RESEARCH LETTER

Reporting of Results in ClinicalTrials.gov and High-Impact Journals

The 2007 Food and Drug Administration (FDA) Amendments Act expanded requirements for ClinicalTrials.gov, a public clinical trial registry maintained by the National Library of Medicine, mandating results reporting within 12 months of trial completion for all FDA-regulated medical products. Reporting of mandatory trial registration information on ClinicalTrials.gov is fairly complete, although there are concerns about its specificity; optional trial registration information is less complete. To our knowledge, no studies have examined reporting and accuracy of trial results information. Accordingly, we compared trial information and results reported on ClinicalTrials.gov with corresponding peer-reviewed publications.

Methods | We conducted a cross-sectional analysis of clinical trials for which the primary results were published between July 1, 2010, and June 30, 2011, in Medline-indexed, high-impact journals (impact factor ≥10; Web of Knowledge, Thomson Reuters) and that were registered on ClinicalTrials.gov and reported results. For each trial, we assessed reporting of the following results information on ClinicalTrials.gov and corresponding publications and compared reported information in both sources: cohort characteristics (enrollment and completion, age/sex demographics), trial intervention, and primary and secondary efficacy end points and results. Results information was considered concordant if the described end point, time of ascertainment, and measurement scale matched. Reported results were categorized as concordant (ie, numerically equal), discordant (ie, not numerically equal), or could not be compared (ie, reported numerically in one, graphically in the other). For discordant primary efficacy end points, we determined whether the discrepancy altered study interpretation. Descriptive analyses were performed using Excel (version 14.3.1, Microsoft).

Results | We identified 96 trials reporting results on ClinicalTrials.gov that were published in 19 high-impact journals. For 70 trials (73%), industry was the lead funder. The most common conditions studied were cardiovascular disease, diabetes, and hyperlipidemia (n = 21; 23%); cancer (n = 20; 21%); and infectious disease (n = 19; 20%). Trials were most frequently published by New England Journal of Medicine (n = 23; 24%), Lancet (n = 18; 19%), and JAMA (n = 11; 12%). Cohort, intervention, and efficacy end point information was reported for 93% to 100% of trials in both sources (Table 1). However, 93 of 96 trials had at least 1 discordance among reported trial information or reported results.

Among trials reporting each cohort characteristic and trial intervention information, discordance ranged from 2% to 22% and was highest for completion rate and trial intervention, for which different descriptions of dosages, frequencies, or duration of intervention were common.

There were 91 trials defining 156 primary efficacy end points (5 trials defined only primary safety end points), 132 (85%) of which were described in both sources, 14 (9%) only on ClinicalTrials.gov, and 10 (6%) only in publications. Among 132 end points described in both sources, results for 30 (23%) could not be compared and 21 (16%) were discordant. The majority (n = 15) of discordant results did not alter trial interpretation, although for 6, the discordance did
Table 2. Discordant Primary Efficacy End Point Results Reported on ClinicalTrials.gov and in Corresponding Publication That Altered Trial Interpretation (n = 6)

