HHS Public Access

Author manuscript

Clin Sci (Lond). Author manuscript; available in PMC 2022 March 12.

Published in final edited form as:

Clin Sci (Lond). 2021 March 12; 135(5): 671–681. doi:10.1042/CS20201573.

How accurate are citations of frequently cited papers in biomedical literature?

V Pavlovic[1], **T Weissgerber**[2], **D Stanisavljevic**[1], **T Pekmezovic**[3], **O Milicevic**[1], **J Milin Lazovic**[1], **A Cirkovic**[1], **M Savic**[1], **N Rajovic**[1], **P Piperac**[4], **N Djuric**[1], **P Madzarevic**[1], **A Dimitrijevic**[5], **S Randjelovic**[1,5], **E Nestorovic**[5], **R Akinyombo**[1], **A Pavlovic**[1], **R Ghamrawi**[6], **V Garovic**[6], **N Milic**[1,6]

[1]Institute for Medical Statistics and Informatics, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

[2]Charité - Universitätsmedizin Berlin, Berlin Institute of Health, QUEST Center, Berlin

[3]Institute of Epidemiology, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

[4]Department of Humanities, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

[5]Clinical Center of Serbia, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

[6]Division of Nephrology and Hypertension, Department of Internal Medicine, Mayo Clinic, Rochester, MN USA

Abstract

Citations are an important, but often overlooked, part of every scientific paper. They allow the reader to trace the flow of evidence, serving as a gateway to relevant literature. Most scientists are aware of citations errors, but few appreciate the prevalence of these problems. The purpose of this study was to examine how often frequently cited papers in biomedical scientific literature are cited inaccurately. The study included an active participation of the first authors of included papers; to first-hand verify the citations accuracy. Findings from feasibility study, where we reviewed 1,540 articles containing 2,526 citations of 14 most cited articles in which the authors were affiliated with the Faculty of Medicine University of Belgrade, were further evaluated for external confirmation in an independent verification set of articles. Verification set included 4,912 citations identified in 2,995 articles that cited 13 most cited articles published by authors affiliated with the Mayo Clinic Division of Nephrology and Hypertension. A citation was defined as being accurate if the cited article supported or was in accordance with the statement by citing authors. At least one inaccurate citation was found in 11% and 15% of articles in the feasibility study and verification set, respectively, suggesting that inaccurate citations are common in biomedical literature. The most common problem was the citation of nonexistent findings (38.4%), followed by an incorrect interpretation of findings (15.4%). One fifth of inaccurate citations were due to chains of inaccurate citations. Based on these findings, several actions to reduce citation inaccuracies have been proposed.

Corresponding author: Natasa Milic, milic.natasa@mayo.edu, Dr. Subotica 15, 11000, Belgrade, Serbia, +38163367700.

Data Availability Statement

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Keywords

Citation; Biomedical literature; Accuracy

Introduction

Citations are an important, but often overloooked, part of every scientific paper. They allow the reader to trace the flow of ideas and evidence through a paper, serving as a gateway to other relevant literature. Citations also allow readers to confirm that the cited information supports the authors' hypotheses and suppositions. Many deficits in citation practices unfortunately have been reported. A recent meta-analysis demonstrated that 25.4% of papers contained a citation error (1). Van der Vet and Nijveen (2) reported that even retracted articles continue to be cited approvingly years after they have been retracted. Citation errors can have serious implications. Millions of Americans suffer from opioid addiction and opioids contributed to 47,600 drug overdose deaths in the United States in 2017 (3). A team of Canadian researchers proposed that uncritical citations of a letter published in the New England Journal of Medicine (4) may have contributed to the opioid crisis (5). The letter stated that narcotic addiction was rare in hospital inpatients with no histories of addiction (4). This five sentence letter contained no detailed methods or results (4); yet it was cited hundreds of times as evidence that addiction risk was low when opioids were perscribed for chronic pain (5). Some citations clearly distorted the findings of the letter, and 81% of the letter's 608 citations did not mention that the study only included hospital inpatients. Leung et al. (5) argue that these uncritical and misleading citations may have helped to shift perscribing practices by convincing doctors that addiction risk was low with chronic opioid use.

A published letter attempting to correct an overestimate of the number of Cochrane reviews on rehabilitation interventions (6) provides another example of the dangers of citation copying (7). Although the article (6) was cited 62 times, all of these citations were related to meta-analyses of genetic risk factors (7). Authors cited the letter to support their use of the Cochran Q-statistic for exploring heterogeneity of effect sizes, whereas the letter was written to highlight the need for Cochrane reviews on rehabilitation.

While the academic literature is replete with similar errors, assessing citation accuracy takes considerable effort. Most scientists are aware of citation errors and copying, but few appreciate their prevalence or consequences. The purpose of this study was to: 1) examine how often frequently cited papers in biomedical literature are cited inaccurately, 2) explore factors associated with inaccurate citations and 3) discuss actions that authors, mentors, and journals can take to eliminate citation errors.

