

What Proportion of Systematic Reviews and Meta-Analyses Published in the *Annals of Surgery* Provide Definitive Conclusions—A Systematic Review and Bibliometric Analysis

Matthew G. Davey ^{1,*}, Martin S. Davey ², Aoife J. Lowery ^{1,2} and Michael J. Kerin ^{1,2}

¹ Discipline of Surgery, The Lambe Institution of Translational Research, National University of Ireland, H91 YR71 Galway, Ireland; aoife.lowery@nuigalway.ie (A.J.L.); michael.kerin@nuigalway.ie (M.J.K.)

² Department of Surgery, Galway University Hospitals, H91 YR71 Galway, Ireland; martindavey@rcsi.ie

* Correspondence: m.davey7@nuigalway.ie; Tel.: +353-91524411

Abstract: Objective: To perform a systematic review and bibliometric analysis of systematic reviews and meta-analyses published in the *Annals of Surgery* during a 10-year eligibility period and determine the unambiguity of concluding statements of these reviews published in the journal. **Background:** Systematic reviews and meta-analyses integrate clinically pertinent results from several studies to replicate large-volume, ‘real world’ scenarios. While the assimilation of results from multiple high-quality trials are at the summit of the evidence-base, the increasing prevalence of reviews using low-to-moderate levels of evidence (LOE) limit the ability to make evidence-based conclusions. In surgery, increasing LOE are typically associated with publication in the highest impact surgical journals (e.g., *Annals of Surgery*). **Methods:** A systematic review was performed as per PRISMA guidelines. An electronic search of the *Annals of Surgery* for articles published between 2011 and 2020 was conducted. Descriptive statistics were used. **Results:** In total, 186 systematic reviews (with or without meta-analyses) were published in the *Annals of Surgery* between 2011 and 2020 (131 systematic reviews with meta-analyses (70.4%) and 55 without meta-analyses (29.6%)). Study data were from 22,656,192 subjects. In total, 94 studies were from European research institutes (50.5%) and 58 were from North American institutes (31.2%). Overall, 75.3% of studies provided conclusive statements (140/186). Year of publication ($P = 0.969$), country of publication ($P = 0.971$), region of publication ($P = 0.416$), LOE ($P = 0.342$), surgery performed ($P = 0.736$), and two-year impact factor (IF) ($P = 0.251$) failed to correlate with conclusive statements. Of note, 80.9% (106/131) of meta-analyses and 61.8% of systematic reviews (34/55) provided conclusive statements ($P = 0.009$, \dagger). **Conclusions:** Over 75% of systematic reviews published in the *Annals of Surgery* culminated in conclusive statements. Interestingly, meta-analyses were more likely to provide conclusive statements than systematic reviews, while LOE and IF failed to do so.

Keywords: systematic review; meta-analysis; academic surgery



Citation: Davey, M.G.; Davey, M.S.; Lowery, A.J.; Kerin, M.J. What Proportion of Systematic Reviews and Meta-Analyses Published in the *Annals of Surgery* Provide Definitive Conclusions—A Systematic Review and Bibliometric Analysis. *Publications* **2022**, *10*, 19. <https://doi.org/10.3390/publications10020019>

Academic Editor: Oleg V. Mikhailov

Received: 16 February 2022

Accepted: 25 April 2022

Published: 28 April 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Synthetic reviews (i.e., systematic reviews and meta-analyses) involve a thorough interrogation of studies published by previous authors to provide a comprehensive consensus based on real-world findings in relation to a predetermined research question. The value of such studies is their ability to integrate clinically pertinent results from several studies or trials, using the robustness of larger data to inform results, outcomes, and overarching consensus. In the world of surgery, the overarching intention of such analysis is to synthesize realistic, large-volume approximations of clinical reality, which may then inform best-practice for prospective candidates requiring surgical interventions.