<table>
<thead>
<tr>
<th>Trial ID</th>
<th>Primary Efficacy Outcome</th>
<th>ClinicalTrials.gov-Reported Results</th>
<th>Publication-Reported Results</th>
<th>Why Trial Interpretation Was Altered</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCT00094887</td>
<td>Median time to resolution of vaso-occlusive pain crisis</td>
<td>Inhaled nitric oxide: 61.83 h (95% CI, 41.75 to 78.00); placebo: 55.16 h (95% CI, 46.00 to 72.00); no statistical analysis provided</td>
<td>Inhaled nitric oxide: 73.0 h (95% CI, 46.0 to 91.0); placebo: 65.5 h (95% CI, 48.1 to 84.0); $P = .87$</td>
<td>Time to resolution in both groups was substantially shorter on ClinicalTrials.gov than in publication</td>
</tr>
<tr>
<td>NCT00108953</td>
<td>Median time to progression</td>
<td>Sorafenib + doxorubicin: 263 d (95% CI, 146 to 384); placebo + doxorubicin: 147 d (95% CI, 66 to 244); $P = .016$</td>
<td>Sorafenib + doxorubicin: 6.4 mo (95% CI, 4.8 to 9.2); placebo + doxorubicin: 2.8 mo (95% CI, 1.6 to 5); $P = .02$</td>
<td>Median time to progression in both groups was substantially longer on ClinicalTrials.gov than in publication</td>
</tr>
<tr>
<td>NCT00177671</td>
<td>Participants with recurrence of major depression</td>
<td>Donepezil: 19/67 (95% CI, 16 to 31); placebo: 11/63 (95% CI, 6 to 18); HR, 3.97; SD, 2.09 (95% CI, 1.00 to 4.41); $P = .05$</td>
<td>Donepezil: 35% (95% CI, 24% to 46%); placebo: 19% (95% CI, 9% to 29%); HR, 2.09 (95% CI, 1.00 to 4.41), $\chi^2 = 3.97$; $P = .05$</td>
<td>Percentage of participants with major depression recurrence was lower on ClinicalTrials.gov than in publication and HR on ClinicalTrials.gov was 2-fold greater</td>
</tr>
<tr>
<td>NCT00281918</td>
<td>Median progression-free survival</td>
<td>Fluorobabine + cyclophosphamide: 981.0 d (range, 1-1343); fluorobabine + cyclophosphamide + rituximab: 1212.0 d (range: 1-1372); $P = .001$</td>
<td>Fluorobabine + cyclophosphamide: 32.8 mo (95% CI, 29.6 to 36.0); fluorobabine + cyclophosphamide + rituximab: 51.8 mo (95% CI, 46.2 to 57.6); $P &lt; .001$</td>
<td>Progression-free survival was substantially shorter in the rituximab group reported on ClinicalTrials.gov than in publication</td>
</tr>
<tr>
<td>NCT00404079</td>
<td>Roland Morris Disability Questionnaire score, 1 y</td>
<td>Glucosamine sulfate: 9, SD, 4; odds ratio, 4.5, SD, 4; $P = .05$</td>
<td>Glucosamine sulfate: 4.8 (95% CI, 3.9 to 5.6); placebo: 5.5 (95% CI, 4.7 to 6.4); $P = .50$</td>
<td>ClinicalTrials.gov score was higher for both trial groups than in publication and statistical testing results were different in 2 sources</td>
</tr>
<tr>
<td>NCT00426751</td>
<td>Participants with complete sum ST resolution 60 min after percutaneous coronary intervention (intent-to-treat population)</td>
<td>Epifibatide: 124/214; abciximab: 103/196; adjusted difference: 6.8% (95% CI, −3.0% to 16.6%); CI of adjusted difference between groups crossed 0 on ClinicalTrials.gov and did not in publication, suggesting a difference in statistical testing of results between the 2 sources</td>
<td>CI of adjusted difference between groups crossed 0 on ClinicalTrials.gov and did not in publication, suggesting a difference in statistical testing of results between the 2 sources</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: HR, hazard ratio.

(Table 2). Overall, 81 of 156 (52%) primary efficacy end points were described in both sources and reported concordant results.

There were 96 trials defining 2089 secondary efficacy end points, 619 (30%) of which were described in both sources, 421 (20%) only on ClinicalTrials.gov, and 1049 (50%) only in publications. Among 619 end points described in both sources, results for 228 (37%) could not be compared, whereas 53 (9%) were discordant. Overall, 338 of 2089 (16%) secondary efficacy end points were described in both sources and reported discordant results.

Discussion | Among clinical trials published in high-impact journals that reported results on ClinicalTrials.gov, nearly all had at least 1 discrepancy in the cohort, intervention, or results reported between the 2 sources, including many discordances in reported primary end points. For discordances observed when both the publication and ClinicalTrials.gov reported the same end point, possible explanations include reporting and typographical errors as well as changes made during the course of the peer review process. For discordances observed when one source reported a result but not the other, possible explanations include journal space limitations and intentional dissemination of more favorable end points and results in publications.3

Our study was limited to a small number of trials that were not only registered and reported results, but also published in high-impact journals. However, because articles published in high-impact journals are generally the highest-quality research studies and undergo more rigorous peer review, the trials in our sample likely represent best-case scenarios with respect to the quality of results reporting. Our findings raise questions about accuracy of both ClinicalTrials.gov and publications, as each source’s reported results at times disagreed with the other. Further efforts are needed to ensure accuracy of public clinical trial result reporting efforts.

Jessica E. Becker, AB
Harlan M. Krumholz, MD, SM
Gal Ben-Josef, BA
Joseph S. Ross, MD, MHS

Author Affiliations: Yale University School of Medicine, New Haven, Connecticut (Becker, Ben-Josef); Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut (Krumholz, Ross).

Corresponding Author: Joseph S. Ross, MD, MHS, Department of Internal Medicine, Yale University School of Medicine, PO Box 208093, New Haven, CT 06520 (joseph.ross@yale.edu).

Author Contributions: Ms Becker and Dr Ross had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Becker, Ben-Josef, Ross.

Acquisition of data: Becker, Ben-Josef, Ross.

Analysis and interpretation of data: Becker, Krumholz, Ross.

Drafting of the manuscript: Becker, Ross.
Critical revision of the manuscript for important intellectual content: Becker, Krumholz, Ben-Josef, Ross.

Statistical analysis: Becker, Ross.