Methods

The study was designed to include an active participation of first authors of the frequently cited papers in biomedical scientific literature, to first-hand verify the accuracy of the citations of their original work. The approach was to determine most cited original articles and their parent authors, which could be feasible to access, and identify, collect and review

all citations of their original work throughout the biomedical scientific literature. As this approach resulted in a time consuming project that is complex to manage, we had conducted a feasibility study, whose results were than further evaluated for external confirmation in an independent verification set of articles. The study was approved by the Ethical Committees of the Faculty of Medicine University of Belgrade (2650/IV-6) and the Mayo Clinic, Rochester, MN, USA (19-005085).

Feasibility study

The sample was formed in two stages (Figure 1). We first chose “source articles” - the most cited articles in which the 1st authors were affiliated with the Faculty of Medicine University of Belgrade, according to a SCOPUS bibliographic database search on October 1, 2017. Based on our hypothesis that the frequency of citation inaccuracies would be 10%, we calculated that 1500 citing articles would be needed to estimate the frequency of inaccurate citations with a precision of 1.5% (alpha=0.05). Fourteen source articles were chosen to reach this predetermined sample size (n=1500 citing articles). Characteristics of the source articles, including the number of citations, are presented in Table 1. Source articles were published between 1994 and 2009. The time elapsed from the publication of the article to the beginning of this study was between 8 and 23 years, and the total number of source article citations for this period ranged from 63 to 393. The publication field was determined according to journal classification from Journal Citation Reports (JCR). In the second stage, we collected all “citing articles” which cited the source articles (according to a SCOPUS bibliographic database search on October 1, 2017). Citing articles were included in the study if they were published in English and we could obtain a full text version of the article. Articles written in other languages, as well as books and book chapters were excluded. The final sample included 1540 of the 1565 citing articles published in English; 25 citing papers (1.6%) were excluded because a full-text version of the manuscript could not be retrieved.

Assessing Citation Accuracy: A citation was defined as being accurate if the reference (source article) supported or was in accordance with the statement by citing authors. Each citing article was first reviewed for citation accuracy and discussed by three reviewers. If inaccuracies were detected, citations were further evaluated by one of the authors of the source article, including first authors (n=11) or another co-author and active member of the research team (n=3). Each author checked the accuracy of citations of his or her own paper and classified the type of inaccurate citation as follows: citation of nonexistent findings, incorrect interpretation of findings, incorrectly cited method, incorrectly cited numerical data/results, citation of nonexistent numerical data/results, wrong context, cited findings from another source, or reference listed in the bibliography but not cited in the text. We extracted the following data from each citing article to identify factors associated with citation inaccuracies: publication year, article type as defined by the journal in which the article was published (i.e. original article, review, perspective, editorial, etc.), self-citation, referencing style, number of authors, impact factor of the journal, and number of references in the bibliography. The impact factor was extracted from Journal Citation Reports (JCR), if available, for the year in which the citing paper was published. Citation was considered as a self-citation if a citing paper and source paper had at least one author in common.

Verification set

Source articles were the most cited articles published by authors affiliated with the Mayo Clinic Division of Nephrology and Hypertension (Rochester, Minnesota, USA), whose research focus is hypertension and peripheral vascular disease. Most cited articles were determined by a SCOPUS database search on May 1, 2019. It was planned to include twice as many citing articles in the verification set compared to the feasibility study. The final verification sample included 2,995 of the 3,096 citing articles published in English; 3.3% of articles were excluded due to unavailability of a full-text version of the manuscript. The procedure of reviewing citing articles was the same as in the feasibility study. Characteristics of the source articles in the verification set are presented in Table 2.

Statistical analysis

Descriptive statistics, including numbers and percentages for categorical data, and median and range for numerical data, were calculated to describe the study sample. The sampling scheme used in this study, in which affiliations and source articles are clusters, introduces multilevel dependency or correlation among the observations that can affect model parameter estimates. Therefore, we used a multilevel regression model for binary data to determine predictors of inaccurate citations. The model had a three-level data structure; the first level was citing articles, the second level was the source articles and the third level was affiliations. Statistical analyses were performed using the R environment for statistical computing (RRID:SCR_001905) (8) with the lme4 package (RRID:SCR_015654) (9). Significance level (alpha) was set at 0.05.

