CORRESPONDENCE COLUMNS provide an important place for comments, questions, and criticisms of work published in journals. The International Committee of Medical Journal Editors1 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.2-4

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 wriggings, ignoring the main point but expansively countering less important ones . . . and sometimes by being implicitly abusive."5 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,6 the Captopril Prevention Project (CAPPP),7 and the Swedish Trial in Old Patients with Hypertension 2 (STOP-2).8 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 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 18790 hypertensive patients. In the published correspondence,9 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.

JAMA, 2002;287:2843-2847

www.jama.com
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.”44 Important potential problems were left undiscussed (eg, the accuracy of blood pressure measurement, the omission of the study’s main end point, 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, surface in discussion of the trial report.10

**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 6614 elderly patients between the ages of 70 and 84 years. 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

### 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 end point 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.
dramatic 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%.

**COMMENT**

This preliminary study shows that research reports leave deep critical footprints in the medical literature. Yet more than half of all criticisms made in letters went unanswered by authors. Important weaknesses in these trials were ignored in subsequently published practice guidelines. The failure to preserve the critical footprint left by a research article weakens the validity of guidelines and distorts clinical knowledge. This study suggests that criticism is undervalued by both the medical research and practice communities.

The limitations of this study are at least 2-fold. First, this was an exploratory investigation: only a small number of trials were chosen, and these were taken from one journal. Second, the quality of the criticisms was not independently validated. Nevertheless, these observations point to important questions not only about the peer-review process in scientific research, but also about the validity of our clinical knowledge.

The weaknesses of existing peer-review processes at medical journals have been described before. For example, Rothwell and Martyn found that reproducibility among peer reviewers of clinical neuroscience journals was poor. Journal editors have recognized these problems, and a few have tried to devise ways to enhance the reliability of their peer-review processes. Some of the most innovative work has been conducted at the Medical Journal of Australia. Editors at the Medical Journal of Australia had found that readers were “not taking full advantage of the opportunity for post-publication peer review provided by correspondence columns” in the journal. They have used the Internet to widen the peer-review process and encourage greater interaction between readers and authors.

**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.
POSTPUBLICATION CRITICISM AND CLINICAL KNOWLEDGE

Box 3. Published Criticisms, Comments, and Questions About the Swedish Trial in Old Patients With Hypertension 2 (STOP-2) Trial

Criticisms
STOP-2 was not designed to compare individual drugs
There are missing data about how many patients received β-blockers and diuretics
Pindolol is harmful
STOP-2 was an open trial
Adverse drug reaction reporting was invalid
There are biased reports of end points
A masked end-point committee does not rule out bias
True differences between drugs are obscured by combinations and crossovers
The conclusion concerning equivalence is incorrect
There is heterogeneity among treatment groups
Cox regression analysis may show an inconclusive result
Randomization has produced bias regarding diabetes rates

Comments
Physicians must wait for trials of drugs, not strategies
Future practice depends on the results of the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial
Wide pulse pressures may explain higher frequencies of end points
Conventional anti-hypertensive drugs may not be good enough
The only valid comparison is between combinations
A surprisingly large decrease in blood pressure was seen despite only dual therapy
Calcium channel blockers are equivalent to other anti-hypertensive drugs
Higher rates of myocardial infarction with calcium channel blockers are disturbing
Dose frequency of isradipine may have been wrong

Questions
Did myocardial infarctions occur more frequently in the mornings?
Were there any differences in heart rate between groups?
Could results of the ABCD trial be reproduced in STOP-2?

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

REFERENCES
14. Rothwell PM, Martyn CN. Reproducibility of peer reviewers 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.
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 previously determined characteristics of the full articles on the outcome measures. We calculated 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 predict 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 presence of a control group, the subjective newsworthiness score, and sample size predicted citation frequency (24.3%, 26.0%, and 26.5% as strongly, respectively). The ability to predict 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, followed by the newsworthiness score and sample size predicted citation frequency (24.3%, 26.0%, and 26.5% as strongly, respectively).

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 of 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.

©2002 American Medical Association. All rights reserved.