

META-ANALYSIS IN MEDICINE

WHERE WE ARE AND WHERE WE WANT TO GO*

MILOS JENICEK

Department of Social and Preventive Medicine, Faculty of Medicine, University of Montreal,
Quebec, Canada

(Received in revised form 27 June 1988)

Abstract—An epidemiologically impeccable study does not bring answers to all the important questions. A structured and systematic integration of information from different studies of a given problem with a view to answering the original question or bringing additional information is the essence and objective of the meta-analytic approach to health problem solving. Original studies in medicine, being very heterogeneous in nature and structure require not only a quantitative approach (as in classical meta-analysis) but also an additional “qualitative meta-analysis” as well. The latter represents not only a systematic accumulation of both information and the characteristics of different studies, but also an assessment of quality, uncertainty, missing data, random error and bias across studies of interest. The greatest challenge of meta-analysis in medicine lies in the integration of the qualitative and quantitative assessment of given information (scoring of quality, weighing of the effect size by quality score, etc.). Meta-analysis in medicine must go beyond a simple pooling of data. It should become the “epidemiology of results of independent studies of a common topic of interest”. Further development of meta-analysis in such an expanded way may have an important impact on decision-making in clinical medicine, and in health policies.

INTRODUCTION

The best possible synthesis of available information is essential for all decision-makers. It is needed in clinical medicine when one has to face a patient or establish common strategies for groups of similar patients. It is valuable in medical research, where new hypotheses should follow first class information. It is necessary in health planning and in administration, where the most efficient and effective programs and policies have to be established. In addition, classical or field epidemiology needs such a synthetic view, for better etiological studies of disease and for a better control of the spread of disease.

The critical review and assessment of clinical problems and questions across different studies is an essential element in the acquisition of medical knowledge. It is essential for answering

questions such as these: does jogging do more harm than good? Is diethyl-stilbestrol a predominant cause of clear cell vaginal carcinoma? Do beta-blockers prevent death after myocardial infarction?

During the last decade, a more systematic way of evaluating and synthesizing information than a simple narrative review was worked out in other areas. An increasing number of original studies and methodological articles appeared on this subject under different terms [1]: meta-analysis, integrative research review, research integration, research consolidation, data synthesis, research synthesis, quantitative synthesis, quantitative assessment of research domains, combining studies, combining results, empirical cumulation, empirical evidence and others. A new quantitative, organized, statistical approach to research synthesis was born.

The objective of this article is to review the domain of the integration of the results of research (meta-analysis) in medicine. The latter is a result of a migration from psychology and education into the health sciences. The future of

*This article is an expanded version of the paper read at the 11th Scientific Meeting of the International Epidemiological Association, Helsinki, Finland, 11 August 1987.

meta-analysis and its relevance to the advancement of medicine will be discussed. Hence, we wish to discuss basic concepts and strategies that may apply to medicine rather than make an extensive historical review or a condensed capsule of quantitative meta-analytic techniques.

Traditionally, research synthesis was done in a fairly simple way. *Narrative reviews* [2] express the personal opinions of their authors and depend heavily on the perspicacity and personal experience of the reviewer. *Secondary replicative analysis* [3] is based on the return to original data gathered on individuals in various studies; individual record linkage allows them greater statistical power. *Box-score analysis*, or the vote counting method [4] as in sports statistics, compares numbers of studies which confirm a hypothesis under study to numbers of studies which reject it. Such a simplistic approach is not satisfactory either.

Often, a more satisfactory research synthesis such as meta-analysis should go beyond the above mentioned methods. Meta-analysis in medicine is any structured and systematic qualitative and/or quantitative integration of the results of several independent studies on a health problem. To accomplish this, we may pool results of different studies or original observations in individual patients, but we do not gather any new, original observations. Quite a wide array of studies, such as descriptive, etiological or intervention studies, or studies validating clinical tools such as diagnostic methods may be the subject of meta-analysis.

The objectives of meta-analysis are:

- to confirm information (hypothesis, proof, initial findings),
- to find errors,
- to search for additional findings and
- to develop new ideas (hypotheses) for further research and future original studies.

The well organized analysis and synthesis of independent research findings on medical questions were reviewed in several journals: in clinical medicine and therapeutics in particular [5, 6], in community health [7], in epidemiological research [8] and elsewhere, as in education [9, 10], clinical psychology [11, 12], occupational therapy [13, 14] and nursing [15, 16].

ORIGINS OF META-ANALYSIS

More than 10 years ago, several outstanding researchers in psychology and education re-

sented the fragility of even a careful individual and narrative approach to the synthesis of research results. After the first stimulating article on the subject by Light and Smith [17] came Glass's first definition of meta-analysis [18]: the "analysis of analyses", or better, "the statistical analysis of a large collection of analyses results from individual studies for the purpose of integrating the findings". Quantitative methods and techniques were further developed by Rosenthal [19, 20], Hedges [21] and others. The 1983 review annual of Evaluation Studies, edited by R. J. Light [22], brings together an important array of methodological articles and many important original studies. Basic methodological textbooks [23–27] soon followed as well as monographs dealing with statistical methods [28] and computer software [29]. Only recently, an introductory text on meta-analysis in medicine [30] appeared. Ultimately, special chapters on meta-analysis also found their way in recent books on medical statistics [31] and clinical epidemiology [32].

