Relative Citation Impact of Various Study Designs in the Health Sciences

Nikolaos A. Patsopoulos, MD
Apostolos A. Analatos
John P. A. Ioannidis, MD

Several authors and organizations have proposed hierarchies of evidence, based on the relative reliability of various types of study designs.1-4 Although many people recognize that expert opinions and nonsystematic reviews provide the least reliable level of information,5,6 such articles continue to have a massive influential presence.7 Controlled studies assume higher places in hierarchies of evidence than uncontrolled studies, and randomized trials are considered the gold standard for clinical research.1-4 However, randomized trials cannot be conducted for all questions of interest8 and there is debate on whether they give different results than nonrandomized studies.9-14 Finally, meta-analyses are becoming increasingly frequent in the literature. Meta-analyses are often placed at the highest level of evidence,1,4 despite their critics.15,16 No hierarchy of evidence is unanimously accepted.

An important issue is whether the impact of various studies is different and changing over time. Impact on clinical practice and decision making is difficult to measure comprehensively. However, one important measure of impact is the use of citations in the published literature. Citations have limitations,17 but they provide an objective measurement of how often scientists use a specific published work. One may ask: What is the relative citation impact of published articles using various types of designs? Is this impact commensurate with the proposed hierarchies of evidence? Has it changed over time? We aimed to answer these questions using citation analysis.

**Context** The relative merits of various study designs and their placement in hierarchies of evidence are often discussed. However, there is limited knowledge about the relative citation impact of articles using various study designs.

**Objective** To determine whether the type of study design affects the rate of citation in subsequent articles.

**Design and Setting** We measured the citation impact of articles using various study designs—including meta-analyses, randomized controlled trials, cohort studies, case-control studies, case reports, nonsystematic reviews, and decision analysis or cost-effectiveness analysis—published in 1991 and in 2001 for a sample of 2646 articles.

**Main Outcome Measure** The citation count through the end of the second year after the year of publication and the total received citations.

**Results** Meta-analyses received more citations than any other study design both in 1991 (P<.05 for all comparisons) and in 2001 (P<.001 for all comparisons) and both in the first 2 years and in the longer term. More than 10 citations in the first 2 years were received by 32.4% of meta-analyses published in 1991 and 43.6% of meta-analyses published in 2001. Randomized controlled trials did not differ significantly from epidemiological studies and nonsystematic review articles in 1991 but clearly became the second-cited study design in 2001. Epidemiological studies, nonsystematic review articles, and decision and cost-effectiveness analyses had relatively similar impact; case reports received negligible citations. Meta-analyses were cited significantly more often than all other designs after adjusting for year of publication, high journal impact factor, and country of origin. When limited to studies addressing treatment effects, meta-analyses received more citations than randomized trials.

**Conclusion** Overall, the citation impact of various study designs is commensurate with most proposed hierarchies of evidence.

**METHODS**

**Identification and Eligibility of Relevant Studies**

We compared the citation impact across various study designs and between studies published in 1991 and 2001. We searched the Institute for Scientific Information (ISI) Science Citation Index at the Web of Science Database (www.isinet.com) for meta-analyses, randomized controlled trials (RCTs), cohort studies, case-control studies, case reports, nonsystematic reviews, and decision analysis or cost-effectiveness analysis records published in 1991 and 2001. These types of publications cover the major, readily identifiable designs used in collecting data. Citations have limitations,17 but they provide an objective measurement of how often scientists use a specific published work. One may ask: What is the relative citation impact of published articles using various types of designs? Is this impact commensurate with the proposed hierarchies of evidence? Has it changed over time? We aimed to answer these questions using citation analysis.
and synthesizing medical information. Secondarily, meta-analyses were also classified as meta-analyses including RCTs vs others. Both meta-analyses and RCTs were also classified according to their subject or purpose (treatment effect [therapy or prevention], prognosis, diagnosis, and etiology or association for meta-analyses; treatment effect and diagnosis for RCTs).