While the assimilation of results from multiple well-designed, high-quality trials (i.e., randomized controlled trials, or RCTs) are placed at the peak of the evidence-base [1], there has been a recent increase in the number of systematic reviews and meta-analyses

being performed to integrate low-to-moderate levels of evidence (LOE), with the ambition to provide consensus from the included data [2,3]. However, higher risks of bias, lower methodological quality, and heterogeneous results impact the validity and the meaningfulness of the conclusions which may be drawn. Despite this, the publication of such studies continues to increase, which often leads to the authors of such studies being unable to report definitive results of their synthetic review. Thus, many published systematic reviews and meta-analyses have evoked scrutiny among expert members of the scientific community and has led to the growing perception that the mass production of such articles has reached ‘epidemic proportions’ [4]. Moreover, some consider the majority of such articles to be ‘unnecessary, misleading, and/or conflicted’, leading to the dissemination of redundant and uninformative data [4]. An illustrative example of this conceptualization has been captured in the recent work by Harris et al., which was published in *Arthroscopy* [5]. Following their extensive review of what the authors considered to be the six top-ranking orthopedic science journals, the authors concluded that nearly one-third of published systematic reviews and meta-analysis provided ambiguous conclusions.

The authors of the current study acknowledge the reputation of the *Annals of Surgery* and recognize the quality and importance of the data published in this top-ranking journal in the field of surgery. While Harris et al. focused solely on investigating the conclusiveness of orthopedic research, the authors of the current study sought to evaluate whether this concept was pertinent to the synthetic reviews on topics published in the *Annals of Surgery*. Accordingly, the primary aim of the current study was to assess the concluding statements of systematic reviews and meta-analyses published in the *Annals of Surgery* to assess the impact on the concluding statements on the practice of surgery. In surgery, increasing LOE is typically associated with publication in the highest impact surgical journals (i.e., *Annals of Surgery*). Therefore, our secondary aim was to determine the impact of LOE and other factors (such as country of origin, year of publication, type of review, etc.) on the concluding statements from the studies published in the journal. Our hypothesis was that at least 50% of included systematic reviews and meta-analyses would provide comprehensive conclusive statements and that LOE would correlate with the conclusive statements of studies, due to the high quality of research articles published in the *Annals of Surgery*.

2. Methods

2.1. Preparation and Study Criteria

Studies meeting the following inclusion criteria were included in this analysis: (1) articles had to be published as full-text manuscripts in the *Annals of Surgery*; (2) articles had to have been published between the years 2011 and 2020; and (3) articles must have been a systematic review with or without meta-analysis. Studies meeting the following criteria were excluded from this analysis: (1) articles not published in the *Annals of Surgery* journal; (2) any study that was not a systematic review with or without meta-analysis; (3) studies published outside the determined search period; or (4) published conference proceedings or abstracts (including proceedings from the European Surgical Association and the American Surgical Association annual conferences).

This review was not prospectively registered with the international prospective register of systematic reviews (PROSPERO) as the results of this review do not have a direct link to human health.

2.2. Search Strategy

A systematic review was performed in accordance with the preferred reporting items for systematic reviews and meta-analysis (PRISMA) checklist [6]. A formal search was performed by two independent reviewers using the predefined search strategy, which was designed by the senior author (M.J.K.). The first and second authors (M.G.D. and M.S.D.) conducted a comprehensive manual electronic search of the *Annals of Surgery* journal for systematic reviews and meta-analyses published in the journal between the years 2011 and 2020. All titles published in the journal were initially screened, and all systemic reviews

(with or without meta-analysis) published in the journal were included and had their abstracts and full texts reviewed. Each reviewer read the title of each manuscript published in the *Annals of Surgery* between the years 2011 and 2020 and identified whether the studies were systematic review with or without meta-analysis. Each reviewer read the retrieved manuscripts to ensure all inclusion criteria was met, before extracting the following data: (1) first author name; (2) year of publication; (3) country; (4) region; (5) level of evidence; (6) whether it was a systematic review or meta-analysis; (7) number of included patients or participants; (8) the two-year impact factor (IF) of each systematic review; and (9) the conclusiveness of each study. In case of discrepancies in opinion between both reviewers, a third reviewer was asked to arbitrate (A.J.L.).

