CORRESPONDENCE COLUMNS provide an important place for comments, questions, and criticisms of work published in journals. The International Committee of Medical Journal Editors has declared that all biomedical journals should have such a section because the absence of one "denies readers the possibility of responding to articles in the same journal that published the original work." Editors of leading journals have repeatedly endorsed this view.

Yet there is a strain of skepticism about letters to the editor. As one commentator noted, "Criticism is a challenge to be overcome, usually by a mixture of semantic wriggles, ignoring the main point but expansively countering less important ones . . . and sometimes by being implicitly abusive." A contrary view is that published correspondence is part of the continual process of peer review. How important is such postpublication criticism in shaping clinical knowledge?

METHODS

My aim was to complete a preliminary study of the critical footprint made in the medical literature by 3 research studies published in The Lancet: the Hypertension Optimal Treatment (HOT) trial, the Captopril Prevention Project (CAPPP), and the Swedish Trial in Old Patients with Hypertension 2 (STOP-2). These studies were selected because they were likely to have a substantial impact on clinical practice. I also wanted to investigate the extent to which their critical footprints were preserved in the subsequent shaping of clinical knowledge about hypertension treatment.

Letters to The Lancet are collected in the 8 weeks following publication of the original article. They are read by a correspondence editor and me, and we share the decision about what to publish. Our aim is to print important criticism without duplication. About half of all letters received are published.

For each study, I prepared a taxonomy of criticism from the letters published in the journal. In each case, the authors were invited to respond to correspondents. No specific guidance was given to the original authors of each trial about how to prepare their response. I devised a list of agreed weaknesses and unanswered criticisms. Next, I searched under the following key words in the MEDLINE database for practice guidelines published from June 13, 1998, through October 25, 2000, after the initial randomized trial report: hypertension, English, human, meta-analysis, and guideline. Finally, I sought evidence for the incorporation of these criticisms into the interpretations of trials discussed in these guidelines.

RESULTS

Hypertension Optimal Treatment Trial

This randomized trial set out to assess the optimum target diastolic blood pressure in 18,790 hypertensive patients. In the published correspondence, 14 criticisms were made, 5 comments offered, and 3 questions asked (Box 1). The HOT

Context Letters to the editor are an important means for ensuring accountability of authors and editors. They form a part of the postpublication peer review process. I studied the critical footprint made in the medical literature by 3 randomized trials (Hypertension Optimal Treatment [HOT], Captopril Prevention Project [CAPPP], and Swedish Trial in Old Patients with Hypertension 2 [STOP-2]) published in The Lancet and investigated the extent to which that footprint was preserved in shaping clinical knowledge.

METHODS Qualitative appraisal of the criticism of each trial, taken from published letters. Agreed weaknesses and unanswered criticisms were identified from the authors' reply. I searched MEDLINE for practice guidelines published after the trial report and sought evidence for incorporation of criticism into these guidelines.

RESULTS From the time of publication to October 2000, HOT was cited in 9 of 36 practice guidelines; CAPPP, in 6 of 36; and STOP-2, not at all. HOT received 14 published criticisms, 5 comments, and 3 questions, of which 15 were responded to. Only 1 criticism, lack of power, was referred to in 1 guideline. CAPPP received 14 criticisms, 9 comments, and 3 questions, of which 8 were responded to. Only 1 criticism, imbalances between groups, was referred to in 1 guideline. STOP-2 received 12 criticisms, 9 comments, and 3 questions, of which only 6 were responded to.

Conclusions More than half of all criticism made in correspondence went unanswered by authors. Important weaknesses in trials were ignored in subsequently published practice guidelines. Failure to recognize the critical footprint of primary research weakens the validity of guidelines and distorts clinical knowledge.

©2002 American Medical Association. All rights reserved.
investigators responded to 8 criticisms, replied to 4 comments, and answered all 3 questions. They acknowledged 3 weaknesses in their study. First, their data were incomplete, partly because editors wished to exclude information to reduce the article’s length. Second, they were unable to separate the effects on outcomes for individual drugs. Third, “the HOT study has not definitively solved all the problems it set out to investigate.” Important potential problems were left undiscussed (eg, the accuracy of blood pressure measurement, the omission of the study’s main endpoint, and the report’s generalizability to patients most at risk).