It is impractical to identify and analyze all the tens of thousands of publications fitting in these study designs. Often it is impossible to accurately classify the study design unless the whole article is carefully scrutinized. Sometimes even this may not suffice. Thus we used a strategy that aimed to yield an adequate number of relevant publications for each design with high specificity in characterizing design. The search strategies for each type of publication sought the appearance of the relevant study design terms in the article title (TI). Meta-analysis was searched with TI=meta-analysis* or metaanaly*, randomized controlled trial with TI=random* AND TI=trial, decision analysis or cost-effectiveness analysis with TI=decision analy* OR TI=cost effectiveness analy* OR TI=cost-effectiveness analy*, nonsystematic review with TI=review NOT TI=systemat* NOT TI=meta-analy* NOT TI=overview NOT TI=case report*, case-control study with TI=case control study, cohort study with TI=cohort study, and case report with TI=case report NOT TI=review NOT TI=overview. When the search algorithm yielded an excessive number of articles, we screened systematically 1:5 or 1:10 batches of records, for study designs with 1200 to 3000 records and more than 3000 records retrieved in a year, respectively.

Two investigators (N.A.P. and A.A.A.) independently screened both the title and abstract of identified articles. Articles were eligible if they represented applications of the type of study design under which they were identified. We excluded ISI records without abstract; letters; editorials; news and meeting abstracts; methodology-and-theory articles; and articles not on human subjects or material, not on health, or both. Discrepancies were discussed between investigators; a third investigator (J.P.A.I.) resolved disagreements.

Data Extraction
For each article eligible for citation analysis, we recorded total citations until December 10, 2004; citations received up to the end of the second year after publication (1991-1993 and 2001-2003, respectively); country of authors; and journal.

Analysis
The main analyses addressed citation counts for 1991-1993 and 2001-2003 (early citations). Most articles are rarely cited, if at all, during the same year in which they were published, but the citation count of the 2 subsequent years is representative (it forms the basis of estimating journal impact factors). Secondary analyses counted total citations until December 10, 2004 (long-term impact); this time frame unavoidably differed between the 1991 and 2001 publication cohorts.

Citation counts per publication type and year were summarized with medians and interquartile ranges (citation distributions are left-skewed).\(^{18}\) Mann-Whitney U tests and Kruskal-Wallis analysis of variance compared 2 or several groups, respectively. We also identified articles that received more than 10 citations in the first 2 years (approximately the top 10% most-cited ISI-indexed articles in Clinical Medicine).\(^{19}\) Logistic regressions addressed the year and type of publication (dummy variables) as predictors of more than 10 citations in 2 years, adjusting also for country of authors and high journal impact factor.

Analyses were conducted using SPSS statistical software version 12.0 (SPSS Inc, Chicago, Ill). \(P\) values are 2-tailed. Statistical significance was considered at the .05 level.

RESULTS
We identified 17 813 articles (6052 from 1991, 11 761 from 2001) and screened 5769 of those for eligibility (1936 and 3833, respectively); 2646 articles (904 and 1742, respectively) were eligible for citation counting (Table 1).
Table 1. Screened and Eligible Studies and Citation Counts Per Study Design and Year of Publication*  

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Articles, No. (%)</th>
<th>Articles, No. (%)</th>
<th>Articles, No. (%)</th>
<th>Articles, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-analysis</td>
<td>71 (171)</td>
<td>266 (608)</td>
<td>40 (56)</td>
<td>135 (51)</td>
</tr>
<tr>
<td>Randomized controlled trials</td>
<td>328 (513)</td>
<td>264 (420)</td>
<td>127 (39)</td>
<td>85 (32)</td>
</tr>
<tr>
<td>Cohort</td>
<td>92 (224)</td>
<td>401 (1023)</td>
<td>45 (49)</td>
<td>131 (33)</td>
</tr>
<tr>
<td>Case-control</td>
<td>172 (302)</td>
<td>361 (659)</td>
<td>57 (33)</td>
<td>118 (33)</td>
</tr>
<tr>
<td>Case report</td>
<td>167 (320)</td>
<td>249 (463)</td>
<td>44 (26)</td>
<td>57 (23)</td>
</tr>
<tr>
<td>Nonsystematic review</td>
<td>44 (310)</td>
<td>124 (480)</td>
<td>18 (41)</td>
<td>58 (47)</td>
</tr>
<tr>
<td>Decision or cost-effectiveness analysis</td>
<td>30 (96)</td>
<td>77 (180)</td>
<td>22 (73)</td>
<td>39 (51)</td>
</tr>
</tbody>
</table>

