The Hidden Research Paper

Richard Horton, FRCP

What happens when scientists disagree? Most times, readers of research papers never know. However, in 1995, a dispute among the writing committee of the Italian Multicentre Acute Stroke Trial spilled out onto the pages of The Lancet. During peer review, it became clear that 2 committee members interpreted the results of the trial very differently from their colleagues. They had, for the good of the collaboration, self-censored their own views. However, this fragile truce broke down once an editor asked for signatures confirming each contributor's assent for the paper to be published. Tognoni and Roncaglioni described their disagreement as "unfortunate."

Such disagreements have come to light before. For example, divided interpretations about a polio outbreak in Israel led to separate commentaries in The Lancet. In an even more protracted dispute, a competing manuscript based on the same study was eventually published 4 years after the original article appeared. Harmony among authors cannot be relied on.

A crisis over definitions of authorship during recent years has led several medical journals to discard rigid rules for who can or who cannot be an author. Instead, the idea of the contributor has emerged. In place of the assumption that scientists named on the byline of an article are true authors, journals such as The Lancet, BMJ, and Annals of Internal Medicine now require contributors to state explicitly what part they played in the research being reported. This conceptual shift was best summarized by Rennie et al: "Contribution is the activity of science that is most relevant to publication because its disclosure can identify who is accountable for what part of the research and allows the reader to assign credit fairly."

However, important as contributorship is, this mechanism of disclosure does not take account of the ideas or interpretations contributed to the research being reported. I wanted to know whether the views expressed in a research paper are accurate representations of contributors' opinions.

### Context
To determine whether the views expressed in a research paper are accurate representations of contributors' opinions about the research being reported.

### Methods
Purposive sampling of 10 research articles published in The Lancet; qualitative analysis of answers to 6 questions about the meaning of the study put to contributors who were listed on the byline of these articles. Fifty-four contributors listed on the bylines of the 10 articles were evaluated, and answers to questions were compared between contributors within research groups and against the published research report.

### Results
A total of 36 (67%) of 54 contributors replied to this survey. Important weaknesses were often admitted on direct questioning but were not included in the published article. Contributors frequently disagreed about the importance of their findings, implications, and directions for future research. I could find no effort to study systematically past evidence relating to the investigators' own findings in either survey responses or the published article. Overall, the diversity of contributor opinion was commonly excluded from the published report. I found that discussion sections were haphazardly organized and did not deal systematically with important questions about the study.

### Conclusions
A research paper rarely represents the opinions of those scientists whose work it reports. The findings described herein reveal evidence of (self-)censored criticism, obscured meanings, confused assessment of implications, and failures to indicate directions for future research. There is now empirical support for the introduction of structured discussion sections in research papers. Editors might also explore ways to recover the plurality of contributors' opinions.

JAMA, 2002;287:2775-2778

---

©2002 American Medical Association. All rights reserved.
RESULTS

All corresponding authors gave permission for me to contact their co-contributors. However, one corresponding author, although agreeing that “the results of your project will shed light on whether the true strength of a collaborative research group is being fully achieved,” declined permission for me to contact more junior members of her research team. She wrote, “I would ask that you refrain from contacting three of the authors on our paper . . . since they are still under my supervision.” In all, 36 (67%) of 54 contributors contacted replied to the survey.

In reporting these results, I will take one study and describe the responses of the contributors in detail. The study I will focus on is a randomized trial, and I do so because it is this study design that is central to establishing evidence for or against interventions in clinical practice. Supportive or contradictory findings, together with further issues, will be explored by describing the remaining replies.

The trial concerned the efficacy of ondansetron, a peripherally active serotonin antagonist, in patients with an eating disorder. Forty-three patients were screened and 26 were randomized to receive either ondansetron or placebo. The primary outcome measure was a composite of the number of bingeing and vomiting episodes per week. For patients receiving ondansetron, at 4 weeks the mean number of episodes was 6.5 (SD, 3.9) per week. For patients receiving placebo, the mean number was 13.2 (SD, 11.6) per week.

When asked to summarize the paper, contributors seemed to reply according to their underlying interest in the research question. At the extremes, for example, one contributor took a purely clinical view: “Our study found that ondansetron significantly reduced binge eating and vomiting compared to pill placebo in women with severe bulimia nervosa.” Another took a more pathophysiologic perspective: “Blocking vagal neurotransmission primarily at the gastric level by ondansetron produces a statistically significant reduction in bulimia symptoms in a group of severely ill bulimic patients.” Different summaries suggest different interests and perhaps different motivations for doing this work. The published article, especially the discussion section, did not clearly separate these interests.