2.3. Definitions

- The hierarchical levels of evidence-based medicine (LOE) were considered in accordance to the previous work of Nguyen et al. [7]. In brief, level I evidence consisted of high-quality RCTs which were adequately powered and the systematic reviews of such studies. Level II studies consisted of lesser quality RCTs and predominantly consisted of prospective cohort studies, and systematic reviews of those studies. Level III studies consisted of retrospective comparative studies. Level IV studies were typically of the case-series variety, and level V articles were usually case reports or expert opinions.
- ‘Higher level of evidence’ including systematic reviews and meta-analyses which included prospective studies and RCTs only.
- Systematic reviews included and were not limited to pooled analyses (without meta-analysis).
- Included meta-analyses included those of network meta-analysis methodology.
- When reporting two-year IF, this was objectively measured as the number of manuscripts citing the study in the first two years from the month of publication, as linked and available through the PubMed electronic database.
- For synthetic reviews included, which included studies of varying LOE, the study with the lowest included LOE was used to represent the LOE of the synthetic review.
- Conclusive conclusions were concluding statements to a synthetic review which provided a clear, concise, and informative message based on the results of the synthetic review as adjudicated by the independent reviewers. Studies reporting the requirement for ‘further’ investigation or research were considered to be inconclusive.

2.4. Statistical Analyses

Descriptive statistics were used to determine the association between the study details and conclusiveness of studies. Chi-squared (χ^2) and Fisher’s exact (†) tests were used, as appropriate [8]. Differences in two-year IF between conclusive and inconclusive studies were measured using independent samples *t*-test (‡). Subgroup analysis was performed based on region of publication, surgical specialties, type of study (i.e., systematic review or meta-analysis), and on LOE. All tests of significance were two-tailed, with $P < 0.050$ indicating statistical significance. Data were analyzed using SPSS™ (IBM SPSS Statistics for Mac, Version 26.0. Armonk, NY, USA) version 26.

3. Results

3.1. Study Characteristics

In total, 186 systematic reviews (with or without meta-analyses) were published in the *Annals of Surgery* between the years 2011 and 2020. These studies included data in relation to an estimated 22,656,192 subjects/patients. Overall, 94 of the included studies were from European research institutes (50.5%), 58 from North American research institutes (31.2%), 18 from Asian research institutes (9.7%), and 16 from research institutes in Australia and New Zealand (8.6%). Research facilities in the United Kingdom published the most systematic reviews and meta-analyses (18.8%, 35/186), followed closely by the United States (16.1%, 30/186), and Canada (15.1%, 28/186). The mean number of published systematic reviews and meta-analyses was 18.6 per year (median 19, range: 15–25 studies).

The majority of included studies were level III evidence (59.7%, 111/186), with 24.2% of included studies providing level I evidence (45/186). General and gastrointestinal surgery was the most common type of surgery with systematic reviews and meta-analyses published in the *Annals of Surgery* (55.4%, 103/186). In total, there were 131 systematic reviews with meta-analyses (70.4%) and 55 systematic reviews without meta-analyses (29.6%) included. Study characteristics from the 186 included studies are outlined in Supplementary Material Table S1.

3.2. Conclusive Conclusions

Of the 186 included systematic reviews and meta-analyses, 75.3% provided conclusive evidence in their conclusions (140/186). Year of publication ($P = 0.969$, χ^2), country of publication ($P = 0.971$, χ^2), region of publication ($P = 0.416$, χ^2), overall LOE ($P = 0.342$, χ^2), higher LOE ($P = 0.465$, †), type of surgery ($P = 0.736$, χ^2), and two-year IF ($P = 0.251$, ‡) were not associated with yielding conclusive statements in their publications in the *Annals of Surgery* (Supplementary Material Table S2). Meta-analyses were more likely to yield conclusive conclusions than systematic reviews ($P = 0.009$, †).