Results

In total, we reviewed 4,535 citing articles (1,540 in the feasibility study and 2,995 in the verification set). The characteristics of these citing articles are shown in Table 3. The most common article types were original research (54.9%) and reviews (29.2%). The Vancouver or mixed citation style was used in most of the articles (92.0%). The median number of authors of the citing articles was five (range 1 to 65). The median impact factor of the 3,995 articles (88.1%) published in journals that had an impact factor at the time of publication was 3.262 (minimum 0.049, maximum 79.60). The source article was cited once (68.7%) in most cases. There were no discrepancies in the main characteristics between the feasibility study and verification set. The total number of citations of source articles was 2,526 and 4,912 in the feasibility study and verification set, respectively. The proportion of inaccurate citations in the feasibility study was 7.2% (183/2,526), while the proportion of articles containing at least one inaccurate citation was 11.1% (171/1,540). The presence of inaccurate citations was confirmed in the verification set, where the frequency of inaccurate citations was 10.3% (505/4,912), with the frequency of articles containing at least one inaccurate citation of 15.0% (449/2,995). Table 4 describes the types of citation inaccuracies in both sets. The most common finding was the citation of nonexistent findings (38.4%), followed by inaccurately cited numerical data/results (16.6%), inaccurate interpretation of findings (15.4%) and citations of quoted findings of another source (15.1%). The frequencies of the other types of inaccurate citations were below ten percent. In structured research articles, inaccurate citations mostly appeared in the introduction and discussion

sections. Reviewers identified 13 chains of inaccurate citations in the feasibility study, in which the inaccurate citations appeared to have been copied from previous articles that had made the same citation error. These 13 chains included 44 articles with inaccurate citations, which accounted for approximately one fourth of all inaccurate citations (24%). The presence of chains of inaccurate citations was confirmed in the verification set, where 14 chains were identified, including 89 articles with inaccurate citations. Inaccurate citations included in the chains accounted for 19.3% of all inaccurate citations.

Binary logistic regression models, with the presence of an inaccurate citation as a dependent variable in the model, are shown in Table 5. Statistically significant predictors in the univariate analyses were review articles, time elapsed time from publication to citation, impact factor and number of citations of the source article. Review articles, longer time elapsed from publication to citation, and a higher number of citations of the source article were associated with a greater risk of inaccurate citations in a multivariate model.

Binary logistic regression models, with presence of chains of inaccurate citations as a dependent variable in the model, are shown in Table 6. Statistically significant predictors for the presence of chains in the univariate analyses were number of authors, self citation and number of references. In a multivariate model, higher number of references was associated with the occurrence of chains of inaccurate citations in biomedical literature.

Discussion

In this study, we found that inaccurate citations are common in biomedical scientific literature. At least one inaccurate citation was found in 11% of the reviewed articles in feasibility study. This finding was confirmed in the verification set of articles, where citation inaccuracies were detected in 15.0% of articles. The study was designed to determine the presence and types of inaccurate citations of the most cited original research articles from authors affiliated with two major research centers and to explore factors associated with inaccurate citations. The strengths of this study included collaboration with authors of the source articles to confirm and classify citation inaccuracies. Previous studies have used a “journal based approach” to determine the percentage of papers containing inaccurate citations (10). In contrast, we used a “source article based approach” to quantify the proportion of inaccurate citations for the most cited articles published by authors affiliated by our institutions. This approach yielded several important findings. Our results suggest that approximately one in ten citations of a highly cited article is inaccurate. Almost half of the citation inaccuracies in our sample were due to the citation of a non-existent finding, whereas 13.8% were due to an inaccurate interpretation of research findings. One fifth of the citation inaccuracies were due to chains of inaccurate citations, in which citation errors appeared to have been copied from previous papers. Review articles were more likely to contain inaccurate citations.

Although many studies have examined citation inaccuracies, our results may not be directly comparable due to differences in study design. Porrino et al. (11) used a similar approach to examine inaccuracies in citations of the Knirk and Jupiter (12) article, and found that 40% of citations were inaccurate by the time of the study (64/159). However, the generalizability of

this finding was limited due to the fact that this study examined citations of a single article, which was selected because the authors were aware of the high rate of citation inaccuracies. Studies using the traditional journal based approach have reported that between 10% and 50% of papers contain citation inaccuracies (1,13). Only a few studies have reported a rate of inaccurate citations below 10% (14,15). One possible reason for this variability may be differences in the complexity (16) and scientific fields of the source articles, covering topics ranging from pure basic to applied clinical research.

In contrast to our results, previous studies have reported associations between citation inaccuracies and citation style (Harvard vs. Vancouver) (1), the number of authors (one vs. more than one) (17) or the number of references (18). These divergent results may also be partially due to study design differences. In contrast to other studies (14,19,20), we have found an association between inaccurate citations and journal impact factor. There were large differences in rates of inaccurate citations among our source articles (from 3.2% to 28.6%).

Our findings, along with previous studies demonstrating that citation inaccuracies are common, have several important implications for authors. Authors should adopt good citation practices, including those outlined in Table 7, when preparing manuscripts. These practices are important for all types of publications, including review articles, which were more likely to contain citation inaccuracies in our study. Inclusion of full texts of all citations in reference manager libraries should become prerequisite. Practices such as sharing libraries and asking multiple authors to check and confirm each citation may help to prevent common inaccuracies, including citations of non-existent findings and inaccurate interpretations of research findings.