The purpose of quantitative, let us say "classical" meta-analysis was usually to assess effectiveness of treatments, programmes and interventions, and less often to assess cause-effect relationships. For instance, is there any relation between intelligence and schizophrenia in young subjects [33]? Does a full moon affect mental health and behavior [34]? Is psychotherapy effective [32, 35–42]? Can hyperactivity in children be controlled pharmacologically [43, 44] or by dietary regimens [45–47]?

"Classical" or quantitative meta-analysis addresses two basic questions:

- How is one variable related to another?
- How strong is the evidence for the relationship?

To obtain adequate answers, meta-analysts gathered as many published and unpublished studies as possible. The quality of the studies was not taken into consideration; meta-analysts were criticized for that later [24]. Experimental studies in fields other than medicine, such as psychology and education may be more uniform in design and execution; target populations may be less heterogeneous. Medical research is much more heterogeneous and one must adapt meta-analysis to this heterogeneity. Let us see what strategies might be adopted first in quantitative meta-analysis, second, in qualitative meta-

analysis and finally, let us foresee an appropriate synthesis of both.

MEASUREMENT OF EFFECT OR QUANTITATIVE META-ANALYSIS

First, we may define quantitative meta-analysis in medicine as a general, systematic and uniform evaluation of *dimensions*. These may be across studies dealing with the following:

- the magnitude of a health problem,
- the strength and specificity of a causal relationship in etiological research,
- the strength and specificity of the impact of a preventive (contrapathic) or therapeutic (contratrophic) intervention (=“effect size”) or
- the internal and external validity of clinical tools (e.g. diagnostic methods).
- the costs and benefits of diagnostic methods and treatments.

Hence, the subject should not be exclusively the effect of an interesting treatment. Other above mentioned topics may also be studied. The study of causal relationships between independent and dependent variables is based on logic and measurement. This applies just as well to the study of undesirable factors in the causing of disease as to that of the beneficial effect of treatment on outcome.

“Classical” meta-analysis

“Classical” meta-analytic methodology was developed for synthesizing studies of some causal relationship, where the independent variable was some uncontrolled extraneous factor (such as a full moon) or some controlled intervention (psychotherapy, diet, etc.). The dependent variable is most often some quantitative variable such as a score of performance or a level (concentration) of a biochemical or hematological component (e.g. blood lipids). The “effect size” in a study (d , ES) is given by a basic formula, which is the difference in mean outcomes of the experimental (exposed, treated) group and control (unexposed) group, divided by the standard deviation of the outcome in the control group (or by the pooled standard deviation of both groups) [18, 48].

The effect size of each meta-analyzed study becomes a new unit of analysis: the heterogeneity of effect sizes may be assessed and an average effect size across studies computed. Cohen originally considered the effect size of 0.2

as small, 0.5 as average and 0.8 as large [5, 48]. Such an assessment of the importance of effect size is purely arbitrary.

At this moment, the clinical implication could be simply whether the intervention under study “is better” or “has no advantage”.

Greenland [8] warns, that by expressing effects in standard deviation units, one can make studies with identical results spuriously appear to yield different results.

Another problem stems from the habit which meta-analysts have of mathematically converting heterogeneous statistical parameters into effect sizes (d 's). Such a mix of original heterogeneous results, even though properly mathematically blended, is not easy to interpret. What operational meaning of effect should be drawn for clinical decisions from a study giving a p result, another a χ^2 , another a t -value etc.? All these analyses were done primarily to evaluate our confidence in what we saw. These analyses are not a measurement of effect by itself. This part of quantitative meta-analysis is for the moment uninterpretable both in practical terms and as a basis for clinical decisions in practice.

The evaluation and integration of various epidemiological measures of risk are more suitable for meta-analysis in medicine.

How strong is the available evidence—the “file drawer problem”

In many fields of research and in some less respected journals, two kinds of studies have a smaller chance of being published: those which do not reject a null hypothesis and/or studies whose findings are not coherent with current prevailing paradigms or a body of knowledge.

Consequently, a portion of the complete body of information, in the form of unpublished studies, remains “in file drawers” [49–51]. Rosenthal [49] proposed a mathematical estimation of how many additional studies (unpublished or to be published) whose average result would be contrary to conclusions based on published information, would be needed to change such conclusions. If, for a given subject, only two or three studies would be needed to disprove current findings, the evidence would not be considered as strong as for another, in which 50 additional studies would perhaps be needed to change current conclusions.

The strength of the evidence across studies may be evaluated this way.

This procedure obviously does not clarify additional questions such as: have all relevant

studies been found and what should be done about unpublished studies?

“Classical” epidemiology

In classical epidemiology, the effect size or the quantitative dimension of the strength and specificity of causal relationships is given by risk ratios (relative risk, RR), risk differences (attributable risk, AR, RD) or by the proportional expression of the latter relative to exposed subjects or to the target population (etiological fraction, EF, or attributable risk percent, AR%).

These measures and their interpretation highlight any contemporary textbook of epidemiology (see also Refs [52] and [53]).

An alternative methodology based on Efron and Morris’ empirical Bayes approach was proposed and used by Gilbert *et al.* [54] to evaluate benefits and risks in surgery and anesthesia.

