Mulrow C (1987). The medical review article: state of the science. Ann Intern Med 106:485-488.

Whole Article

The Medical Review Article: State of the Science

CYNTHIA D. MULROW, M.D., M.Sc.; San Antonio, Texas

Fifty reviews published during June 1985 to June 1986 in four major medical journals were assessed in a study of the methods of current review articles. Assessments were based on eight explicit criteria adapted from published guidelines for information syntheses. Of the 50 articles, 17 satisfied three of the eight criteria; 32 satisfied four or five criteria; and 1 satisfied six criteria. Most reviews had clearly specified purposes (n = 40) and conclusions (n = 37). Only one had clearly specified methods of identifying, selecting, and validating included information. Qualitative synthesis was often used to integrate information included in the review (n = 43); quantitative synthesis was rarely used (n = 3). Future research directives were mentioned in 21. These results indicate that current medical reviews do not routinely use scientific methods to identify, assess, and synthesize information. The methods used in this systematic assessment of reviews are proposed to improve the quality of future review articles.

GOOD REVIEW articles are precious commodities. In this era of a plentiful and burgeoning medical literature, the individual's capacity to read and absorb information has not changed. Reduction of large quantities of information into palatable pieces is essential for digestion.

Adequate refinement of information calls for critical exploration, evaluation, and synthesis. The accomplished reviewer must tediously sift and sort data sources, systematically appraise data quality, and thoughtfully integrate essential data into a unified whole. Such reviewers must synthesize data in a cogent and illuminating way, not just shuffle documents (1).

Unfortunately medical reviews are often subjective, scientifically unsound, and inefficient (2). Strategies for identifying and selecting information are rarely defined. Collected information is reviewed haphazardly with little attention to systematic assessment of quality. Under such circumstances, cogent summarization is an arduous, if not insurmountable, task. The purpose of this article is twofold: to describe the methods of recently published review articles, and to demonstrate a systematic method for preparing review articles.

Methods

Of the 117 journals indexed in the *Abridged Index Medicus*, the four peer-reviewed American medical journals with circulations of greater than 50 000 were identified (3). Articles classified as review or progress articles published in these journals during June 1985 to June 1986 were selected. For *Annals of Internal Medicine* and *Archives of Internal Medicine*, 10 (4-13) and 17 (14-30) review articles, respectively, were identified; for *The Journal of the American Medical Association*, 4 (31-34) state of the art/review articles were identified; and for *The New England Journal of Medicine*, 19 (35-53) medical progress articles were identified. A single assessment was done on reviews published in two parts (54-57).

All reviews were assessed with eight explicit criteria adapted from published guidelines for information syntheses (2, 58, 59): Was the specific purpose of the review stated? Were sources and methods of the citation search identified? Were explicit guidelines provided that determined the material included in and excluded from the review? Was a methodologic validity assessment of material in the review performed? Was the information systematically integrated with explication of data limitations and inconsistencies? Was the information integrated and weighted or pooled metrically? Was a summary of pertinent findings provided? Were specific directives for new research initiatives proposed? Each criterion was categorized as specified, unclear, or not specified.

Results

Assessments of the methods of the 50 review articles are presented in Table 1. No single review clearly specified all eight criteria. One met six criteria; 32 met four or five criteria; and the remaining 17 met three criteria. Most reviews were written by more than 1 author (range, 1 to 12), and the average number of references cited was 100 (range, 18 to 381).

SPECIFIED PURPOSE

The purpose of the review was stated in 40 articles. These purposes were often broad and exhaustive. For example, some reviews addressed multiple aspects of different diseases, such as epidemiology, natural history, physical manifestations, pathogenesis, diagnosis, therapy, and prevention (38, 42). Other reviews had narrower purposes and addressed only treatment (15, 21, 22, 24), pathogenesis (35, 49, 52), clinical presentation (36), or prevention (30). In 10 reviews, no clearly defined purpose was found. Although 1 of these failed to state any purpose whatsoever (44), most were nebulously defined reviews of "current concepts," "aspects of disease," or "lessons derived from 15 years of studies" (23, 45, 51).

Clearly stated purposes are important for several reasons. They give the reader a frame of reference for deciding whether to read further. Specific purposes help determine strategies to select information. For example, reviews concerning therapeutic efficacy might be limited to data from controlled clinical trials; reviews specifically

From the Division of General Internal Medicine, Department of Medicine, University of Texas Health Science Center at San Antonio; San Antonio, Texas.