The HOT report was cited in 9 of 36 practice guidelines published between August 1998 and October 2000. These guidelines included 3 versions (long and short) of the World Health Organization/International Society of Hypertension (WHO/ISH) guidelines on management of hypertension, 2 versions (1 complete) of guidelines from the British Hypertension Society, and 1 full version each of joint British recommendations, Canadian guidelines, Japanese guidelines for hypertension in the elderly, and a consensus document on preserving renal function in adults with hypertension (a reference list can be obtained from the author). In only 1 practice guideline did 1 criticism, lack of power, which was not cited in The Lancet correspondence, survive in discussion of the trial report.10

**Box 1. Published Criticisms, Comments, and Questions About the Hypertension Optimal Treatment (HOT) Trial**

**Criticisms**
- Precision of blood pressure measurements was not given*
- Method of blood pressure measurement produces distortion
- Aspirin treatment may introduce bias
- Intention-to-treat analysis shows no advantage to low diastolic blood pressure
- The HOT trial was stopped early, well below target number of events*
- A total of 114 events were not validated*
- Different antihypertensive agents may produce different outcomes*
- Original main endpoint was not used
- Results do not justify the conclusion about a lower target blood pressure*
- Estimate of optimum blood pressure did not take into account baseline values
- Analysis of nondiabetic population confirms the J-shaped curve*
- Marginal clinical benefits do not justify costs*
- Data are not corrected for placebo treatment*
- Patients at highest risk were excluded

**Comments**
- More prescriptions will mean higher costs*
- Extrapolation of benefits is fraught with difficulty*
- Interpretation is compatible with caution over aggressive blood pressure reduction*
- Recommendation for a particular drug regimen in hypertensive patients with diabetes
- Comparison with one other randomized trial*

**Questions**
- Does lack of data about pulse pressure affect the interpretation of the HOT trial?*
- What were the outcomes in patients with a history of ischemic heart disease?*
- What were the outcomes among patients who achieved their target blood pressure?*

*Items with an asterisk indicate that they were responded to by the authors.

**Swedish Trial in Old Patients With Hypertension 2**

This clinical trial aimed to compare the effects of conventional vs newer antihypertensive drugs on cardiovascular mortality and morbidity in more than 10,000 patients with hypertension. In the published correspondence,11 14 criticisms were made, 9 comments offered, and 3 questions asked (Box 2).

The CAPPP investigators responded to 5 criticisms, replied to 2 comments, and answered 1 question. They acknowledged 3 weaknesses in their study. First, the dosing schedule of captopril was inferior. Second, insulin sensitivity, their proposed mechanism to explain fewer cases of diabetes in patients taking captopril, was unlikely. Third, different treatments seemed to produce different patterns of clinical effects, which they were unable to separate. A large number of important issues were left undiscussed (eg, about post hoc analyses, mean doses of drugs given, cost of treatments, and the clinical implications of the study).

The CAPPP study was cited in 6 of 36 practice guidelines published between August 1998 and October 2000. These guidelines included 3 versions of the WHO/ISH guidelines on management of hypertension, 2 versions (1 long, 1 short) of guidelines from the British Hypertension Society, and 1 set of Canadian recommendations. In only 1 of these guidelines was a criticism acknowledged and discussed. The WHO/ISH guidelines refer to “imbalances in the assignment of treatment.”12
dividual drugs, and end-point reporting). No practice guidelines had incorporated the results of STOP-2, most likely because insufficient time had elapsed for the study to be included. The overall response rate in HOT, CAPPP, and STOP-2 to criticisms, comments, and questions was 40%.

**POSTPUBLICATION CRITICISM AND CLINICAL KNOWLEDGE**

**Box 2. Published Criticisms, Comments, and Questions About the Captopril Prevention Project (CAPPP) Trial**

**Criticisms**
- The dosing schedule of captopril was inadequate
- Angiotensin-converting enzyme inhibitor therapy may be harmful
- Trial design was open
- Randomization led to important imbalances between groups
- Mean daily doses of drugs are missing
- Details of combined treatments are not given
- Cost comparisons are not provided
- Inclusion criteria are violated
- Changes in blood pressure are not specified
- Reliability of screening for diabetes is questioned
- Post hoc analyses are unreliable
- Statements about clinical implications are premature
- The stroke risk increases with captopril therapy
- In considering implications, the stroke rate cannot be ignored

**Comments**
- Diuretics and β-blockers remain first-line treatment
- Angiotensin-converting enzyme inhibitors are add-on, not monotherapy
- It is hard to compare CAPPP with the United Kingdom Prospective Diabetes Study
- Diabetes incidence was low in this population
- Reduction in diabetes was unrelated to insulin sensitivity
- CAPPP is difficult to interpret
- Different outcomes would be expected from different drug classes
- Excess strokes are still present after adjusted analysis
- One could compare stroke events in databases

**Questions**
- Which patients received β-blockers as initial treatments?
- Which patients received diuretics as initial treatments?
- Was the outcome influenced by choice of drug?

*Items with an asterisk indicate that they were responded to by the authors.