Despite this, different epidemiological expressions of “risk” are most often used in etiological research of noxious factors and less often in the assessment of causal relationships between intervention and cure, where sole statistical significance still prevails in many studies and fields.

There are three possible approaches to the estimation of the effect size across studies:

(1) A simple averaging of results of original studies. This is definitely a very simplistic approach, even if studies are stratified and/or weighted according to some preselected criteria.

(2) The characteristic effect across studies is recalculated from cell frequencies of events in two by two tables (or their extensions) from original studies [55–57]. This approach, developed by British authors, stems from the original Mantel–Haenszel approach to stratified observations in case–control studies [52]. Conceptually, strata are replaced by original studies [58]. The resulting relative risk (risk ratio) across studies [55] and typical odds ratio [56, 57] represent another assessment of effect size.

Cochran’s guidelines for combining rates from the individual trials are also used [59].

Also, in terms of attributable risk, an average risk difference across studies may be sought by DerSimonian and Laird’s method [60]. It was used, for example, by Himel *et al.* [61] in the evaluation of adjuvant chemotherapy for breast cancer.

Elsewhere, as proposed by Wortman and Yeaton [62], a difference in frequencies of success between treated and control groups is ex-

pressed as ratios to some denominator, since this is conceptually close to the etiological (preventable) fraction [53]. Some other proportional expression of therapeutic success may be sought [63].

(3) An evaluation of risk may be strengthened by using regression models. The possibility of doing meta-analysis on these findings is reviewed in depth by Greenland [8].

QUALITATIVE META-ANALYSIS

In medicine, contrary to many other fields of research, variables of interest such as methodology, health problems, target populations and etiological factors are considerably diversified and heterogeneous. In such a situation, an adequate assessment of the quality of original studies must be made before a quantitative meta-analysis is performed. Unacceptable studies must be rejected. Some weighing of studies according to quality (score) may be attempted or some other stratification based on quality. The effect size may be evaluated in different strata or across studies with appropriate attention to quality. Such an approach represents at present one of the greatest challenges of meta-analysis in medicine: to find a good integration of qualitative and quantitative aspects of meta-analysis in the synthesis of research. Until now, the question of what studies to include and how to weight them has not been at all settled.

Qualitative meta-analysis in medicine may be defined as “a method of assessment of the importance and relevance of medical information coming from several independent sources through (by) a general, systematic and uniform application of pre-established criteria of acceptability to original studies representing the body of knowledge of a given health problem or question” [30].

The objectives of qualitative meta-analysis are:

- to determine the prevalence, homogeneity and distribution of quality attributes,
- to expand the knowledge of missing and or imperfect data and
- to evaluate and interpret “outliers” (e.g. observations beyond a customary range).

For example, Feinstein’s qualitative criteria for case–control studies [64] represent a structured guideline in the assessment of an “initial state–manoeuvre–subsequent state” model. His

20 prerequisites, if respected, should guarantee freedom from major biases and random errors. In many cases, scoring for quality of the study would be useless if the presence of an important protopathic or susceptibility bias were to invalidate the interpretation of quantitative results of the study.

Elsewhere, Lichtenstein *et al.* [65] use 34 criteria to assess the quality of case-control studies. Twenty items such as methods of data collection, sources of cases and controls, blinding of interviewers, description of sampling and analytic methods, diagnostic procedures and criteria, information on exposure etc. are considered as essential. A quality score is not given.

In many cases, it may not be enough to simply list present or absent attributes of each study. An appropriate dimension must be given to these facts where necessary. Scoring of quality is of particular interest. Chalmers *et al.* [66] proposed a qualitative assessment of clinical trials, where, from a total score of 100, a maximum of 60 points is given to the data base, design and "protocol" with emphasis on blinding, 30 points are given to statistical analysis and 10 points to the way the study is presented.

The McMaster group used another type of scoring to evaluate the quality of compliance research reports [67, 68]. The highest score is given to studies having good internal and external validity (e.g. a randomized trial which is based on a random population sample, in which replicable diagnostic inclusion and exclusion criteria are clearly stated, where direct longitudinal measures of compliance are taken and where compliance and therapeutic regimens are completely described to the reader and replicable by others).

Unfortunately, the published literature does not contain an equivalent method of undertaking the quality assessment of observational analytical studies, giving some "quantitative assessment of quality", which might be used in the meta-analysis of a given problem, where both qualitative and quantitative aspects of evaluation are integrated.

DerSimonian *et al.* [69] evaluate the quality of clinical trials in another way: the use of barycentric-coordinate plots allows these authors to establish a sort of "qualitative somatotype" of studies according to selected qualitative criteria or axes. (N.B. They also give a point score based on the number of items reported).

A recent study of secondary health effects of oral contraceptives, meticulously executed by

Realini and Goldzieher [70], is a good example of a qualitative meta-analysis of a given problem in etiologic research. (The authors' very systematic approach did not include scoring for quality.)

The quality of studies may be considered in two ways in meta-analysis. It may first be viewed as an independent variable and one may examine whether study results depend on their good or bad quality. Secondly, results may be scored by quality of studies, if and only if there is an association. Would such a scoring be necessary if there were not such an association? Should the effect size observed in each study be weighted by the study's quality score and the overall effect size across studies determined thereafter?

