

12. Hatfield C, Østbye T, Sori C. Sex of editor in medical journals. *Lancet*. 1995;345:662.
13. Garrow J, Butterfield M, Marshall J, Williamson A. Amateur editors. Paper presented at: Third International Congress on Peer Review in Biomedical Publication; September 20, 1997; Prague, Czech Republic.
14. Misakian AL, Williard R, Bero L. Gender representation in the editorial and peer review process. Paper presented at: Third International Congress on Peer Review in Biomedical Publication; September 19, 1997; Prague, Czech Republic.
15. Parlette N, Brand R, Gentry D, Gemmill M. *Longitudinal Study of Graduates, Schools of Public*

- Health 1956-1985*. Washington, DC: Association of Schools of Public Health; 1992.
16. Levy J. Women's leadership status in the American Public Health Association. *Am J Public Health*. 1987;77:1537-1538.
17. Yoels WC. The structure of scientific fields and the allocation of editorships on scientific journals: some observations on the politics of knowledge. *Sociol Q*. 1974;15:264-276.
18. Beyer J. Editorial policies and practices among leading journals in four scientific fields. *Sociol Q*. 1978;19:68-88.
19. Fox MF. Gender, environmental milieu, and productivity in science. In: Zuckerman H, Cole JR,

- Bruer JT, eds. *The Outer Circle: Women in the Scientific Community*. New York, NY: WW Norton & Co Inc; 1991.
20. Yokopenic PA, Bourque LB, Brogan D. Professional communication networks: a case study of women in the American Public Health Association. *Soc Problems*. 1975;22:493-509.
21. Tedeschi JT. Overview and introduction. In: Tedeschi JT, ed. *Perspectives on Social Power*. Chicago, Ill: Aldine Publishing; 1974.
22. Tedeschi JT, Schlenker BR, Lindskold S. The exercise of power and influence: the source of influence. In: Tedeschi JT, ed. *The Social Influence Process*. Chicago, Ill: Aldine-Atherton Inc; 1972.

# Reviewing the Reviews

## The Example of Chronic Fatigue Syndrome

John Joyce, MB, BCh, MRCPsych; Sophia Rabe-Hesketh, PhD;  
Simon Wessely, MA, BM, BCh, MSc, MD, FRCP, MRCPsych

**Objective.**—To test the hypothesis that the selection of literature in review articles is unsystematic and is influenced by the authors' discipline and country of residence.

**Data Sources.**—Reviews in English published between 1980 and March 1996 in MEDLINE, EMBASE (BIDS), PSYCHLIT, and *Current Contents* were searched.

**Study Selection.**—Reviews of chronic fatigue syndrome (CFS) were selected. Articles explicitly concerned with a specialty aspect of CFS and unattributed, unreferenced, or insufficiently referenced articles were discarded.

**Data Extraction.**—Record of data sources in each review was noted as was the departmental specialty of the first author and his or her country of residence. The references cited in each index paper were tabulated by assigning them to 6 specialty categories, by article title, and by assigning them to 8 categories, by country of journal publication.

**Data Synthesis.**—Of 89 reviews, 3 (3.4%) reported on literature search and described search method. Authors from laboratory-based disciplines preferentially cited laboratory references, while psychiatry-based disciplines preferentially cited psychiatric literature ( $P = .01$ ). A total of 71.6% of references cited by US authors were from US journals, while 54.9% of references cited by United Kingdom authors were published in United Kingdom journals ( $P = .001$ ).

**Conclusion.**—Citation of the literature is influenced by review authors' discipline and nationality.

*JAMA*. 1998;280:264-266

MANY narrative reviews and reviews that describe themselves as systematic have been shown to be nonreproducible and to be of low mean scientific quality.<sup>1-3</sup> A lack of clearly specified methods of identifying, selecting, and validating included information has been among the problems noted.<sup>4</sup> Experts could not

agree, even among themselves, about whether other experts who wrote review articles had conducted a competent search or generated a bias-free list of citations.<sup>5,6</sup>

Few things are certain about chronic fatigue syndrome (CFS) other than that it is controversial. Both public and professional opinions are often debated passionately. In such circumstances both physicians and interested members of the public may turn for guidance and information to review articles. Such articles fulfill an important function for professionals, journalists, and patients unable to find, locate, or evaluate primary sources of information. This is particularly important in CFS, since poten-

tially relevant research spans many disciplines, with important contributions coming from specialties as diverse as immunology, virology, internal medicine, psychiatry, psychology, and neurology. Thus, our aim was to examine the quality of current reviews of CFS. Our hypothesis was that use of the literature would show the following biases: the identification and selection of literature for review is unsystematic, it fails to reflect the broad range of literature, and it is influenced by the author's discipline and country of residence.

