A Lesson From Neonatology

Yoav Littner, MD; Francis B. Mimouni, MD, FAAP, FACN; Shaul Dollberg, MD, FACN; Dror Mandel, MD

Objective: To test the hypothesis that articles with negative results are more likely than articles with positive results to be published in journals with lower impact factor.

Design and Setting: We selected all of the randomized, placebo-controlled trials conducted during the neonatal period between October 1, 1998, and October 1, 2003. Trials were classified as having positive results or negative results (significant or no significant difference, respectively). Only studies dealing with primary outcomes (efficacy) were included.

Main Outcome Measures: The impact factor of each journal was determined, and the sample size for each study was noted.

Results: There were 233 articles that fulfilled the inclusion criteria. There was a significant difference between the 2 groups in terms of impact factor ($P = .03$) but not sample size ($P = .30$). Impact factor correlated with both sample size and the type of study results (positive results vs negative results; $P < .05$).

Conclusion: Articles with negative results are more likely than articles with positive results to be published in journals with lower impact factor.

Arch Pediatr Adolesc Med. 2005;159:1036-1037

A recognized weakness of meta-analyses is the so-called “publication bias,” ie, the greater likelihood of a study with “positive” results (PR; significant difference between treatment groups) to be published. The impact factor (IF) has been developed as a means of assessing the quality of a journal (ie, the impact it has on the scientific community), is mostly based on the quotation rate of the articles published in it, and may vary from year to year. The higher it is, the more prestigious the journal. We tested the hypothesis that articles with negative results (NR) are more likely than articles with PR to be published in journals with lower IF.

Methods

We selected all of the randomized, placebo-controlled trials conducted during the neonatal period (birth to age 1 month) that were published between October 1, 1998, and October 1, 2003, using PubMed to search placebo, limiting to newborn, randomized controlled trials, and human. We classified the studies as having PR or NR (significant or no significant difference, respectively). Only the primary outcomes (efficacy), not secondary outcomes (such as adverse effects), were included. To ensure consistency, only 1 of us (Y.L.) reviewed the articles.

The IF of each journal was determined, and the sample size for each study was noted as a potential confounder. We gave an arbitrary classification of 0 for an IF when a journal was not included in the citation index.

Minitab version 13.1 software (Minitab Inc, State College, Pa) was used for statistical analyses. Since the distribution of IF was not normal and the variance of IF in studies with NR was significantly different ($P < .05$) from that in studies with PR, we used nonparametric tests (Kruskal-Wallis tests) to study the difference in IF and in sample size between groups. Regression analysis was used to study the independent effect of sample size and type of study (PR vs NR) on the ranking of IF. A $P$ value of less than .05 was considered significant.

Results

We identified 286 articles by the search, and 233 of them fulfilled the inclusion criteria. The primary outcome was identified in all of the articles. There was a significant difference between the 2 groups...
in terms of IF ($P = .03$), but not in terms of sample size ($P = .30$) (Table). In ranked multiple regression analysis using the IF rank as the dependent variable, both sample size and the type of study results (findings of PR vs NR) were significant ($P < .05$).

We found that in our field of expertise (newborn medicine), placebo-controlled, prospective clinical trials with NR (ie, no significant difference between control and study groups) are more likely to appear in peer-reviewed medical journals of lower IF than trials with PR (ie, treatment superior to placebo).

There are several possible explanations for this finding. One is related to publication bias. Indeed, a study with NR is less likely to be published than a study with PR. Although the IF of a journal is only 1 way of measuring its quality (and is a controversial one), the competition for acceptance in high-IF journals is fierce, and it is possible that reviewers and editors are biased toward articles with PR. In a recent study of articles submitted to JAMA, this was not found to be the case, as there was no statistically significant difference in the publication rates of submitted articles with positive vs negative results. This study theoretically indicates that the editorial board of JAMA is not likely to be influenced by the PR or NR of a study, but this may not be true for the editorial boards of other journals. The issue of publication bias is greater than just the problem of combining results in meta-analyses. It also directly impacts clinical care, as the published literature will be biased and the results in meta-analyses. It also directly impacts clinical care, as the published literature will be biased and the results in meta-analyses.

2 error, and increase the likelihood of having their articles accepted for publication. On the other hand, it is also possible that statistical design (eg, sequential analyses) or interim analyses lead to stopping trials with PR early. Alternatively, acceptance bias may favor the publication of trials with small sample sizes only if they have PR. Nevertheless, in the multiple regression analysis, the impact of PR vs NR on the IF of the journal in which the study was published remained significant, even after taking the sample size into account.

Accepted for Publication: April 7, 2005.
Correspondence: Shaul Dollberg, MD, Department of Neonatology, Lis Maternity Hospital, 6, Weitzman St, Tel Aviv, Israel 64239 (dollberg@tasmc.health.gov.il).

## REFERENCES

5. Seglen PO. Why the impact factor of journals should not be used for evaluating research. BMJ. 1997;314:498-502.

### Table. Impact Factor and Sample Size in Studies With Positive and Negative Results

<table>
<thead>
<tr>
<th>Impact factor</th>
<th>Mean ± SD</th>
<th>Median</th>
<th>Range</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies with positive results</td>
<td>4.567 ± 5.878</td>
<td>2.778</td>
<td>0-31.736</td>
<td>.03</td>
</tr>
<tr>
<td>Studies with negative results</td>
<td>4.242 ± 7.153</td>
<td>2.126</td>
<td>0-31.736</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Studies with positive results</td>
<td>421 ± 1525</td>
<td>80</td>
<td>8-11 619</td>
<td>.30</td>
</tr>
<tr>
<td>Studies with negative results</td>
<td>600 ± 2207</td>
<td>90</td>
<td>12-15 832</td>
<td></td>
</tr>
</tbody>
</table>