Another alternative well worth considering might be choosing the best available evidence only, as proposed by Slavin [71]. Should experimental studies of a causal relationship of interest be the only ones retained for meta-analysis if evidence is also available from observational cohort and/or case-control research?

Gerbarg and Horwitz [72] in their guidelines for the meta-analyses of clinical trials stress as the first step the conduct of a structured and consistent *methodologic analysis* of all available clinical trials and consider for pooling only those that adhere to current standards of methodologic rigor.

In all the above-mentioned approaches, an appropriate sequence appears to be qualitative assessment of studies first, followed by a selection of acceptable or best evidence and the quantitative meta-analysis of the latter.

PROPOSAL FOR A BASIC ARCHITECTURE OF A META-ANALYTIC STUDY IN MEDICINE

From the epidemiological point of view, meta-analysis is and must be the "epidemiology of results". The individual as a unit of observation is replaced by some result of an original study, which is itself becoming a new unit of subsequent study in meta-analysis.

In medicine, a collaborative, multicenter clinical trial or etiologic study (case-control studies are often done this way) is a classical terrain for a meta-analytic approach. Observations of individuals are pooled for the sake of statistical power, separate results from participating centers or hospitals being integrated and further submitted to epidemiological analysis. The classical epidemiological paradigm "persons-

time-place" becomes a new meta-analytic one: "studies-times-places". Meta-analysis in medicine should not only involve the pooling of data, but also the pooling of results, as well as the integration, epidemiological exploration and evaluation of these results.

A meta-analytic study should bring answers to some problem or question which was clearly formulated at the beginning. Such a deductive approach to the subject of interest is by far superior to the inductive one. We should not wait for what will "pop out" from a meta-analytic study. On the other hand, questions other than the original one will appear during the study. These should become the subject of an inductive approach. The deductive and inductive approaches are iterative in research but the meta-analytic study should be developed to answer some specific problem first.

The logical sequence of meta-analysis is qualitative assessment, then, quantitative assessment. Both should be integrated in a logical sequence and frame. The flow chart in Fig. 1 sums up one possible kind of procedure. This model also assumes that the typical or overall effect is not the only subject of interest. The homogeneity of results should also be tested [20], and an additional examination of results carried out. Are there any outliers [73, 74]? Are there some strata of particular interest? Such questions may be at the origin of new hypotheses for further studies and research. We totally agree with Light and Pillemer [26] and Greenland [8] that an analysis of the heterogeneity of studies is likely to bring more important information than some "typical" or "average" effect.

RECOMMENDATIONS FOR THE USE AND DEVELOPMENT OF META-ANALYSIS IN MEDICINE

Two basic recommendations may be proposed, one for qualitative meta-analysis and another for quantitative meta-analysis.

In qualitative meta-analysis, we need an adequate dimensional assessment of the quality of studies other than controlled randomized clinical trials. How are we to assess the quality of descriptive studies (often wrongly called prevalence studies) and the quality of observational analytic cohort studies?

In quantitative meta-analysis, classical epidemiology should explore to a greater extent the measurement of effect size where quantitative variables are studied. Elsewhere, if needed, what

dimension should be given to a descriptive measure such as the incidence or prevalence rate across descriptive studies? What about sensitivity or specificity or the predictive values of a diagnostic tool? How to solve the dilemma between diagnostic subgroups or strata according to some clinimetric classification or an overall "challenging clinical spectrum"?

In addition to this, we should establish qualitative criteria for meta-analytic studies themselves. These criteria might be the presence of explicit information such as:

- a clear definition of the health problem under scrutiny,
- the subject of meta-analysis (effect? outliers? other?),
- questions, hypotheses and objectives of the meta-analytic study,
- identification of the expected contribution to medical practice,
- a list of integrated studies,
- operational criteria of the selection and assessment of studies,
- a clear description of the methodology of the meta-analysis itself,
- a presentation of all relevant qualitative and quantitative results of meta-analysis,
- a suitable review of bias,
- a critical assessment of results for further research, for practice and for health policies if relevant [30].

CONCLUSIONS

Meta-analysis in medicine is in the embryonic stage. The field is barely defined and its specific methods are undeveloped. All this should be done. However, meta-analysis is a rigorous and structured approach to health problems across studies, and therefore in direct opposition to an educated guess.

The notion of the "importance" of analytic and meta-analytic findings is ambiguous. What is statistically important merits epidemiological analysis. What is epidemiologically important is not necessarily clinically important and such scientific importance does not necessarily change clinical decisions. This is true as much for an original study as for a meta-analytic study.

Meta-analysis is still presented mainly as a method of evaluating whether treatment works [75-77]. It should be more than that. Indeed, meta-analysis should be developed and used far