Table 1. Assessments of Methods Used in Fifty Current Medical Reviews, June 1985 to June 1986

	Specified	Unclear	Not Specified
	←	*n*	→
Purpose	40	1	9
Data identification	1	2	47
Data selection	1	0	49
Validity assessment	1	1	48
Qualitative synthesis	43	0	7
Quantitative synthesis	3	1	46
Summary	37	12	1
Future directives	21	4	25

addressing natural history might emphasize data from inception cohort studies; and reviews addressing etiology might include data from case-control studies. Finally, the specific purpose of the review can determine appropriate methods of information assessment. Questions concerning where and in whom a particular diagnostic test, intervention, or prevention strategy works might require careful assessment of population and setting. Questions concerning optimal timing of a diagnostic test, intervention, or prevention strategy might concentrate on the actual methods of the test or intervention.

DATA IDENTIFICATION AND SELECTION

The source of the information reviewed was clearly identified in only one review (22), which described a computer-assisted search of English-language articles on MEDLINE. Two reviews (8, 33) stated that published reports had been reviewed but did not specify search processes or time periods. Selection procedures that determined which specific studies to include in the review were provided only by Zarrabi and Rosner (30).

Most reviewers left to conjecture whether included data were identified from automated databases, expert consensus, textbooks, present contents of reference files, or personal-favorite selections. Most also failed to report whether included data were selected on the basis of predetermined criteria, such as particular study designs or population characteristics. As a result, it was impossible to determine whether reviewed material represented information available on a given subject. Selection biases, where authors preferentially cite data that support their opinions, could not be evaluated. Moreover, it was difficult to evaluate whether relevant material had been excluded. This judgment required either considerable personal knowledge of the reviewed topic or considerable faith in the objectivity of the reviewers.

VALIDITY ASSESSMENT

A standardized methodologic assessment of data was used in only one review (22). This article had three authors; whether the assessment ratings represented a consensus or one author's view was not reported. No other review appraised validity in a systematic manner.

Systematic assessment of quality or validity determines, on the basis of a critical examination of the methods used to produce the findings, what conclusions are justifiable. Appraisal of research designs, implementations, and analyses is required. Because authors of most reviews failed to appraise systematically the information regarding population, program, or setting characteristics, judgments on generalizability of the information were limited. Failure to examine details of study design (such as diagnostic and measurement techniques; disease, exposure, and outcome definitions; and intervention and analytic approaches) left the quality of data included open to question.

DATA SYNTHESIS

Although some reviewers (4, 10, 21, 23, 25, 34, 42) merely listed findings, most ($n = 43$) provided some degree of qualitative integration by mentioning limitations and inconsistencies in existing data. Although this qualitative synthesis was often scant and haphazard, three reviews (15, 33, 52) exemplified the value of critically integrating prima-facie conflicting evidence. Hussey (15) elucidated differences in published reports concerning a therapeutic intervention by exploring the differences in study populations, intervention approaches, and outcome measures. Valuable insights about the generalizability of findings were gained: benefits of the therapy were limited to particular populations and settings. Brewster (33) showed that systematic errors in research implementation, such as selection and reporting biases, explained much of the variance in reports on the prevalence of a particular health problem. Clouse and Comp (52) carefully related current knowledge of a pathophysiologic mechanism to differences in laboratory techniques and limitations of measurement methods.

In contrast, synthesis involving quantitative methods was provided in only three reviews. Brewster (33) compared prevalence information among eight studies by converting data to a common numerator and denominator. Houston (24) combined data from four studies involving a drug therapy to determine average response rates, average magnitudes of response, and average times to response. Sakata and colleagues (8) pooled results of several studies to characterize the clinical features of a particular patient group. In these reviews, quantitative methods were used to provide a common unit of comparison and to identify average effects or average characteristics. Other advantages of quantitative techniques, such as identification of interactions and identification of small effects not readily detectable by individual small studies, were not used.

SUMMARIES AND FUTURE DIRECTIVES

Approximately 75% of reviews summarized pertinent findings in remarks in either the initial abstract or the final paragraphs. Although these summaries were helpful in compressing review results into an easily manageable form, they required cautious interpretations. Some resembled "bottom-line" remarks based on "armchair" reviewing techniques; conclusions were not always supported by valid review processes.