Scientists should also take steps to prevent the propagation of chains of inaccurate citations. These include carefully reading all papers prior to citation to prevent an inaccurate citation “domino effect,” not copying citations from other sources, and raising awareness about known chains of citation inaccuracies in the scientists’ fields. The practice of citing original papers without reading them has been already recognized in the literature as “lazy author syndrome” (21). Teixeira et al. demonstrated that 15% of citations in ecology journals inappropriately referenced reviews instead of the original articles of authors who proposed the idea or reported research findings (22).

Citation inaccuracies undermine the integrity of the scientific literature and can have serious consequences, however, good citation practices are rarely taught. Principal investigators can promote better practices by establishing standard citation protocols for their laboratories and engaging trainees in the process of verifying citation accuracy for their publications. The citations section of Table 7 includes references that provide more information regarding many of the practices described in the table. Other members of the scientific community can also develop incentives and implement strategies to improve citation accuracy. Table 8 provides an overview of strategies that journal editors can consider emphasizing the importance of citation accuracy and promoting good citation practices.

Limitation of our study is that source articles in the feasibility study and verification set of articles were each selected from one institution or department. This limitation only applies to the source articles because the citing articles came from different institutions and journals worldwide. However, findings were similar in both sets, suggesting that they may be generalizable to other institutions or departments. Selection of highly cited articles may limit the generalizability of the findings to articles with fewer citations. Additional limitations are the exclusion of articles not published in English and the use of a single–database based methodology.

Acknowledgement

To the memory of Professor Goran Trajkovic, MD, PhD, Faculty of Medicine, University of Belgrade (1963-2019).

Sources of Funding

This study was supported by the National Institutes of Health (NIH) (R01HL 136348; Vesna Garovic). Natasa Milic has received funding from the QUEST - Quality | Ethics | Open Science | Translation, Charité - Universitätsmedizin Berlin, Berlin Institutes of Health (BIH).

References

1. Jergas H, Baethge C. Quotation accuracy in medical journal articles-a systematic review and meta-analysis. PeerJ. 2015;3:e1364. [PubMed: 26528420]
2. van der Vet PE, Nijveen H. Propagation of errors in citation networks: a study involving the entire citation network of a widely cited paper published in, and later retracted from, the journal Nature. Res Integr Peer Rev. 2016;1(1):3. [PubMed: 29451542]
3. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and Opioid-Involved Overdose Deaths - United States, 2013–2017. MMWR Morb Mortal Wkly Rep. 2018;67(5152):1419–27. [PubMed: 30605448]
4. Porter J, Jick H. Addiction rare in patients treated with narcotics. N Engl J Med. 1980;302(2):123.
5. Leung PTM, Macdonald EM, Stanbrook MB, Dhalla IA, Juurlink DN. A 1980 Letter on the Risk of Opioid Addiction. N Engl J Med. 2017;376(22):2194–5. [PubMed: 28564561]
6. Handoll HH. Systematic Reviews on Rehabilitation Interventions. Arch Phys Med Rehabil. 2006;87(6):875. [PubMed: 16731227]
7. Handoll HHG, Atkinson G. Snowballing citations. BMJ. 2015;351:h6309. [PubMed: 26666656]
8. R Core Team. A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: www.R-project.org
9. Bates DM, DebRoy S. Linear mixed models and penalized least squares. J Multivar Anal. 2004;91(1):1–17.
10. Smith N, Cumberledge A. Quotation errors in general science journals. Proc R Soc A Math Phys Eng Sci. 2020;476(2242).
11. Porrino JA, Tan V, Daluiski A. Misquotation of a Commonly Referenced Hand Surgery Study. J Hand Surg Am. 2008;33(1):2.e1–2.e9. [PubMed: 18261657]
12. Knirk JL, Jupiter JB. Intra-articular fractures of the distal end of the radius in young adults. J Bone Joint Surg Am. 1986;68(5):647–59. [PubMed: 3722221]
13. Singh S, Chaudhary R. Accuracy of references cited in articles published in Indian Journal of Dermatology, Venereology and Leprology: a pilot study. Indian J Dermatol Venereol Leprol. 2009;75(5):488–91. [PubMed: 19736428]
14. Buijze GA, Weening AA, Poolman RW, Bhandari M, Ring D. Predictors of the accuracy of quotation of references in peer-reviewed orthopaedic literature in relation to publications on the scaphoid. Bone Joint J. 2012;94-B(2):276–80.
15. Reddy MS, Srinivas S, Sabanayagam N, Balasubramanian SP. Accuracy of references in general surgical journals - An old problem revisited. Surgeon. 2008;6(2):71–5. [PubMed: 18488770]

