
22. Annalingam A, Damayanthi H, Jayawardena R, et al. Determinants of the citation rate of medical research publications from a developing country. *Springerplus* 2014;3:140.
23. Callahan M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA* 2002;287:2847–50.
24. Nieminen P, Rucker G, Miettunen J, et al. Statistically significant papers in psychiatry were cited more often than others. *J Clin Epidemiol* 2007;60:939–46.
25. Duyx B, Urlings MJE, Swaen GMH, et al. Selective citation in the literature on swimming in chlorinated water and childhood asthma: a network analysis. *Res Integr Peer Rev* 2017;2:17.
26. Urlings MJE, Duyx B, Swaen GMH, et al. Citation bias in the literature on dietary trans fatty acids and serum cholesterol. *J Clin Epidemiol* 2019;106:88–97.
27. Duyx B, Urlings MJE, Swaen G, et al. Protocol for citation analysis of the published literature on the hygiene hypothesis. 2017 <http://hdl.handle.net/10411/ZKGGOG>
28. van Eck NJ, Waltman L. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *J Informetr* 2014;8:802–23.
29. Leng RI. A network analysis of the propagation of evidence regarding the effectiveness of fat-controlled diets in the secondary prevention of coronary heart disease (CHD): Selective citation in reviews. *PLoS One* 2018;13:e0197716.
30. 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;69:174–8.
31. Cummings P. The relative merits of risk ratios and odds ratios. *Arch Pediatr Adolesc Med* 2009;163:438–45.