The strengths of the research were identified as follows: study design (by 6 contributors), an identified mechanism of action (4), the double-blind nature of the trial (3), daily patient contact (3), well-matched controls (2), cyclicity of symptoms taken into account (2), the large treatment response (2), and the interpretation (2). The first 3 of these strengths were clearly identified as such in the article. Similar transparency was not found for weaknesses. In the published report, highlighted weaknesses were self-reporting of symptoms and the risk of a higher motivation to succeed among study participants. However, on direct questioning, small sample size (7 contributors), short duration of study (4), no long-term follow-up (2), and poor generalizability (2) were emphasized. Concerns about the study, freely stated by the scientists undertaking this research, had not been incorporated into the article.

The views about interpretation in the context of the totality of available evidence matched those found for the summary of findings. That is, contributors ranged between strongly clinical (“ondansetron is effective in the treatment of bulimia nervosa”) and more pathophysiologic conclusions. Again, these distinctions, although clear from individual replies, were not made in the published article, where clinical and mechanistic issues were mixed together in the discussion section. The implications of the study findings were also poorly expressed in the published report of the trial, according to the responses of individual authors. The main implications concerned vagal nerve research (3 contributors), vagal influences over psychiatric symptoms (3), ondansetron as a treatment for bulimia (2), the therapeutic value of vagal blockade (2), and the need for a broader vision for research into bulimia (2). Only ondansetron as a treatment was highlighted in the final article. Indeed, according to a senior author, “the most im-
portant implication” of the trial was that the results “would help remove the negative social connotations associated with this disorder.” Nowhere was this implication mentioned in the article published in The Lancet.

Finally, in considering lines for future research, several possibilities were identified: the physiology of bulimia (4 contributors), comparisons of ondansetron with other treatments (3), the inclusion of patients with less severe conditions in subsequent trials (2), and a longer study duration (2). None of these ideas was discussed in the published article.

Many of these omissions and patterns of reporting were found in the other articles studied (data not shown). However, there were exceptions. For example, in a randomized trial of folic acid plus vitamin B₆ to lower plasma homocysteine concentrations and perhaps to ameliorate atherosclerosis, the weaknesses cited in the survey responses (small size, use of surrogate measures, and short study duration) were all discussed in the published article. The striking fact, therefore, was the inconsistency across this sample of articles. For instance, in a study of how El Niño affects diarrheal diseases in Peruvian children, several contributors pointed out that only one El Niño event had been studied. This weakness was not discussed in the article. Similarly, although wide confidence intervals were cited as a weakness in a study of cancer in individuals with Down syndrome, this weakness was not highlighted in the article published in The Lancet.

The question that yielded the most uniformly disappointing response concerned interpretation in the context of the totality of available evidence. In neither the survey responses nor the published articles were any efforts made to describe systematically evidence that related to the investigators’ own findings. Anecdotal reporting of other work was the norm in both settings. The consistent failure by scientists to provide a more rigorous overview of past evidence when considering their own findings has been pointed out before.⁹

Confusion was also common when implications of new research were being considered. For example, in the folate and vitamin B₆ randomized trial, one contributor concluded that the results “provide some justification” for treating patients at high risk of atherothrombotic disease with folic acid. Another contributor simply considered that this “first (little) piece of evidence . . . should be seen as an encouragement for other trials” only. In a systematic review of stress hyperglycemia and risk of death after myocardial infarction, one contributor drew a diagnostic conclusion: “a simple, early available, and cheap plasma glucose identifies patients at a high risk for in hospital complications and death.” Another drew a treatment lesson: “clinicians should recognise hyperglycaemia as an important prognostic marker and take an aggressive therapeutic approach for patients who have elevated blood glucose readings at the time of [myocardial infarction].” The article itself does not explore this range of opinion and takes a more conservative line, emphasizing glucose as a risk factor only.

Contributors differed in their views about future research. In response to a direct question, contributors to a paper on risk factors for suicide suggested looking at sex, family history of illness, age, socioeconomic background, psychiatric diagnosis, and life events. Yet none of these ideas was discussed in the published report. In a study of Helicobacter pylori transmission among siblings, readers were given no direction about where future research might be directed. However, in their survey responses, the contributors suggested work on the natural history of the infection, factors (especially those in the family) that influence the dynamics of infection, and a focus on early childhood years.