16. Mogull SA. Accuracy of cited “facts” in medical research articles: A review of study methodology and recalculation of quotation error rate. PLoS One. 2017;12(9):1–17.
17. Schulmeister L. Quotation and reference accuracy of three nursing journals. Image J Nurs Sch. 1998;30(2):143–6. [PubMed: 9775555]
18. Eichorn P, Yankauer A. Do authors check their references? A survey of accuracy of references in three public health journals. Am J Public Health. 1987;77(8):1011–2. [PubMed: 3079520]
19. Mertens S, Baethge C. The virtues of correct citation. Dtsch Arztebl Int. 2011;108(33):550–2. [PubMed: 21912574]
20. Awrey J, Inaba K, Barmparas G, Recinos G, Teixeira PGR, Chan LS, et al. Reference accuracy in the general surgery literature. World J Surg. 2011;35(3):475–9. [PubMed: 21161653]
21. Gavras H Inappropriate attribution: The “Lazy Author Syndrome”. Am J Hypertens. 2002;15(9):831. [PubMed: 12722775]
22. Teixeira MC, Thomaz SM, Michelan TS, Mormul RP, Meurer T, Fasolli JVB, et al. Incorrect Citations Give Unfair Credit to Review Authors in Ecology Journals. PLoS One. 2013;8(12):e81871. [PubMed: 24349143]
23. Lowry SR. How accurate are quotations and references in medical journals? Br Med J (Clin Res Ed). 1985;291(6499):1421.
24. Gupta P, Yadav M, Mohta A, Choudhury P. References in Indian Pediatrics: authors need to be accurate. Indian Pediatr. 2005;42(2):140–5. [PubMed: 15767709]
25. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Citeseer. 2015;(12):1–17.
26. Goldberg R, Newton E, Cameron J, Jacobson R, Chan L, Bukata WR, et al. Reference accuracy in the emergency medicine literature. Ann Emerg Med. 1993;22(9):1450–4. [PubMed: 8363119]
27. Lee SY, Lee JS. A survey of reference accuracy in two Asian dermatologic journals (the Journal of Dermatology and the Korean Journal of Dermatology). Int J Dermatol. 1999;38(5):357–60. [PubMed: 10369545]
28. Fenton JE, Brazier H, De Souza A, Hughes JP, Mcshane DP. The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. Clin Otolaryngol Allied Sci. 2000;25(1):40–4. [PubMed: 10764235]
29. Lukić IK, Lukić A, Glunčić V, Katavić V, Vučenik V, Marušić A. Citation and quotation accuracy in three anatomy journals. Clin Anat. 2004;17(7):534–9. [PubMed: 15376293]
30. de Lacey G, Record C, Wade J. How accurate are quotations and references in medical journals? Br Med J (Clin Res Ed). 1985;291(6499):884–6.
31. Ingelfinger FJ. Seduction by citation. N Engl J Med. 1976;295(19):1075–6. [PubMed: 972662]

Clinical perspectives

- Most scientists are aware of citation errors and copying, but few appreciate their prevalence or consequences. The purpose of this study was to examine how often frequently cited papers in biomedical literature are cited inaccurately.
- In this study, we found that inaccurate citations are common in biomedical scientific literature. At least one inaccurate citation was found in 11% of the articles in feasibility study. This finding was confirmed in the verification set of articles, where citation inaccuracies were detected in 15.0% of articles. The most common problem was the citation of nonexistent findings (38.4%), followed by an incorrect interpretation of findings (15.4%).
- Citation inaccuracies can undermine the integrity of the scientific literature. To reduce citation inaccuracies, several actions that authors, mentors and journals can take have been proposed.