*The total number of retrieved articles in the case report search was 1605 in 1991 and 2343 in 2001 and the total number of retrieved articles in the nonsystematic review search was 3141 in 1991 and 4811 in 2001 while the total number of retrieved articles in the randomized controlled trial search was 2139 in 2001. For these study designs and years, statistically 1 in 5 or 1 in 10 batches of records were screened (see “Methods” section for more details). The x coefficient between the 2 investigators for eligibility of the screened articles was 0.94. Final reasons for exclusions included: no abstract (n = 352), letter (n = 239), editorial (n = 159), news or meeting abstract (n = 1082), methods or theory article (n = 277), not a practical application of the study design being searched (n = 720), not on human health (n = 294).

†Journal impact factor exceeding 10 according to the latest ratings (Institute for Scientific Information, Journal Citation Reports 2003).

‡For the comparison of the 1991 vs 2001 publication cohort in terms of the 2-year citation count.

Figure 1. Percentage of Articles by Study Design With at Least 10 Citations in First 2 Years of Publication

RCTs, OR, 0.49; 95% CI, 0.36-0.68; for cohort studies, OR, 0.46; 95% CI, 0.34-0.63; for case-control studies, OR, 0.37; 95% CI, 0.27-0.52; for case reports, OR, 0.01; 95% CI, 0.00-0.04; for nonsystematic reviews, OR, 0.47; 95% CI, 0.31-0.73; and for decision or cost-effectiveness analysis articles, OR, 0.29; 95% CI, 0.16-0.51 vs meta-analyses).

Long-term Impact

Both in 1991 and in 2001, there was a statistically significant difference in citation count between various designs (P<.001, FIGURE 2). For 1991, meta-analyses were statistically significantly cited more times than all other designs (P<.05 for all comparisons). Conversely, RCTs had significantly more citations only from case reports (P<.001) and possibly decision or cost-effectiveness analysis articles (P=.05). However, for 2001, meta-analyses had greater impact than all other designs (P<.001 for all comparisons) and RCTs were cited significantly more times than all the remaining designs (P<.05 for all comparisons).

Case reports had once again a very low impact (P<.001 for all comparisons). All other comparisons of designs were not statistically significant.

Subgroups

Citations of subgroups of meta-analyses and RCTs are shown in TABLE 2. There were no statistically significant differences in the citations received by meta-analyses including or not including RCTs, both in 1991 and 2001 and both for the 2 years and for the long term (P=.08 for 2-year citations, P=.10 for long-term citations), and the difference became more clear in 2001 (P=.001 and P=.001, respectively).

Comment

The citation impact of various study designs follows the order proposed by most current theoretical hierarchies of evidence.1-3 On average meta-analyses currently receive more citations than any other type of study design. Meta-analyses have clearly surpassed in citation impact both decision or cost-effectiveness analysis articles and RCTs,
against which they had mostly modest differences, if any, in the early 1990s. Although RCTs have become the second most cited study design, decision or cost-effectiveness analysis has not followed this growth. Epidemiological studies are now lagging behind randomized research; however, this was not as evident for articles published in 1991. Non-systematic reviews continue to have a citation impact similar to that of epidemiological studies. Finally, case reports have negligible impact.