The Lancet introduced contributors’ descriptions of the parts each person played in the research being reported in 1997. These descriptions are written by the authors themselves. In this study, I relied on these self-reports to find links between stated contributions and survey responses. No such associations could be made, mostly because contributor statements lacked sufficient descriptive detail.

In reviewing these 10 published articles, the most frustrating aspect of comparing survey responses with published reports was the chaotic nature of discussion sections. There was no clear or consistent approach by contributors to the discussion of their results. Limitations were frequently omitted, clinical interpretations were often mixed with mechanistic reflections, and repetition of key results was common.

**COMMENT**

The results of this qualitative study show that a research paper rarely represents the full range of opinions of those scientists whose work it claims to report. I have found evidence of censored criticism; obscured views about the meaning of research findings; incomplete, confused, and sometimes biased assessment of the implications of a study; and frequent failure to indicate directions for future research. Some papers have more complete evaluations of findings than others. What was striking was the inconsistency in published evaluations, especially regarding weaknesses. The strengths of this work are its qualitative design, which produced a rich data set, and a purposeful sampling technique that confirmed the findings across a range of subject areas and study designs.

This work also had several limitations. First, these data are preliminary. The sample of articles was small and came from one journal only. The risk of bias is substantial. Second, since I surveyed contributors after publication of their studies, I could not rule out the possibility that contributors discussed their responses with one another before replying. Third, variance of opinion was determined by one person (R.H.) using a nonvalidated survey instrument. Multiple independent assessments, perhaps adopting a quantitative scale, could improve the validity of these findings.

A scientific research paper is an exercise in rhetoric; that is, the paper is designed to persuade or at least con-
Box 2. Proposal for Structured Discussion in Medical Research Papers

Summary of Key Findings
Primary outcome measure(s)
Secondary outcome measure(s)
Results as they relate to a prior hypothesis

Strengths and Limitations of the Study
Study question
Study design
Data collection
Analysis
Interpretation

Interpretation and Implications in the Context of the Totality of Evidence
Is there a systematic review to refer to?
If not, could one be reasonably done here and now?
What this study adds to the available evidence
Effects on patient care and health policy
Possible mechanisms

Controversies Raised by This Study
Future Research Directions
For this particular research collaboration
Underlying mechanisms
Clinical research

vey to the reader a particular point of view. When one probes beneath the surface of the published report, one will find a hidden research paper that reveals the true diversity of opinion among contributors about the meaning of their research findings. For both readers and editors, the views expressed in a research paper are governed by forces that are clear to nobody, perhaps not even to the contributors themselves. Who determines what is written and why? Despite the introduction of contributors’ sections to research reports, this question remains unanswered.

The gaps identified in published research reports reveal not only the range of opinions among contributors, but also the weaknesses of editorial procedures. In particular, the omission of limitations from the discussion sections must be judged a potential failure of journal peer review.

What more could editors do to recover the plurality of contributors’ opinions? The discussion section is a neglected part of the research paper. The 6 questions I asked in the survey described herein seem to set a minimum standard for addressing the central scientific issues concerning the validity and meaning of a piece of research. A first step would be to ensure that all 6 questions are answered explicitly in the discussion section, with the full range of contributors’ opinions being offered. To enable authors to answer these questions satisfactorily, longer papers may be necessary. Editors should probably relax their usual word limits for research papers or consider publishing expanded versions of the paper on the Web. Indeed, one could go further. These data provide empirical support for structured discussions. I have raised this possibility before, and other editors have substantially supported and developed this proposal.

The results reported herein indicate that more careful organization of the discussion section of a research paper might provide the framework for not only a fairer and more accurate representation of contributors’ views, but also a more complete analysis of the data being presented. The importance of omissions in research reports has been described. A proposal for the elements to be considered in a structured discussion is shown in Box 2.

Future research might aim to confirm these findings in a larger sample from multiple clinical and nonclinical journals, perhaps using a quantitative scale or a more formal method of linguistic analysis. More interestingly, an ethnographic approach would reveal how papers are actively put together. Whatever route this work takes, current definitions of “contributor” should be widened to include the appraisal and interpretation of research—what is thought as well as what is done.

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

Acknowledgment: I thank Mrs Hopelyn Goodwin, who provided secretarial assistance.

REFERENCES