3.3. Subgroup Analyses—Region of Publication

We performed a subgroup analysis based on region of publication of the synthetic reviews included in this study. When evaluating each region independently, the LOE failed to significantly impact the conclusiveness of studies, irrespective of region (all $P > 0.050$). Additionally, when analyzing studies published from Australia and New Zealand, 100.0% of studies performed in the fields of breast surgery, academic surgery, and gastrointestinal surgery all yielded conclusive conclusions ($P = 0.026$, χ^2). For each of the other regions, surgical specialty failed to impact the conclusiveness of studies performed (all $P > 0.050$). For studies published from European surgical facilities, meta-analyses trended towards significance for being more likely to yield conclusive conclusions ($P = 0.074$, χ^2). The type of study performed failed to impact the conclusiveness of studies published from other regions (Supplementary Material Table S3).

3.4. Subgroup Analyses—Level of Evidence

When performing a subgroup analysis based on the LOE, surgical specialty failed to significantly impact the conclusiveness of studies included in this systematic review (all $P > 0.050$, χ^2). For studies included that were of level III evidence, meta-analyses were significantly more likely to provide conclusive conclusions compared to traditional systematic reviews ($P = 0.016$, †). Otherwise, the type of study performed failed to impact the conclusiveness of the studies (all $P > 0.050$, †). All correlations between other subgroups and LOE are outlined in Supplementary Material Table S3.

3.5. Subgroup Analyses—Study Type

In this study, the type of study (systematic review or meta-analysis) failed to influence the conclusiveness of studies based on surgical specialty (both $P > 0.050$, χ^2). For systematic reviews and meta-analyses independently, all other study parameters failed to impact the conclusiveness of included studies, as outlined in Supplementary Material Table S3.

4. Discussion

The most important finding in this systematic review of systematic reviews and meta-analyses published in the *Annals of Surgery* over a 10-year eligibility period is that over 75% of the 186 included studies that yielded conclusive conclusions to their articles. This result highlights the value of synthetic reviews published in the *Annals of Surgery* and supports the authors' null hypothesis suggesting that over 50% of such studies published in the journal would provide indecisive conclusive statements. These results support the journal as one that provides strong definitive conclusions on most synthetic reviews, particularly when compared to similar, previously conducted studies (e.g., Harris et al. reported one in three

studies that failed to provide definitive conclusions in their previous analysis of orthopedic literature). Interestingly, study characteristics (such as country of publication, region of publication, LOE, type of surgery, and two-year IF) failed to inform the conclusiveness of published synthetic reviews during this time period in the *Annals of Surgery*. Conversely, meta-analyses published in the journal were more likely to yield conclusive conclusions when compared to traditional systematic reviews ($P = 0.009$, †). While the number of synthetic reviews published in the *Annals of Surgery* increased marginally during the 10-year eligibility period, the proportion of studies with conclusive conclusions remained stable ($P = 0.969$, χ^2), indicating consistency of these published studies during this time period.

As previously outlined, more than three-quarters of systematic reviews and meta-analyses published in this journal provided conclusive statements to their study. This is an interesting finding, albeit one that is somewhat predictable as the *Annals of Surgery* has traditionally been renowned as the most prestigious academic journal in the field of academic surgery and is consistently ranked as the highest ranking surgical journal using both SCImago and Resurchify journal ranking metrics [9,10]. Furthermore, Agha et al. previously reported the median IF of 1.526 for the 193 surgical Thomas Reuters Journal Citation Reports (2014) [11], which is considerably lower than the current IF for the *Annals of Surgery* (IF of the journal at the time of writing is 12.969 (2021)). Therefore, it is fair to assume that synthetic reviews of higher LOE are more likely to be published in this journal, which one may intuitively expect to impact the authors' likelihood to provide conclusive statements to their review.