The superiority in citation impact of meta-analyses and secondarily RCTs is consistent with the prominent role given to these designs by evidence-based medicine,1-4 despite the criticisms leveled against both designs.15,20 The further dissemination of hierarchies of evidence may further increase the citations for meta-analyses and RCTs. If the proposal that each study should start and end when a meta-analysis is adopted,21 meta-analyses may become even more highly cited. Interestingly, high citations for meta-analyses extend to meta-analyses of nonrandomized research. Of course, we acknowledge that primary studies are required for quantitative syntheses ever to be performed.

The relative impact of epidemiological research has lost ground recently.

Table 2. Citations of Subgroups of Meta-analyses and Randomized Controlled Trials

<table>
<thead>
<tr>
<th></th>
<th>Median (Interquartile Range)</th>
<th>Two-Year Citations</th>
<th>Long-term Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-analyses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per inclusion of randomized controlled trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>39 (55)</td>
<td>146 (55)</td>
<td>5 (1-13)</td>
</tr>
<tr>
<td>No</td>
<td>32 (45)</td>
<td>120 (45)</td>
<td>6 (3-17)</td>
</tr>
<tr>
<td>Per subject or purpose*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment effect</td>
<td>46 (65)</td>
<td>164 (62)</td>
<td>6 (2-15)</td>
</tr>
<tr>
<td>Prognosis</td>
<td>4 (6)</td>
<td>18 (7)</td>
<td>10 (2-24)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>1 (1)</td>
<td>22 (8)</td>
<td>2</td>
</tr>
<tr>
<td>Etiology or association</td>
<td>18 (25)</td>
<td>62 (23)</td>
<td>3 (3-8)</td>
</tr>
<tr>
<td>Randomized Trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment effect</td>
<td>323 (98)</td>
<td>261 (99)</td>
<td>4 (1-10)</td>
</tr>
<tr>
<td>Diagnosis†</td>
<td>5 (2)</td>
<td>3 (1)</td>
<td>1 (1-9)</td>
</tr>
</tbody>
</table>

*Two meta-analyses on the physiological range of blood pressure published in 1991 were not counted in any category.  
†Clinical trials randomizing participants to 2 diagnostic methods and comparing accuracy.
Perhaps there is increasing uncertainty due to the refutation of several key cohort studies on important questions such as vitamins or hormone therapy. \(^1\) Decision or cost-effectiveness analysis has also not managed to keep a high impact. Nevertheless, many important questions simply cannot be answered with randomized research.

Also many nonsystematic reviews continue to be published. In our study, we excluded nonquantitative reviews that seemed to use some systematic approaches. Empirical evaluations of orthopedic and general medical journals have shown that systematic reviews received double the number of citations compared with nonsystematic ones.\(^2\) Efforts to enhance the accuracy and usefulness of all reviews are important because even nonsystematic expert reviews are still extensively read by practitioners.\(^2\)

Some caveats should be discussed. First, higher citation rates in articles published in 2001 than in those published in 1991 probably reflect simply the increase of journals worldwide (especially journal articles listed by ISI). Second, we excluded several types of reports such as nonhuman studies and hybrid designs (eg, reports describing both cohort and case-control studies). However, we wanted to focus sharply on the key study designs. Third, we did not exclude self-citations. Fourth, we used very strict screening criteria to ensure high specificity for characterizing study designs. Most studies probably still do not mention their design in their title. It is unknown whether among studies having the same design, those that state it in the title would get more citations or less. Nevertheless, even if such differences exist, they probably would not affect selectively some study designs over others.

Finally, a citation does not guarantee the respect of the citing investigators. A study may be cited only to be criticized or dismissed. Nevertheless, citation still means that the study is active in the scientific debate. Moreover, we should acknowledge that citation impact does not necessarily translate into clinical or scientific impact, but this is extremely difficult to measure and could vary on a case-by-case basis. Allowing for these caveats, our evaluation provides empirical evidence on the relative impact of various study designs in the literature.

**REFERENCES**