Directives for future studies were proposed in only 21 reviews. Specific research recommendations promoted identification of the most promising areas for future re-

search and discouraged duplicative and wasteful efforts. These reviews efficiently enhanced attainment of knowledge by suggesting to the designer of the 100th study what had been learned from the first 99.

Conclusions

This article presents a systematic review method and has shown that current medical reviews do not routinely use such systematic methods to identify, assess, and synthesize information. Although purposes (often broad) were stated in most reviews, the sources and selection methods of reviewed data were rarely defined. Standardized methodologic criteria for assessing the validity of the data were not used. Synthesis of data was weighted toward occasional and informal qualitative critiques with little use of quantitative methods. Summaries were made without showing careful reviewing techniques, and future research directives were often neglected. Although some reviews provided useful overviews of topics, their methods were not replicable and their conclusions might not be valid.

To improve the scope, impact, and quality of reviews, several steps are needed. First, a well-conceived review always answers a question. This question should be made clear at the beginning of the review (60). It should be precisely formulated, rather than broad or ill defined.

Second, efficient strategies for identifying relevant material of substantive quality and excluding irrelevant or poor-quality material are needed. Generally, computer searches of the literature cross-checked with references from other review articles can be used to identify pertinent literature. Experts in the area to be reviewed can be used to verify that all pertinent citations have been identified, but these persons should not serve as the sole source of information. Explicit guidelines for determining what data to include in the review should be stated. These guidelines should coincide with the purpose of the review. In fact, the precise definition of the review's purpose may determine whether characteristics such as study designs, study populations, disease definitions, or time frames should be used as criteria for selecting information for the review.

Third, to manage large quantities of data objectively and effectively, standardized methods of appraising information should be included in more review processes. This involves appraisal of research designs, implementations, and analyses. Standardized appraisal forms addressing these issues can be used by reviewers to optimize uniform assessment of data. To avoid single-reviewer biases, data assessments may be consensus ranked by more than one reviewer. Experts from different areas, such as appropriate specialists, statisticians, and research methodologists, can be used both to help develop the standardized appraisal forms and to rank data.

Fourth, final synthesis of information should involve systematic rather than selective integration. Data regarded as scientifically unsound on the basis of the standardized appraisal should be discarded. Other data can sometimes be assigned a weight or relative value based on its quality as determined by the standardized appraisal (61). Insights gained from careful explorations of divergent findings in scientifically valid data sets should be sought, and limitations of data sets identified. More reviewers should consider quantitative synthesis techniques to complement and supplement their qualitative techniques. Quantitative methods can be used to provide a common unit of comparison between studies and to combine data from several studies. These methods can be used to evaluate generalizability, consistency, interactions, and small effects that are not readily recognizable from individual studies.

Fifth, conclusions are justified only when the aforementioned processes of collecting, analyzing, and integrating information are systematically and thoughtfully applied. The conclusions should be succinct and logically ordered summarizations of data. If the appraisal and synthesis of data involves weighting of evidence according to some type of quality assessment, the conclusions too should reflect the relative weighting.

Finally, reviewers should capitalize on their intensive efforts by clearly identifying gaps in present knowledge and suggesting future initiatives. Unsolved issues and problems can be delineated, and appropriate methods for addressing these issues can be suggested. In this way, the reader finishes the review with a view of what is not known about the subject as well as what is known (60).

By using these systematic methods of exploration, evaluation, and synthesis, the good reviewer can accomplish the task of advancing scientific knowledge.

ACKNOWLEDGMENTS: The author thanks Drs. A. Diehl, R. Bauer, J. Pugh, R. Velez, and P. Mullen for their assistance with this manuscript.

Grant support: by a Teaching and Research Scholar Award from the American College of Physicians.

▶ Requests for reprints should be addressed to Cynthia D. Mulrow, M.D., M.Sc.; Ambulatory Care (11C), Audie Murphy Veterans Administration Hospital, 7400 Merton Minter Drive; San Antonio, TX 78284.