©2002 American Medical Association. All rights reserved.
Questions about the basis of clinical knowledge. Following the work of Popper,23 scientific knowledge accumulates according to the resistance of hypotheses to falsification. This test, Popper argued, is a more useful way to judge the success of a scientific theory than verification. Yet it was John Stuart Mill who laid the philosophical groundwork to understand the importance of criticism in shaping knowledge. In On Liberty,24 he anticipated Popper when he wrote, “The beliefs which we have most warrant for have no safeguard to rest on but a standing invitation to the whole world to prove them unfounded.” Specifically, knowledge depends on fostering a culture of criticism: “However unwillingly a person who has a strong opinion may admit the possibility that his opinion may be false, he ought to be moved by the consideration that, however true it may be, if it is not fully, frequently, and fearlessly discussed, it will be held as a dead dogma, not a living truth.” The practical importance of this view is direct, according to Mill: “Complete liberty of contradicting and disproving our opinion is the very condition which justifies us in assuming its truth for purposes of action; and on no other terms can a being with human faculties have any rational assurance of being right.” Indeed, formal letter writing to journals is now used in some medical curricula to teach critical appraisal skills.25

In sum, criticism in letters to the editor is a neglected genre of writing. Letters enable free expression of opinion, reveal the intellectual vigor of the community concerned, and help shape knowledge. This study suggests that there is a resistance to criticism in medical research. The randomized trial is the best means medicine has to secure reliable knowledge about interventions. However, the label randomized trial seems to have blunted our desire to think critically when devising practice guidelines. Editors could do a great deal more to ensure proper accountability to criticism. In addition, editors could help further by adopting a preventive role: to encourage authors to publish fuller and longer articles, perhaps by taking advantage of the Internet, that anticipate likely criticism.

Author Contributions: Richard Horton conceived, designed, conducted, analyzed, and wrote this study.

Acknowledgment: I am grateful to Hopelyn Goodwin, who provided secretarial assistance.

REFERENCES


14. Rothwell PM, Martyn CN. Reproducibility of peer
A

LTHOUGH PUBLICATION IS A
crucial portion of the scien-
tific process, an equally im-
portant part is the subse-
quently use and citation of these
published articles by other resea-
chers and authors. We studied a cohort
of all research submitted to a scien-
tific meeting and subsequently pub-
lished to determine how these studies
were cited by other authors and deter-
mine which characteristics (including
positive results) were associated with
more frequent citation.

METHODS

We previously reported the methods of the
first phase of this study.\textsuperscript{1} To sum-
marize, all abstracts of scientific stud-
ies submitted to the Society for Aca-
demic Emergency Medicine (SAEM)
meeting in 1991 were examined. Each
submitted abstract was categorized in a
blinded fashion according to re-
search design, number of subjects, and
other characteristics (TABLE 1 and

Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized study</td>
<td>30%</td>
</tr>
<tr>
<td>Control group</td>
<td>20%</td>
</tr>
<tr>
<td>Methodology</td>
<td>10%</td>
</tr>
<tr>
<td>Design</td>
<td>5%</td>
</tr>
<tr>
<td>Sample size</td>
<td>2%</td>
</tr>
<tr>
<td>Study quality</td>
<td>1%</td>
</tr>
</tbody>
</table>

Context Citation by other authors is important in the dissemination of published science, but factors predicting it are little studied.

Methods To identify characteristics of published research predicting citation in other journals, we searched the Science Citations Index database for a standardized 3.5 years
for all citations of published articles originally submitted to a 1991 emergency medicine specialty meeting. Analysis was conducted by classification and regression trees, a non-
parametric modeling technique of regression trees, to determine the impact of pre-
viously determined characteristics of the full articles on the outcome measures. We cal-
culated the the number of times an article was cited each year and calculated the mean
impact factor (citations per manuscript per year) in other citing journals.

Results Of the 493 submitted manuscripts, 204 published articles met entry criteria.
The mean citations per year was 2.04 (95% confidence interval, 1.6-2.4; range, 0-20.9)
in 440 different journals. Nineteen articles (9.3%) were never cited. The ability to pre-
dict the citations per year was weak (pseudo $R^2=0.14$). The strongest predictor of
citations per year was the impact factor of the original publishing journal. The pres-
ence of a control group, the subjective newsworthiness score, and sample size pre-
dicted citation frequency (24.3%, 26.0%, and 26.5% as strongly, respectively). The ability to pre-
dict mean impact factor of the citing journals was even weaker (pseudo
$R^2=0.09$). The impact factor of the publishing journal was the strongest predictor, fol-
lowed by the newsworthiness score (89.9% as strongly) and a subjective quality score
(61.5%). Positive outcome bias was not evident for either outcome measure.

Conclusion In this cohort of published research, commonly used measures of study
methodology and design did not predict the frequency of citations or the importance
citing journals. Positive outcome bias was not evident. The impact factor of the original
publishing journal was more important than any other variable, suggesting that
the journal in which a study is published may be as important as traditional measures of
study quality in ensuring dissemination.

JAMA. 2002;287:2847-2850

Author Affiliations: Division of Emergency Medicine, University of California, San Francisco (Dr Callaham and Weber), Department of Emergency Medicine, Univer-
sity of Florida at Jacksonville, Jacksonville (Dr Wears).

Corresponding Author and Reprints: Michael Callaham, MD, Box 0208, University of California, San Fran-
cisco, CA 94143-0208 (e-mail: mlc@medicine.ucsf.edu).

©2002 American Medical Association. All rights reserved.