### METHODS

#### Data Sources

All reviews of CFS between 1980 and 1996 from English-language journals were eligible. We defined a review as an article that made a claim, either implicit or explicit, referring to the range of knowledge known at the time of publication and that represented itself as being able to reach general conclusions about CFS. Articles explicitly labeled as dealing with a specialty aspect of CFS, such as "psychiatric aspects of CFS" or "immunological findings in CFS," were excluded. Reviews with fewer than 15 references (we considered a citation list of at least 15 as necessary evidence of a serious attempt to review the subject) or without any details about authors were excluded. Seventy-three foreign language reviews in 14 different languages found in the same search were excluded due to lack of linguistic expertise and small numbers per country of journal publication.

We searched MEDLINE, EMBASE (BIDS), PSYCHLIT and *Current Contents*. We conducted a free text search using the terms *chronic fatigue syn-*

From the Institute of Psychiatry (Drs Joyce, Wessely, and Rabe-Hesketh), King's College School of Medicine, King's College, and Maudsley Hospitals (Dr Wessely), London, England.

Presented at the Third International Congress on Peer Review in Biomedical Publication, Prague, Czech Republic, September 19, 1997.

Corresponding author: Simon Wessely, MA, BM, BCh, MSc, MD, FRCP, MRCPsych, Department of Psychological Medicine, Denmark Hill, London SE5 9RS, England (e-mail: sphascw@iop.bpmf.ac.uk).

*drome, neurasthenia, myalgic encephalomyelitis, and tiredness* and the truncated terms *chronic fatigue* and *postviral*. More than 4000 references were checked in title and abstract by one of the authors (J.J.). All possible reviews were then confirmed by one of the authors (S.W.).

### Data Extraction

We used 4 phases in extracting and categorizing data. First, using the first 3 of 10 criteria for the assessment of scientific quality of research overviews and how they relate to selecting literature to reviews as recommended by Oxman and Guyatt,<sup>7</sup> we noted comments the authors made on their search methods, each article's comprehensiveness, and the review article's inclusion criteria. Second, we divided the tasks of tabulating (J.J.) and checking (S.W.) between the departmental specialty of the first author and his or her country of residence and found no discrepancies. Third, the references cited in each index paper were tabulated and assigned to 6 specialty or subject categories by article title alone as shown in Table 1. Fourth, the references cited were also tabulated and assigned to 8 categories representing the countries in which they were published, including the United States, United Kingdom, Europe (excluding the United Kingdom), Australia, Canada, New Zealand, other (which included South Africa, Israel, India, Japan, and China), and not listed. The place of publication of each journal title was ascertained by consulting Libertas' list of serials.

### Statistical Methods

The data consist of a set of percentages for each review, representing the number of references that fall into each subject or country of publication. We wished to test how the author's discipline, country of residence, and the country in which the article was published had affected the use of references. Since the subject or country of publication categories may be regarded as repeated measurements within each review and the data are approximately normally distributed, this could be done by repeated-measures multivariate analysis of variance (MANOVA) using the matrix of percentages. In order to display the multivariate data in 2 dimensions, we made a series of biplots,<sup>8</sup> ie, plotting the first 2 principal components of the matrix of percentages together with a set of axes for the reference subject or country of publication category. The repeated MANOVA was carried out using SPSS statistical software (SPSS Inc, Chicago, Ill) and the biplots were made using S-Plus (MathSoft, Seattle, Wash).

Table 1.—References Tabulated According to Specialty by Article Title

Reference Discipline	Content
General medicine	Internal medicine, rheumatology
General psychiatry	Psychiatry, psychology
Laboratory	Microbiology, virology, immunology
Neuromuscular	Neurology, neuromuscular
Neuropsychiatry	Neuroendocrinology, neuropsychology, neuroimaging
Unclassifiable	

### RESULTS

One hundred eighty-six reviews were found on our preliminary search. Eighty-nine of these were counted as eligible for analysis. All were checked by one of the authors (S.W.) and there were no disagreements between the 2 reviewers. Fifty-five of the articles concerned a specialty aspect of CFS, despite a general title, or were primary research, an audit, a case report, or a first-person account (26 of these were reviews specifically dealing with treatment only). Thirteen were not attributed to an author, 7 were without any references, and 18 had fewer than 15 references. Four were unobtainable despite a thorough interlibrary search.