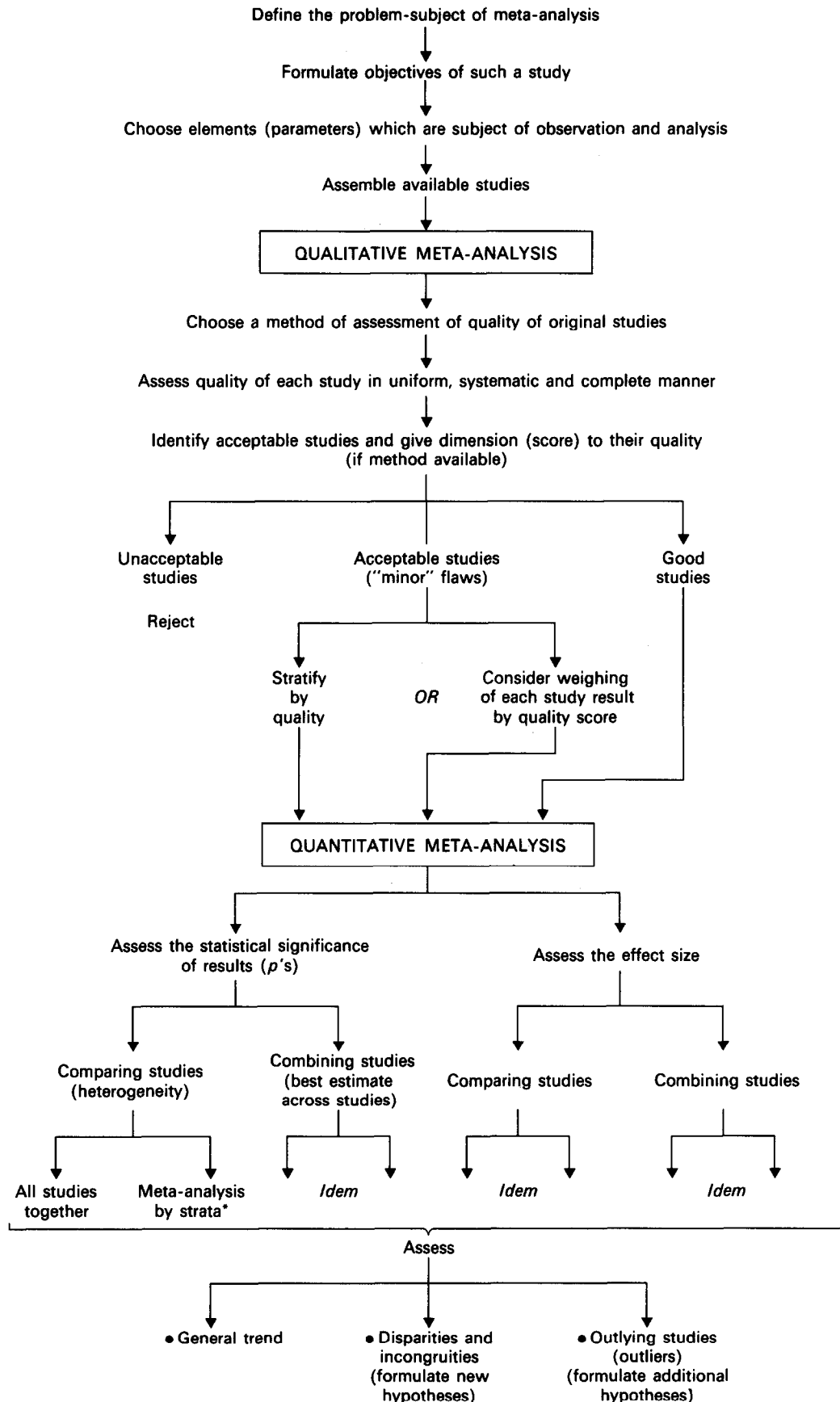


Fig. 1. Flowchart of meta-analysis.

beyond randomized clinical trials. Observational descriptive and analytic studies, and the evaluation of diagnostic methods are also domains of interest. Meta-analysis in these domains should be further developed. In the domain of diagnostic tests, Nierenberg and Feinstein use an innovative approach in assessing the validity of the dexamethasone suppression test. To do this they integrate available studies and test the balance of evidence by making a hierarchical classification of the five phases of development and evaluation for diagnostic marker tests [78].

The analysis of the heterogeneity of studies should not be sacrificed to some global encompassing average or "typical" value. Both should be analyzed. The pooling of results is sometimes inappropriate as the results of some studies may be very heterogeneous. This requires analysis rather than a search for some overall picture which would be hard to interpret. As an example, the preventable fraction and effectiveness of a new vaccine may be evaluated in separate trials in developed and developing countries, in healthy subjects, in malnourished subjects or in individuals suffering from some important comorbidity. In such an instance, it is better to analyze differences, and to draw new hypotheses and test them rather than to try to obtain some universal protective ratio which applies to neither group in the original studies.

Meta-analysis should not be seen exclusively as a new tool in etiological and intervention research. An even more important field may be health policies and health programs, tactics and strategies at the hospital and in the community. Already available studies of medical care [79-83], of process [84] or of the impact of medical decisions [85-87] confirm that. The cost-effectiveness of interventions also becomes open to meta-analysis [88].

The following problems, which have been at least partly tackled in this article, should be better clarified in the future:

- How does one determine if all relevant studies have been found?
- How does one determine which studies should be combined?
- How can the meta-analysts' biases be minimized?
- Which statistical techniques are most appropriate?
- What should be done about unpublished studies?

The limitations and advantages of meta-analysis have been widely discussed in the literature [24, 30, 41, 75-77, 89-92]. One of the main challenges remains the risk that meta-analytic methodology, impressive and at the same time understandable to a larger professional audience, will be used indiscriminately, inductively, and without operational criteria of inclusion and exclusion on both dependent and independent variables. A very hard look at the data must be taken first [71, 72, 93, 94].