References

1. President's Science Advisory Committee. *Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information.* Washington, D.C.: Government Printing Office; 1963.
2. Light RJ, Pillemar DB. *Summing Up: the Science of Reviewing Research.* Cambridge, Massachusetts: Harvard University Press; 1984.
3. *Ulrich's International Periodicals Directory.* 24th ed. New York: R.R. Bowker Co.; 1985.
4. Schafer AI. The hypercoagulable states. *Ann Intern Med.* 1985;**102**:814-28.
5. Gale RP, Foon KA. Chronic lymphocytic leukemia: recent advances in biology and treatment. *Ann Intern Med.* 1985;**103**:101-20.
6. Packer M. Vasodilator therapy for primary pulmonary hypertension: limitations and hazards. *Ann Intern Med.* 1985;**103**:258-70.
7. Ausubel K, Furman S. The pacemaker syndrome. *Ann Intern Med.* 1985;**103**:420-9.
8. Sakata S, Nakamura S, Miura K. Autoantibodies against thyroid hormones or iodothyronine: implications in diagnosis, thyroid function, treatment, and pathogenesis. *Ann Intern Med.* 1985;**103**:579-89.
9. Bass E. Cardiopulmonary arrest: pathophysiology and neurologic complications. *Ann Intern Med.* 1985;**103**(6 pt 1):920-7.
10. Graham DY, Smith JL. Aspirin and the stomach. *Ann Intern Med.* 1986;**104**:390-8.
11. Pasulka PS, Bistrian BR, Benotti DN, Blackburn GL. The risks of surgery in obese patients. *Ann Intern Med.* 1986;**104**:540-6.
12. Meltzer RS, Visser CA, Fuster V. Intracardiac thrombi and systemic embolization. *Ann Intern Med.* 1986;**104**:689-98.
13. Marton KI, Gean AD. The spinal tap: a new look at an old test. *Ann Intern Med.* 1986;**104**:840-8.
14. Burch WM. Cushing's disease: a review. *Ann Intern Med.* 1985;**145**:1106-11.