#### Details of Literature Search

Only 3 (3.4%) of the 89 reviews reported on the database source or sources used to conduct its literature search, none of which were written by any of the current authors. In 2 articles, the authors merely described the search, without further elaboration, as "relevant published research literature," and in 1 article, the author reported the databases used for the search. One of the 3 specifically reported its inclusion criteria.

#### Reference Disciplines by Specialty of Author

After data inspection, 1 review (by a neurologist) was excluded as an outlier, as its large distance from all other data would exert a disproportionate influence.

Because of the insufficient numbers of nurse-therapists, pharmacologists, and those in the other group for testing, we avoided the problems of inference, which would be caused by having too many small groups, by including for analysis only those groups with a number of authors who were mostly physicians, either general practitioners or specialized in infectious diseases or psychiatry. There was a highly significant interaction between reference disciplines and specialty of author ( $P = .01$ ,  $F_{12,185} = 2.84$ ). Figure 1 shows the pattern of reference use by infectious diseases specialists and psychia-

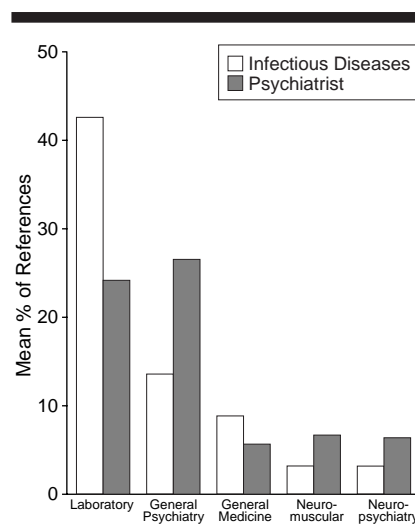


Figure 1.—Use of reference disciplines by infectious disease specialists and psychiatrists.

trists. Those working in infectious diseases quote the laboratory category most often followed by physicians, nurses, and general practitioners. Infectious diseases specialists and general practitioners quote psychiatric articles least. Psychiatrists and pharmacologists quote the laboratory category and psychiatric categories about equally.

The biplot in the left panel of Figure 2, shows the spread of reviews by specialty of author, in this case psychiatrist or infectious diseases specialist according to reference disciplines. Those reviews authored by infectious diseases experts cluster on the left around the laboratory category, and those authored by psychiatrists cluster on the right around general psychiatry. Principal components accounted for 73.5% and 14.4% of the variance.

#### Country of Residence and Reference Selection

The distribution of countries of reference publication was compared by countries where the review author resided using repeated measures-MANOVA. Only references from the United States and United Kingdom were included due to small numbers of articles published in other countries. The interaction between where the author lived and where the reference was published was significant ( $P < .001$ ,  $F_{1,72} = 145.3$ ).

The 2 principal components accounted for 92.1% and 5% of the variance, respectively (97.1% in total). The right panel of Figure 2 shows the spread of reviews according to where the author resides and in what country the references were published. Those reviews with US authors cluster to the right and those with United Kingdom authors cluster to the