We will always be making decisions about whether to do another, possibly better, study of the same subject or, remembering a message from our childhood at a railway crossing, whether to "stop, look and listen". In other words, we must slow down from time to time to reassess currently available medical knowledge. Such a meta-analytic approach is even more important for decision makers in governments and institutions, that is, wherever health policies are decided, proposed and evaluated. Institutions are seldom producers of first hand information. However, they are responsible for the best possible evaluation of current information and for the best possible implementation of adequate health policies and programs. This should not be a subject of improvisation or solely a matter of serendipity and flair. Even if political decisions finally override recommendations, well organized homework is needed. Better political decisions will follow.

Acknowledgements—The author is indebted to Mrs Louise Valois for editorial assistance and to Mrs Johanne Roy for technical assistance in the preparation of this manuscript.

Special thanks go to three reviewers of the Journal. Their expertise in the fast evolving field of meta-analysis and their challenging suggestions played a determining role in the final version of the manuscript.

REFERENCES

1. Feldman KA. Review of "Issues in Data Synthesis". (Yeaton WH, Wortman PM, Eds. San Francisco: Jossey-Bass; 1984). *J Am Stat Ass* 1986; 81: 253-254.
2. Zautra A, Hempel A. Subjective well-being and physical health: A narrative literature review with suggestions for future research. *Int J Aging Hum Dev* 1984; 19: 95-110.
3. George LK, Landerman R. Health and subjective well-being: A replicated secondary data analysis. *Int J Aging Hum Dev* 1984; 19: 133-156.
4. Hedges LV, Olkin I. Vote-counting methods in research synthesis. *Psychol Bull* 1980; 88: 359-369.
5. Einerson TR, McGhan WF, Sabers DL. Meta-analysis: Quantitative integration of independent research results. *Am J Hosp Pharm* 1985; 42: 1957-1964.
6. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316: 450-455.

7. Louis TA, Fineberg HV, Mosteller F. Findings for public health from meta-analysis. *Ann Rev Public Health* 1985; 6: 1-20.
8. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987; 9: 1-30.
9. Pillemer DB, Light RJ. Synthesizing outcomes: How to use research evidence from many studies. *Harvard Educ Rev* 1980; 50: 176-195.
10. Kavale KA, Glass GV. Meta-analysis and the integration of research in special education. *J Learn Dis* 1981; 14: 531-538.
11. Strube JJ, Hartmann DP. Meta-analysis: Techniques, applications, and functions. *J Consult Clin Psychol* 1983; 51: 14-27.
12. Fiske DW. The meta-analytic revolution in outcome research. *J Consult Clin Psychol* 1983; 51: 65-70.
13. Ottenbacher K. Quantitative Reviewing: The literature review as scientific inquiry (research, statistical procedures, methodology). *Am J Occup Ther* 1983; 37: 313-319.
14. Ottenbacher K, York J. Strategies for evaluating clinical change: Implications for practice and research. *Am J Occup Ther* 1984; 38: 647-659.
15. O'Flynn AI. Meta-analysis. *Nurs Res* 1982; 31: 314-316.
16. Curlette WL, Cannella KS. Going beyond the narrative summarization of research findings: The meta-analysis approach. *Res Nurs Health* 1985; 8: 293-301.
17. Light RJ, Smith PV. Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educ Rev* 1971; 41: 429-471.
18. Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976; 5: 3-8.
19. Rosenthal R. Combining results of independent studies. *Psychol Bull* 1978; 85: 185-193.
20. Rosenthal R. Assessing the statistical and social importance of the effects of psychotherapy. *J Consult Clin Psychol* 1983; 51: 4-13.
21. Hedges LV. *Statistical Methodology in Meta-Analysis*. Princeton: ERIC Clearinghouse on Tests, Measurement and Evaluation, Educational Testing Service; 1982.
22. *Evaluation Studies. Review Annual*. Light RJ, Ed. Beverly Hills, Calif.: Sage Publications; 1983: Vol. 8.
23. Glass GV, McGaw B, Smith ML. *Meta-Analysis of Social Research*. Beverly Hills, Calif.: Sage Publications; 1981.
24. Hunter JE, Schmidt FL, Jackson GB. *Meta-Analysis: Cumulating Research Findings Across Studies*. Beverly Hills, Calif.: Sage Publications; 1982.