15. Hussey KP. Vasopressin therapy for upper gastrointestinal tract hemorrhage: has its efficacy been proven? *Arch Intern Med.* 1985;**145:**1263-7.
16. Horowitz M, Collins PJ, Shearman DJC. Disorders of gastric emptying in humans and the use of radionuclide techniques. *Arch Intern Med.* 1985;**145:**1467-72.
17. Tobin MJ. Use of bronchodilator aerosols. *Arch Intern Med.* 1985;**145:**1659-63.
18. Barish CF, Wu WC, Castell DO. Respiratory complications of gastroesophageal reflux. *Arch Intern Med.* 1985;**145:**1882-8.
19. Shively B, Goldschlager N. Progress in cardiac pacing: part I. *Arch Intern Med.* 1985;**145:**2103-6.
20. Houston MC. Sodium and hypertension: a review. *Arch Intern Med.* 1986;**145:**179-85.
21. Hoshino PK, Gaasch WH. When to intervene in chronic aortic regurgitation. *Arch Intern Med.* 1986;**146:**349-52.
22. Levine MN, Sackett DL, Bush H. Heroin vs morphine for cancer pain? *Arch Intern Med.* 1986;**146:**353-6.
23. Peterson CE, Kwaan HC. Current concepts of warfarin therapy. *Arch Intern Med.* 1986;**146:**581-4.
24. Houston MC. Treatment of hypertensive emergencies and urgencies with oral clonidine loading and titration: a review. *Arch Intern Med.* 1986;**146:**586-9.
25. Levin ME, Rigg LA, Marshall RE. Pregnancy and diabetes: team approach. *Arch Intern Med.* 1986;**146:**758-67.
26. Mallette LE, Eichhorn E. Effects of lithium carbonate on human calcium metabolism. *Arch Intern Med.* 1986;**146:**770-6.
27. Fulkerson WJ, Coleman RE, Ravin CE, Saltzman HA. Diagnosis of pulmonary embolism. *Arch Intern Med.* 1986;**146:**961-7.
28. Park GD, Spector R, Goldberg MJ, Johnson GF. Expanded role of charcoal therapy in the poisoned and overdosed patient. *Arch Intern Med.* 1986;**146:**969-73.
29. Chan WC, Winton EF, Waldmann TA. Lymphocytosis of large granular lymphocytes. *Arch Intern Med.* 1986;**146:**1201-3.
30. Zarrabi MH, Rosner F. Rarity of failure of penicillin prophylaxis to prevent postsplenectomy sepsis. *Arch Intern Med.* 1986;**146:**1207-8.
31. Schuckit MA. Genetics and the risk for alcoholism. *JAMA.* 1985;**254:**2614-7.
32. Stone KM, Grimes DA, Magder LS. Primary prevention of sexually transmitted diseases: a primer for clinicians. *JAMA.* 1986;**255:**1763-6.
33. Brewster JM. Prevalence of alcohol and other drug problems among physicians. *JAMA.* 1986;**255:**1913-20.
34. Morrison RG. Medical and public health aspects of boxing. *JAMA.* 1986;**255:**2475-80.
35. Brittigan BE, Cohen MS, Sparling PF. Gonococcal infection: a model of molecular pathogenesis. *N Engl J Med.* 1985;**312:**1683-94.
36. Huston TP, Puffer JC, Rodney WM. The athletic heart syndrome. *N Engl J Med.* 1985;**313:**24-32.
37. Foley KM. The treatment of cancer pain. *N Engl J Med.* 1985;**313:**84-95.
38. Herzog DB, Copeland PM. Eating disorders. *N Engl J Med.* 1985;**313:**295-303.
39. Rojeski MT, Gharib H. Nodular thyroid disease: evaluation and management. *N Engl J Med.* 1985;**313:**428-36.
40. Fielding JE. Smoking: health effects and control (part 1). *N Engl J Med.* 1985;**313:**491-8.
41. Fraser CL, Arieff AI. Hepatic encephalopathy. *N Engl J Med.* 1985;**313:**865-73.
42. Elliot DL, Tolle SW, Goldberg L, Miller JB. Pet-associated illness. *N Engl J Med.* 1985;**313:**985-95.
43. Lemon SM. Type A viral hepatitis. New developments in an old disease. *N Engl J Med.* 1985;**313:**1059-67.
44. Demling RH. Burns. *N Engl J Med.* 1985;**313:**1389-98.
45. Goorin AM, Abelson HT, Frei E III. Osteosarcoma: fifteen years later. *N Engl J Med.* 1985;**313:**1637-43.
46. DeBusk RF, Blomqvist CG, Kouchoukos NT, et al. Identification and treatment of low-risk patients after acute myocardial infarction and corconary-artery bypass graft surgery. *N Engl J Med.* 1986;**314:**161-6.
47. Benatar SR. Fatal asthma. *N Engl J Med.* 1986;**314:**423-9.
48. Colucci WS, Wright RF, Braunwald E. New positive inotropic agents in the treatment of congestive heart failure: mechanisms of action and recent clinical developments: 1. *N Engl J Med.* 1986;**314:**290-9.
49. Ross R. The pathogenesis of atherosclerosis—an update. *N Engl J Med.* 1986;**314:**488-500.
50. Corey L, Spear PG. Infections with herpes simplex viruses. (part 1). *N Engl J Med.* 1986;**314:**686-91.
51. Katzman R. Alzheimer's disease. *N Engl J Med.* 1986;**314:**964-73.
52. Clouse LH, Comp PC. The regulation of hemostasis: the protein C system. *N Engl J Med.* 1986;**314:**1298-304.
53. Riggs BL, Melton LJ III. Involutional osteoporosis. *N Engl J Med.* 1986;**314:**1676-86.
54. Shively B, Goldschlager N. Progress in cardiac pacing: part II. *Arch Intern Med.* 1985;**145:**2238-44.
55. Fielding JE. Smoking: health effects and control (part 2). *N Engl J Med.* 1985;**313:**555-61.
56. Colucci WS, Wright RF, Braunwald E. New positive inotropic agents in the treatment of congestive heart failure: mechanisms of action and recent clinical developments: 2. *N Engl J Med.* 1986;**314:**349-58.
57. Corey L, Spear PG. Infections with herpes simplex viruses (part 2). *N Engl J Med.* 1986;**314:**749-57.
58. *Policy Research Incorporated Literature Review Validation Procedures Manual.* Baltimore: Policy Research Inc.: 1979.
59. Mullen PD, Ramirez G. Information synthesis and meta-analysis. In: *Advances in Health Education and Promotion.* Vol. 2. Greenwich, Connecticut: Jai Press; 1986.
60. Huth EJ. *How to Write and Publish Papers in the Medical Sciences.* Philadelphia: ISI Press; 1982:64.
61. Morgan PP. Review articles: 2. The literature jungle. *Can Med Assoc J.* 1986;**134:**98-9.