Of note, both LOE and the two-year IF failed to correlate with the conclusiveness of studies published in the *Annals of Surgery* in this systematic review. These are interesting findings; members of the academic community have the tendency to rely on IF as a proxy of the quality of a journal compared to competing journals in the same field [12], with the *Annals of Surgery* being considered among those at the summit of surgical journals internationally. Panesar et al. previously established that just 5.6% of studies published in four of the highest ranking surgical journals by IF (*Annals of Surgery*, *Archives of Surgery*, *British Journal of Surgery*, and *Annals of the Royal College of Surgeons*) are RCTs in design (63/1135) [13]. This implies that in the absence of well-designed RCTs being published in such journals, it is plausible that there is an overall 'dilution' of the quality of published studies, with those of moderate methodological quality being published [14]. Even in a high-ranking journal, such as the *Annals of Surgery*, this 'dilution' is somewhat evident in the results of the current study. Overall, 69.4% of included systematic reviews and meta-analyses were of level III evidence or lower (129/186), with just 21.2% of studies representing level I evidence (45/186), a finding that was surprising given the high-ranking and reputable profile of the journal.

Interestingly, meta-analyses published in the *Annals of Surgery* were significantly more likely to provide conclusions than traditional systematic reviews ($P = 0.009$, †). Moreover, meta-analyses of level III evidence were significantly more likely to be conclusive than traditional systematic review articles ($P = 0.016$, †). This emphasizes the value of utilizing meta-analysis methodology in providing consensus for surgical research [15]. While both systematic reviews and meta-analyses are useful in integrating data in large volumes from several clinical studies to replicate 'real world' scenarios, meta-analyses have the advantage of providing outcome measures which provide a more precise estimate of the treatment effect when compared to simple pooled analyses used to report results in conventional systematic reviews [16]. Moreover, meta-analyses also examine the degree of variability (or 'heterogeneity') of included data, while accurately establishing the impact of treatment effects, which is crucial in the field of academic surgery, in order to provide greater scientific rationalization of results yielded. This is evident from the results of the current study where 80.9% of meta-analyses versus 61.8% of systematic reviews provide conclusive statements to their analyses ($P = 0.009$, †). Thus, this study supports the use of meta-analysis methodology where feasible in order to improve the ability to provide conclusive results to surgical research questions.

5. Limitations

The current systematic review of systematic reviews and meta-analyses published in the *Annals of Surgery* suffers from several limitations. Firstly, and most importantly, the authors of this study subjectively adjudicated the conclusiveness of the published systematic reviews and meta-analyses in the journal, without applying a reliable scoring system to fairly judge ‘conclusiveness’. This is due to the requirement of several methodologies used to score ‘conclusiveness’, requiring the analyst to have the raw data from the study. This is an obvious shortcoming of this study. Secondly, although this analysis was performed by two independent reviewers, the study design makes presumptions as to the reliability of this assessment, with no formal appraisal of intra- and inter-observer agreement. Thirdly, the failure for LOE and IF to influence the conclusiveness of studies published may be considered a blunt instrument when determining the actual clinical impact these studies may have on influencing or challenging current practice in the field of surgery. Finally, this study fails to evaluate the overall methodology or the risk of bias/quality assessments of the 186 included studies, which limits the conclusions which may successfully be drawn from the data presented in the current study.

6. Conclusions

In conclusion, over 75% of systematic reviews and meta-analyses published in the *Annals of Surgery* during a 10-year period yielded conclusive conclusions statements. Study characteristics, such as country of origin, region of origin, surgical specialty, LOE, and two-year IF, failed to impact the conclusiveness of published studies. Interestingly, meta-analyses were more likely to provide conclusive statements than systematic reviews. This systematic review emphasizes the value of performing comprehensive synthetic reviews capable of providing informative data to the reader, which may be interpreted in a ‘conclusive’ manner to impact clinical practice.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/publications10020019/s1>, Table S1: List of included systematic review and meta-analysis included in this systematic review by year of publication, Table S2: Correlations between study characteristics and the conclusiveness of systematic review and meta-analyses published in the *Annals of Surgery* which were included in the current study, Table S3: Correlations between study characteristics based on subgroups and the conclusiveness of systematic review and meta-analyses published in the *Annals of Surgery* which were included in the current study. Full reference list for the included 186 studies available in S4.