25. Borg WR, Gall MD. *Critical Evaluation of Research, Educational Research*. An Introduction. New York: Longman; 1983: 4th edn.
26. Light JRJ, Pillemer DB. *Summing-Up. The Science of Reviewing Research*. Cambridge, Mass.: Harvard University Press; 1984.
27. Rosenthal R. *Meta-analytic Procedures for Social Research*. Beverly Hills, Calif.: Sage Publications; 1984.
28. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. New York: Academic Press; 1985.
29. Mullen B, Rosenthal R. *Basic Meta-Analysis: Procedures and Programs*. Hillsdale, N.J.: Lawrence Erlbaum Associates; 1985.
30. Jenicek M. *Meta-analyse en médecine. Evaluation et Synthèse de l'Information Clinique et Épidémiologique*. St-Hyacinthe et Paris: EDISEM et Maloine; 1987.
31. Halvorsen KT. Combining results from independent investigations: Meta-analysis in medical research. In: *Medical Uses of Statistics*. Bailar JC III, Mosteller F. Eds., Waltham, Mass.: NEJM Books; 1986: 392-416.
32. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology—the Essentials*. (Second edition). Baltimore: Williams & Wilkins; 1988: 2nd edn, Chapter 12. Summing up, pp. 226-240.
33. Aylward E, Walker E, Bettles B. Intelligence in schizophrenia: Meta-analysis of the research. *Schizophr Bull* 1984; 10: 430-459.
34. Rotton J, Kelly IW. Much ado about full moon: A meta-analysis of lunar-lunacy research. *Psychol Bull* 1985; 97: 286-306.
35. Smith JL, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977; 37: 752-760.
36. Shapiro DA, Shapiro D. Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychol Bull* 1982; 92: 581-604.
37. Steinbrueck SM, Maxwell SE, Howard GS. A meta-analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults. *J Consult Clin Psychol* 1983; 51: 856-863.
38. Prioleau L, Murdock M, Brody N. An analysis of psychotherapy versus placebo studies. *Behav Brain Sci* 1983; 6: 275-310.
39. Glass GV, Kliegl RM. An apology for research integration in the study of psychotherapy. *J Consult Clin Psychol* 1983; 51: 28-41.
40. Shapiro DA, Shapiro D. Comparative therapy outcome research: Methodological implications of meta-analysis. *J Consult Clin Psychol* 1983; 51: 42-53.
41. Wilson TG, Rachman SJ. Meta-analysis and the evaluation of psychotherapy outcome: Limitations and liabilities. *J Consult Clin Psychol* 1983; 51: 54-64.
42. Weissz JR, Weiss B, Alicke MD, Klotz ML. Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *J Consult Clin Psychol* 1987; 55: 542-549.
43. Kavale K. The efficacy of stimulant drug treatment for hyperactivity: A meta-analysis. *J Learn Disabil* 1982; 15: 280-288.
44. Thurber S, Walker CE. Medication and hyperactivity: A meta-analysis. *J Gen Psychol* 1983; 108: 79-86.
45. Kavale KA, Forness SR. Hyperactivity and diet treatment: A meta-analysis of the Feingold hypothesis. *J Learn Disabil* 1983; 16: 324-330.
46. Rimland B. The Feingold diet: An assessment of the reviews by Mattes, by Kavale and Forness and others. *J Learn Disabil* 1983; 16: 331-333.
47. Mattes J. The Feingold diet: A current reappraisal. *J Learn Disabil* 1983; 16: 319-323.
48. Cohen J. *Statistical Power Analysis For the Behavioral Sciences*. New York: Academic Press; 1977.
49. Rosenthal R. The "File Drawer Problem" and tolerance for null results. *Psychol Bull* 1979; 86: 638-641.
50. Orwin RG, Cordray DS. Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychol Bull* 1985; 97: 134-147.
51. Simes RJ. Confronting publication bias: A cohort design for meta-analysis. *Stat Med* 1987; 6: 11-29.
52. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; 22: 719-748.
53. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol* 1974; 99: 325-332.
54. Gilbert JP, McPeck B, Mosteller F. Progress in surgery and anesthesia: Benefits and risks of innovative therapy. In: *Costs, Risks, and Benefits of Surgery*. Bunker JP, Barnes BA, Mosteller F, Eds. New York: Oxford University Press; 1977: 156-161.
55. Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH. Effect of intravenous streptokinase on acute myocardial infarction. Pooled results from randomized trials. *N Engl J Med* 1982; 307: 1130-1182.
56. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: An