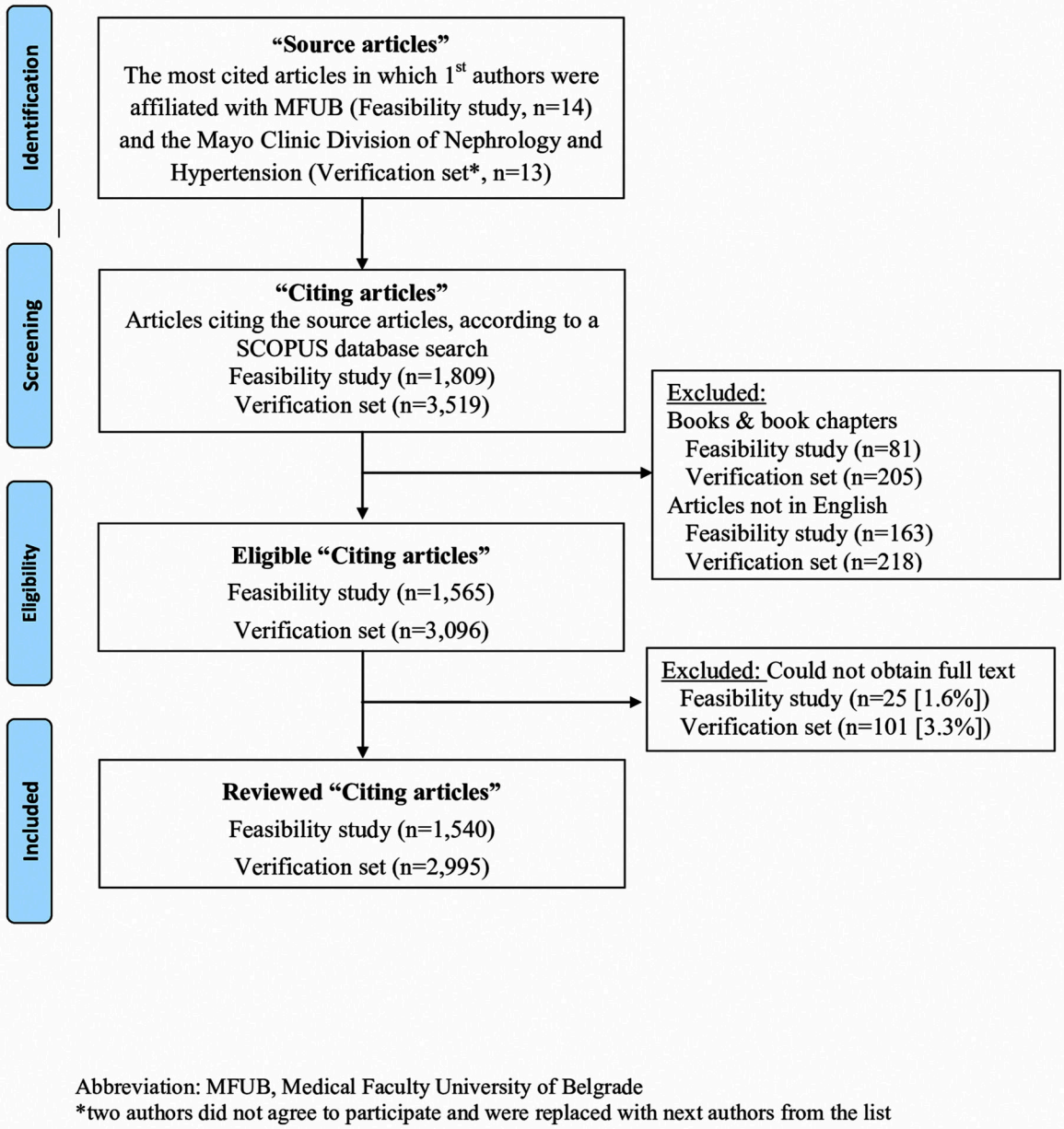

Figure 1.
Study design flow chart

Table 1.

Characteristics of source articles in the feasibility study

Source article	Publication field	Publication year	Years since publication	Number of citations			
				Articles in English	Articles not in English*	Books and book chapters*	Total citations
1	Multidisciplinary Sciences	1997	20	353	16	24	393
2	Cardiac & Cardiovascular Systems, Peripheral Vascular Disease	1994	23	184	35	15	234
3	Endocrinology & Metabolism	1995	22	179	5	1	185
4	Rheumatology	2009	8	127	3	3	133
5	Infectious Diseases	2005	12	102	8	7	117
6	Endocrinology and Metabolism, Nutrition & Dietetics	2009	8	108	1	4	113
7	Clinical Neurology, Psychiatry, Surgery	1994	23	85	18	9	112
8	Clinical Neurology	2002	15	76	9	4	89
9	Clinical Neurology	2001	16	56	21	4	81
10	Urology & Nephrology	1996	21	45	31	2	78
11	Clinical Neurology	2001	16	71	4	1	76
12	Medicine, Legal, Pathology	2003	14	61	6	1	68
13	Immunology, Neurosciences	2001	16	62	2	3	67
14	Peripheral Vascular Disease	2006	11	56	4	3	63
Total				1,565	163	81	1,809

*These categories were excluded from subsequent analyses

Table 2.

Characteristics of source articles in the verification set

Source article	Publication field	Publication year	Years since publication	Number of citations			
				Articles in English	Articles not in English*	Books and book chapters*	Total citations
1	Medicine, General & Internal	2004	15	740	62	22	824
2	Medicine, General & Internal	2012	7	510	39	27	576
3	Surgery, Transplantation	2009	10	413	16	24	453
4	Surgery, Transplantation	2006	13	210	9	17	236
5	Peripheral Vascular Disease	2002	17	183	16	15	214
6	Urology & Nephrology	2010	9	181	13	15	209
7	Physiology	1995	24	162	1	6	169
8	Urology & Nephrology	2007	12	129	16	11	156
9	Gastroenterology & Hepatology	2003	16	103	12	30	145
10	Urology & Nephrology	2006	13	126	3	14	143
11	Medical Laboratory Technology	2005	14	126	7	9	142
12	Obstetrics & Gynecology	2007	12	114	3	13	130
13	Urology & Nephrology	2008	11	99	21	2	122
Total				3,096	218	205	3,519

*These categories were excluded from subsequent analyses

Table 3.