Study supervision: Ross.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Drs Krumholz and Ross receive support from Medtronic to develop methods of clinical trial data sharing, from the Centers of Medicare & Medicaid Services to develop and maintain performance measures that are used for public reporting, and from the Food and Drug Administration to develop methods for postmarket surveillance of medical devices. Dr Krumholz reports that he chairs a scientific advisory board for UnitedHealthcare. Dr Ross reports that he is a member of a scientific advisory board for FAIR Health.

Funding/Support: This project was not supported by any external grants or funds. Dr Krumholz is supported by a National Heart, Lung, and Blood Institute Cardiovascular Outcomes Center Award (IU01HL105270-02). Dr Ross is supported by the National Institute on Aging (K08 AG032886) and by the American Federation for Aging Research through the Paul B. Beeson Career Development Award Program.

Previous Presentation: This study was presented at the Seventh International Congress on Peer Review and Biomedical Publication; Chicago, Illinois; September 9, 2013.


Study Design, Publication Outcome, and Funding of Research Presented at International Congresses on Peer Review and Biomedical Publication

The first International Congress on Peer Review and Biomedical Publication (PRC) was organized in 1989 to “subject the editorial review process to some of the rigorous scrutiny that editors and reviewers demand of the scientists whose work they are assessing.” Since then, peer review research was introduced as a Medical Subject Heading (MeSH), and the number of indexed articles has been continuously increasing. To evaluate the development of peer review research in biomedicine, we analyzed research presented at all PRCs since 1989.

Methods | We established a retrospective cohort of PRC poster and podium abstracts and used author names to search the PRC’s website, Web of Science (WoS), and PubMed for full articles through August 20, 2013. We collected data on authorship, time to publication, declared funding sources, article availability, and citation counts in WoS. Two authors classified study design independently, with acceptable agreement (κ = 0.78). Data were analyzed using MedCalc; 2-sided significance testing included χ² tests and Kruskal-Wallis test (type I error was set to P < .05).

Results | Of 614 presented abstracts, 75% described observational studies; 18%, intervention studies, and 7%, opinion pieces (Table 1). Over time, the number of opinion pieces decreased from 17 in 1989 to 1 in 2013 (χ² for trend = 47.3, P < .001). The number of cohort studies increased from 0 in 1989 to 8 in 2013 (χ² = 10.7, P = .002). Feasibility studies increased from 1 in 1989 to 20 in 2013 (χ² = 11.3, P < .001). The median number of abstract authors increased from 1 (95% CI, 1-1) in 1989 to 4 (95% CI, 4-5) in 2013 (P < .001). Of the 504 abstract presentations from the first 6 PRCs, 305 (61%) led to 294 published articles (Table 2). From abstract presentation to publication, there were no changes in the byline order or number of byline authors in 166 abstracts (56%), whereas 83 abstracts (28%) had changes in the number of authors listed and 45 (15%) had changes in the byline order. One hundred fourteen articles (38%) were published in JAMA, 21 (7%) in BMJ, 12 (4%) in Annals of Emergency Medicine, and 8 (3%) each in the Journal of Clinical Epidemiology and in PLoS ONE (Table 2). The median time to publication was 14 months (95% CI, 12-16), when excluding 110 articles in JAMA theme issues. One hundred articles (63%) were freely available online.

Funding was reported in 106 (36%) published articles that had been presented as abstracts at the 1989-2009 PRCs and in 45 abstracts (41%) presented at the 2013 PRC, most commonly from public or charity sources (Table 1). The absolute number and proportion of articles with declared funding increased over time, with a peak in 2005 (Table 2).

Two hundred eighty-four published articles (97%) were indexed in WoS; 265 (93%) of them received at least 1 citation, with a median of 20 (95% CI, 17-27) citations per article. Articles with the most citations were on a reporting guideline for health research1 (published in 17 journals; n = 1798 citations), synthesis of evidence4 (n = 1016), and publication bias5 (n = 547).

Discussion | Peer review research uses various study designs and is published in a broad spectrum of journals. However, experimental studies aimed at improving methods of peer review and reporting of biomedical research are still underrepresented. Although the peer review research community is aware of the consequences of nonpublication of research,6 39% of studies presented at PRCs have not been fully published. In our cohort, we were unable to determine whether the underreporting was selective (eg, favoring positive results) and were not able to determine its causes. Lack of suitable journal outlets is an unlikely explanation because there was no decrease in publication output after JAMA ceased its PRC theme issues in 2005. Because our cohort represents research presented for more than 20 years at the discipline’s major meeting, it may have limited generalizability to research presented elsewhere.

Peer review and other editorial procedures have the potential to significantly influence the knowledge base of health care. Despite their critical role in biomedical publishing, methods of peer review are still underresearched and lack dedicated funding. Systematic and competitive funding schemes...