- overview of the randomized trials. *Prog Cardiovasc Dis* 1985; 27: 335-371.
57. Collins R, Langman M. Treatment with histamine H₂ antagonists in acute upper gastrointestinal hemorrhage. *N Engl J Med* 1985; 313: 660-706.
 58. Janicak PG, Davis JM, Gibbons RD, Ericksen S, Chang S, Gallagher P. Efficacy of ECT: A meta-analysis. *Am J Psychiat* 1985; 142: 297-302.
 59. Colditz GA, Tuden RL, Oster G. Rates of venous thrombosis after general surgery: Combined results of randomised clinical trials. *Lancet* 1986; 2: 143-146.
 60. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Contr Clin Trials* 1986; 7: 177-188.
 61. Himel HN, Liberati A, Gelber RD, Chalmers TC. Adjuvant chemotherapy for breast cancer. A pooled estimate based on published randomized clinical trials. *JAMA* 1986; 256: 1148-1159.
 62. Wortman PM, Yeaton WH. Cumulating quality of life results in controlled trials of coronary artery bypass graft surgery. *Contr Clin Trials* 1985; 6: 289-305.
 63. Gent M, Roberts RS. A meta-analysis of the studies of dihydroergotamine plus heparin in the prophylaxis of deep vein thrombosis. *Chest* 1986; 89 (Suppl.): 396S-400S.
 64. Twenty scientific principles for trohoc research. In: Feinstein AR, *Clinical Epidemiology. The Architecture of Clinical Research*. Philadelphia: WB Saunders; 1985: 543-547.
 65. Lichtenstein MJ, Mulrow CD, Elwood PC. Guidelines for reading case-control studies. *J Chron Dis* 1987; 40: 893-903.
 66. Chalmers TC, Smith Jr. H, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized clinical trial. *Contr Clin Trials* 1981; 2: 31-49.
 67. Haynes RB, Sackett DL in coll. with Taylor DW, Hackett BC, Luterbach E, Cloak JR. Appendix I: An annotated bibliography (including notes on methodological standards for compliance research). In: Sackett DL, Haynes RB, Eds. *Compliance with Therapeutic Regimens*. Baltimore and London: The Johns Hopkins University Press; 1976: 193-199.
 68. Haynes RB, Taylor DW, Snow JC, Sackett DL in coll. with Tugwell PJ, Walsh M, Hackett BC, Mukherjee J. Appendix I: Annotated and indexed bibliography on compliance with therapeutic and preventive regimens. In: Haynes RB, Taylor DW, Sackett DL, Eds. *Compliance in Health Care*. Baltimore and London: The Johns Hopkins University Press; 1979: 337-343.
 69. DerSimonian R, Charette J, McPeck B, Mosteller F. Reporting methods in clinical trials. *N Engl J Med* 1982; 306: 1332-1337.
 70. Realini JP, Goldzieher JW. Oral contraceptives and cardiovascular disease: A critique of the epidemiologic studies. *Am J Obstet Gynecol* 1985; 152: 729-798.
 71. Slavin RE. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. *Educ Res* 1986; 15: 5-11 (reprinted: *Eval Stud Rev Ann* 1987; 12: 667-673).
 72. Gerbarg ZB, Horwitz RI. Resolving conflicting clinical trials: guidelines for meta-analysis. *J Clin Epidemiol* 1988; 41: 503-509.
 73. Bolton S. Outliers. In: *Pharmaceutical Statistics. Practical and Clinical Applications*. New York: Marcel Dekker; 1984: 294-300, 485.
 74. Canner PL, Huang YB, Meinert CL. On the detection of outlier clinics in medical and surgical trials. I. Practical consideration *Contr Clin Trials*, 1981; 2: 231-240. II. Theoretical considerations. *Ibid.* 241-252.
 75. L'Abbé KA, Detsky SA, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; 107: 224-233.
 76. Naylor CD. Two cheers for meta-analysis: problems and opportunities in aggregating results of clinical trials. *Can Med Ass J* 1988; 138: 891-895.
 77. Thacker SB. Meta-analysis. A quantitative approach to research integration. *JAMA* 1988; 259: 1685-1689.
 78. Nierenberg AA, Feinstein AR. How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 1988; 259: 1699-1702.
 79. Berk AA, Chalmers TC. Cost and efficacy of the substitution of ambulatory for inpatient care. *N Engl J Med* 1981; 304: 393-397.
 80. Blackburn BA, Smith H Jr, Chalmers TC. The inadequate evidence for short hospital stay after hernia or varicose vein stripping surgery. *Mt Sinai J Med* 1982; 49(5): 383-390.
 81. Devine EC, Cook TD. A meta-analytic analysis of effects of psychoeducational interventions on length of postsurgical hospital stay. *Nurs Res* 1983; 32: 267-274.
 82. Mumford E, Schlesinger HJ, Glass GV, Patrick C, Cuedon T. A new look at evidence about reduced cost of medical utilization following mental health treatment. *Am J Psychiat* 1984; 141: 1145-1158.
 83. Sacks HS, Chalmers TC, Berk A, Reitman D. Should mild hypertension be treated? An attempted meta-analysis of the clinical trials. *Mt Sinai J Med* 1985; 52: 265-270.
 84. Posavac EJ, Sinacore JM, Brotherton SE, Helford MC, Turpin RS. Increasing compliance with medical treatment regimens. A meta-analysis of program evaluation. *Eval Health Prof* 1985; 8: 7-22.
 85. Mumford E, Schlesinger HJ, Glass GV. The effects of psychological intervention on recovery from surgery and heart attacks: An analysis of literature. *Am J Public Health* 1982; 72: 141-151.
 86. Mazzuca S. Does patient education in chronic disease have therapeutic value? *J Chron Dis* 1982; 35: 521-529.
 87. Detsky AS, Baker JP, O'Rourke K. Perioperative parenteral nutrition: A meta-analysis. *Ann Intern Med* 1987; 107: 195-203.
 88. Oster G, Huse DM, Delea TE, Colditz GA. Cost-effectiveness of nicotine gum as an adjunct to physician's advice against cigarette smoking. *JAMA* 1986; 256: 1315-1318.
 89. Mintz J. Integrating research evidence: A commentary on meta-analysis. *J Consult Clin Psychol* 1983; 51: 71-75.
 90. Curlette WL, Cannella KS. Going beyond the narrative summarization of research findings: The meta-analysis approach. *Res Nurs Health* 1985; 8: 293-301.
 91. Strube MJ, Hartmann DP. A critical appraisal of meta-analysis. *Br J Clin Psychol*, 1982; 21: 129-139.
 92. Littenberg B. Aminophylline treatment in severe, acute asthma. A meta-analysis. *JAMA* 1983; 249: 2362-2369.
 93. Clemens JD, Chong JJH, Feinstein AR. The BCG controversy. A methodological and statistical re-appraisal. *JAMA* 1983; 249: 2362-2369.
 94. Goldman L, Feinstein AR. Anticoagulants and myocardial infarction. The problems of pooling, drowning and floating. *Ann Intern Med* 1979; 90: 92-94.