Characteristics of citing articles

Characteristics	Total (n=4,535)	Feasibility study (n=1,540)	Verification set (n=2,995)
Article type, * n (%)			
Original research	2,490 (54.9)	836 (54.3)	1654 (55.2)
Review	1324 (29.2)	484 (31.4)	840 (28.0)
Comment/Note	54 (1.2)	8 (0.5)	46 (1.5)
Letter	61 (1.3)	15 (1.0)	46 (1.5)
Brief/short report/communication	50 (1.1)	18 (1.2)	32 (1.1)
Case report/series	136 (3.0)	59 (3.8)	77 (2.6)
Study protocol	19 (0.4)	1 (0.1)	18 (0.6)
Guidelines	35 (0.8)	10 (0.6)	25 (0.8)
Pilot	3 (0.1)	2 (0.1)	1 (0.1)
Opinion	29 (0.6)	0 (0)	29 (1.0)
Editorial, editorial comment	87 (1.9)	11 (0.7)	76 (2.5)
Other	247 (5.4)	96 (6.2)	151 (5.0)
Self-citation, n (%)	484 (10.7)	92 (6.0)	392 (13.1)
Citation style, n (%)			
Vancouver or mixed	4172 (92.0)	1265 (82.1)	2907 (97.1)
Harvard	363 (8.0)	275 (17.9)	88 (2.9)
Number of authors, median (range)	5 (1, 65)	4 (1, 65)	5 (1, 36)
Impact factor †			
Have an impact factor, n (%)	3,955	1,227	2,728
Median (range)	3.262	3.055	3.374
	(0.049, 79.260)	(0.051, 34.833)	(0.049, 79.260)
Number of references in reference list			
Median (range)	41 (1, 1131)	47 (1, 1131)	38 (1, 620)
Time to citation, median (range), years	6 (0, 24)	6 (0, 23)	5 (0, 24)
Number of citations, n (%)			
1	3,114 (68.7)	1,046 (67.9)	2,068 (69.0)
2	775 (17.1)	260 (16.9)	515 (17.2)
3	313 (6.9)	121 (7.9)	192 (6.4)
4	149 (3.3)	52 (3.4)	97 (3.2)
≥5	178 (3.9)	60 (3.9)	118 (3.9)

* Defined by the journal

† Retrieved from Journal Citation Reports for all journals that were indexed at the time when the citing article was published

Table 4.

Citation inaccuracies

Citation inaccuracies	Total n (%)	Feasibility study n (%)	Verification set n (%)
Inaccurate citations, n (%)	688/7,438 (9.2)	183/2,526 (7.2)	505/4,912 (10.3)
Articles with inaccurate citations, n (%)	620/4,535 (13.7)	171/1,540 (11.1)	449/2,995 (15.0)
Type of citation error, n (%) *			
Citation of nonexistent finding	264 (38.4)	86 (47.0)	178 (35.2)
Inaccurate interpretation of findings	106 (15.4)	39 (21.3)	67 (13.3)
Inaccurately cited numerical data/results	114 (16.6)	16 (8.7)	98 (19.4)
Wrong context	41 (6.0)	15 (8.2)	26 (5.1)
Citation of quoted findings of another source	104 (15.1)	11 (6.0)	93 (18.4)
Inaccurately cited method	34 (4.9)	9 (4.9)	25 (4.9)
Citation of nonexistent numerical data/results	18 (2.6)	6 (3.3)	12 (2.4)
Reference listed in bibliography but not cited in the text	6 (0.9)	1 (0.5)	5 (1.0)
Number of structured articles, n (%)	2,439 (53.8)	827 (53.7)	1,612 (53.8)
Inaccurate citations in structured articles, n (%) **			
Introduction	199/1,469 (13.5)	37/451(8.2)	162/1,018 (15.9)
Method	70/399 (17.5)	1/34 (2.9)	69/365 (18.9)
Results	34/159 (21.4)	3/42 (7.1)	31/117 (26.5)
Discussion	280/1,797 (15.6)	80/746 (10.7)	200/1,051 (19.0)
Chains of inaccurate citations			
Number of chains of inaccurate citations	27	13	14
Total number of articles included in chains	133	44	89
Average number of articles included in chain, median (range)	3 (2, 20)	3 (2,7)	4 (2, 20)

* Denominators in this section are total numbers of inaccurate citations

** Denominators in this section list the number of citations of a source article in the specified section of structured original research articles

Table 5.