Author Contributions: Conceptualization, M.G.D. and M.S.D.; methodology, M.G.D. and M.S.D.; software, M.G.D.; validation, M.G.D., M.S.D. and A.J.L.; formal analysis, M.G.D.; data curation, M.G.D. and M.S.D.; writing—original draft preparation, M.G.D.; writing—review and editing, M.G.D., M.S.D., A.J.L. and M.J.K.; visualization, M.G.D. and M.J.K.; supervision, M.J.K.; funding acquisition, M.G.D. All authors have read and agreed to the published version of the manuscript.

Funding: M.G.D. received funding from the National Breast Cancer Research Institute, Ireland.

Data Availability Statement: Data made available upon reasonable request from the corresponding author (M.G.D.).

Conflicts of Interest: None of the authors have any conflict of interest.

References

1. Burns, P.B.; Rohrich, R.J.; Chung, K.C. The levels of evidence and their role in evidence-based medicine. *Plast. Reconstr. Surg.* **2011**, *128*, 305–310. [[CrossRef](#)] [[PubMed](#)]
2. Juhl, C.B.; Lund, H. Do we really need another systematic review? *Br. J. Sports Med.* **2018**, *52*, 1408–1409. [[CrossRef](#)] [[PubMed](#)]
3. Lund, H.; Juhl, C.; Christensen, R. Systematic reviews and research waste. *Lancet* **2016**, *387*, 123–124. [[CrossRef](#)]
4. Ioannidis, J.P. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q.* **2016**, *94*, 485–514. [[CrossRef](#)]
5. Harris, J.D.; Cote, M.P.; Dhawan, A.; Hohmann, E.; Brand, J.C. Nearly One-Third of Published Systematic Reviews and Meta-analyses Yield Inconclusive Conclusions: A Systematic Review. *Arthroscopy* **2021**, *37*, 2991–2998. [[CrossRef](#)] [[PubMed](#)]

6. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* **2009**, *339*, b2535. [[CrossRef](#)] [[PubMed](#)]
7. Nguyen, A.; Mahabir, R.C. An Update on the Level of Evidence for Plastic Surgery Research Published in Plastic and Reconstructive Surgery. *Plast. Reconstr. Surg. Glob. Open* **2016**, *4*, e798.
8. Kim, H.Y. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restor. Dent. Endod.* **2017**, *42*, 152–155. [[CrossRef](#)] [[PubMed](#)]
9. Scimago Journal & Country Rank. 2021. Available online: <https://www.scimagojr.com/journalrank.php?category=2746> (accessed on 25 November 2021).
10. Resurchify. 2021. Available online: https://www.resurchify.com/find/?query=surgery#search_results (accessed on 25 November 2021).
11. Agha, R.A.; Barai, I.; Rajmohan, S.; Lee, S.; Anwar, M.O.; Fowler, A.J.; Orgill, D.P.; Altman, D.G. Support for reporting guidelines in surgical journals needs improvement: A systematic review. *Int. J. Surg.* **2017**, *45*, 14–17. [[CrossRef](#)] [[PubMed](#)]
12. Rawat, S. How is impact factor impacting our research? *Biomed. J.* **2014**, *37*, 415–416. [[CrossRef](#)] [[PubMed](#)]
13. Panesar, S.S.; Thakrar, R.; Athanasiou, T.; Sheikh, A. Comparison of reports of randomized controlled trials and systematic reviews in surgical journals: Literature review. *J. R. Soc. Med.* **2006**, *99*, 470–472. [[CrossRef](#)] [[PubMed](#)]
14. Sanborn, A.N.; Noguchi, T.; Tripp, J.; Stewart, N. A dilution effect without dilution: When missing evidence, not non-diagnostic evidence, is judged inaccurately. *Cognition* **2020**, *196*, 104110. [[CrossRef](#)] [[PubMed](#)]
15. Knight, S.R. The Value of Systematic Reviews and Meta-Analyses in Surgery. *Eur. Surg. Res.* **2021**, *62*, 221–228. [[CrossRef](#)] [[PubMed](#)]
16. Haidich, A.B. Meta-analysis in medical research. *Hippokratia* **2010**, *14* (Suppl. 1), 29–37. [[PubMed](#)]