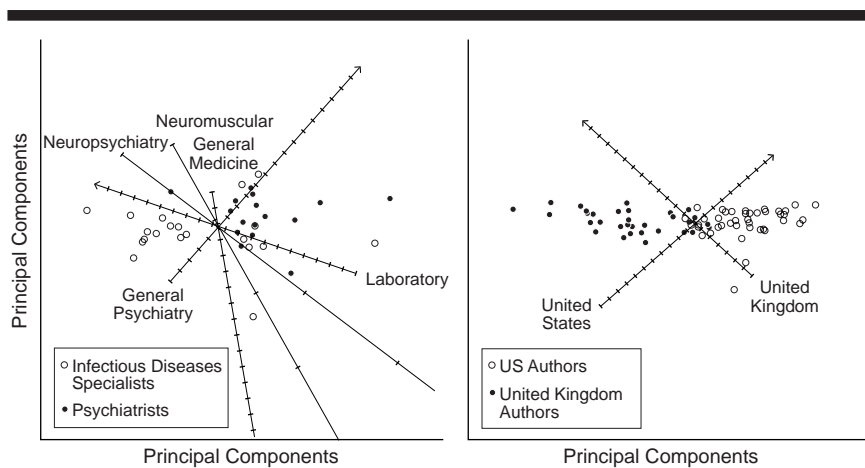


Figure 2.—Left, Biplot of reference disciplines by specialty of author. Right, Biplot of country in which reference was published and in what country the authors reside. The points represent reviews, and distances between points reflect differences between reviews in their citation patterns, according to discipline (left) and country (right). The axes show the direction in which the percentage of references in a reference category increase, and tick marks represent 5% increments. For example, the left side of the figure shows that reviews vary from approximately 0% references in the laboratory category to 100% from right to left.

Table 2.—Mean Percentage of Reference Country of Publication Across All Reviews and by US, United Kingdom, and Australian Authors

Reference Nationalities	All Reviews (n = 89)	Author Country of Residence		
		United States (n = 42)	United Kingdom (n = 32)	Australian (n = 6)
United States	54.7	71.6	34.5	43.7
United Kingdom	33.7	17.4	54.9	34.7
Australia	2.6	1.7	2.1	12.7
Europe	2.5	2.2	3.4	0.6
Canada	1.7	1.7	1.6	1.3
New Zealand	0.5	0.3	0	0.8
Other	0.2	0.2	0.2	0.1
Not listed	4.1	4.9	3.3	6.1

left. We have not included the 6 reviews by Australian authors in this biplot in order to preserve the clarity of presentations. However, Table 2 shows that the Australian reviews were situated midway between the United States and the United Kingdom in terms of citation distribution by country.

### COMMENT

Despite the recent emphasis on the necessity for quality in medical reviews, our results show that in the area of CFS the vast majority of reviews are not based on systematic literature searches and do not use objective criteria for inclusion or exclusion.

One might reasonably expect that reviews of a multidisciplinary subject, such

as CFS, are able to integrate findings from many sources—instead it is possible that they perpetuate preexisting disciplinary biases. We have shown that the choice of articles to cite is influenced by the author's discipline and the country in which he or she resides. We emphasize that we have only looked at review articles that claimed to be comprehensive and have not included any review articles that made such a bias explicit. A reader consulting any of the review articles we have studied expects them to be an objective synthesis of a complex subject. Instead, most display biases toward particular disciplines, usually the one in which the author practices. Reference bias has been previously reported in drug trials<sup>9</sup> but not, to our

knowledge, in the area of reviews. Such biases are not unexpected but are important. Similarly, it is a staple of academic gossip that Americans only cite US literature, Europeans European literature, and so on. We have confirmed that both US and United Kingdom authors are more likely to cite literature published in their own countries. Also of note is the underuse of Continental European literature by US or United Kingdom authors. We are aware of only 1 previous confirmation of this intuition.<sup>10</sup>

Exclusion of references because of language has been shown to introduce bias in randomized controlled trials.<sup>11-13</sup> We acknowledge our findings can only be generalized to the English-language literature. For fairly obvious reasons, language bias alone cannot explain our findings. American English and United Kingdom English may not sound the same, but they do read the same.

Dr Joyce is supported by an educational grant from Pfizer, UK: Sandwich, Kent, England.

We would like to thank Iain Chalmers, MD, for his comments on the manuscript.

### References

1. Dawson NV, Arkes HR. Systematic errors in medical decision making: judgment limitations. *J Gen Intern Med.* 1987;2:183-187.
2. Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol.* 1996;2:235-243.
3. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA.* 1990;263:1392-1395.
4. Mulrow C. The medical review article: state of the science. *Ann Intern Med.* 1987;106:485-488.
5. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol.* 1991;44:92-98.
6. Ingelfinger FJ. Peer review in biomedical publication. *Am J Med.* 1974;56:686-692.
7. Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci.* 1992;703:125-133.
8. Gower JC, Hand DJ. *Biplots.* London, England: Chapman & Hall; 1996.
9. Gotzsche PC. Reference bias in reports of drug trials. *BMJ.* 1987;295:654-656.
10. Campbell F. National bias: a comparison of citation practices by health professionals. *Bull Med Libr Assoc.* 1996;78:376-382.
11. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol.* 1996;48:159-163.
12. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomized controlled trials published in English and German. *Lancet.* 1997;350:326-329.
13. Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet.* 1996;347:363-366.