Factors associated with inaccurate citations

Independent variable	Univariate			Multivariate		
	b	SE	p	b	SE	p
Review article	0.22	0.09	**0.023**	0.22	0.09	**0.022**
Time to citation (years)	0.19	0.08	**0.018**	0.23	0.08	**0.005**
Number of authors	−0.05	0.06	0.340			
Self-citation	0.08	0.14	0.548			
Impact factor, Yes	−0.26	0.13	**0.048**			
Citation style, Vancouver or mixed	−0.14	0.15	0.373			
Number of citations of source article, >1	0.59	0.09	**<0.001**	0.60	0.09	**<0.001**
Reference count	0.11	0.06	0.057			

Data were analyzed by multilevel regression models for binary data, with citation inaccuracy (yes vs. no) as the dependent variable.

b, regression coefficient; SE, standard error; p, p-value

Table 6.

Factors associated with occurrence of chains of inaccurate citations

Independent variable	Univariate			Multivariate		
	b	SE	p	b	SE	p
Review article	0.31	0.19	0.105			
Time to citation (years)	0.30	0.17	0.073			
Number of authors	−0.09	0.00	**<0.001**			
Self-citation	−0.76	0.38	**0.045**	−0.69	0.38	0.070
Impact factor, Yes	−0.39	0.26	0.132			
Citation style	−0.20	0.29	0.483			
Number of citations of source article, >1	−0.09	0.19	0.657			
Reference count	0.39	0.12	**<0.001**	0.37	0.12	**0.001**

Data were analyzed by multilevel regression models for binary data, with citation inaccuracy (yes vs. no) as the dependent variable.

b, regression coefficient; SE, standard error; p, p-value

Table 7:

Actions authors, mentors and readers can take to encourage good citation practices and prevent errors

Recommendations for Authors	Rationale
Use software to manage references **1** Create a separate library for each paper **2** Add full texts of each paper to the library **3** Share the library with all co-authors **4** If the library includes in press citations, pre-prints, abstracts, etc., check for the most recent version before finalizing the bibliography	Citation links automatically update when references are added and removed, reducing the likelihood of errors Authors can easily access full texts to verify citation accuracy
Cite the best available evidence **1** Cite original research or material instead of abstracts or narrative reviews **2** When multiple supporting citations are available, cite more informative studies with stronger designs. For example, cite a Cochrane review instead of a single trial, or a randomized controlled trial instead of observational studies.	**1** Citations refer readers to the most relevant material, not an indirect source (14,24–25) **2** Provides readers with information about the quality of the supporting evidence (14)
Place references close to the word, phrase or sentence to which they refer. Avoid citing groups of references at the end of a sentence or paragraph.	Eliminates confusion about which claim each citation supports (18)
Use the Harvard citation style (author, date) instead of the Vancouver style (numbered) when preparing the manuscript.	Seeing author names and dates may make it easier to detect errors (1)
Check all citations for accuracy **1** Statements should be verified against original papers, not indirect sources **2** Each citation should be checked by at least two authors. Co-authors should be critical about citations.	**1** Avoid propagating errors from inaccurate citation chains **2** Verification is essential and should not be seen as mistrustful (11,14,23)
Recommendations for Mentors & Advisors	**Rationale**
Create & include laboratory citation policies in the lab manual **1** Teach trainees about good citation practices **2** Engage trainees in the process of verifying citation accuracy for their publications	Ensures that there is a clear citation management plan so that trainees are learning best practices
Recommendations for Readers	**Rationale**
Raise awareness among scientists in your field about: **1** The importance of citation inaccuracies & best practices **2** Known problems with chains of inaccurate citations This may include discussing citation inaccuracies in post-publication peer-reviews or on social networks for scientists	Action from individual scientists is needed to promote field-wide improvements in practice and prevent propagation of errors

Table 8:

Options journal editors can consider to encourage good citation practices

Options for Journals		Rationale
The Instructions for Authors should include detailed guidelines, with recommendations for citing literature (for example, see “Recommendations for Authors” table)		Citation skills are rarely taught. Many authors are unaware of best practices.
Consider new policies and practices		
1	When submitting, ask authors to declare that they have checked all references for accuracy and have used primary references instead of indirect or secondary references	Emphasizes the need for good citation practices (26)
2	Restrict the number of references	May make it easier for authors to maintain an overview of what they cite (14,27,28)
3	Encourage editors and reviewers to check selected references	Random checks by editorial staff may remind authors about the importance of citation accuracy. Editors: (13,27); Reviewers: (17,29)
4	Inform authors that citation accuracy is expected and checked	Shows that the editorial staff is committed to good citation practices (26)
5	Institute a misquotations column to present cases of citation errors	Raise awareness of the consequences of inaccurate citations (28,30)
6	Editors may consider whether certain statements really need one or more than one reference, particularly in discussion sections	Pressure to have a citation for every statement may increase the risk that authors include unnecessary or inappropriate citations (31)