Meta-analysis in epidemiology

D. YACH

Summary

Meta-analysis is the structured and systematic qualitative and quantitative integration of the results of several independent studies (i.e. the epidemiology of results). As in any epidemiological study, a meta-analysis needs to start with clearly stated aims and objectives. Attention needs to be paid to selection bias in selecting the study population (all publications on the topic). An initial qualitative assessment (conducted blinded to results) categorises projects on the basis of their methods, as unacceptable (dropped from later analysis) and acceptable or good. Further analysis could be conducted by stratifying or weighting independent studies according to preset quality criteria. The quantitative assessment involves deriving a pooled measure of outcome (usually the relative or attributable risk). Tests for heterogeneity are required before pooling. By pooling the results from many settings using different methods, the ability to generalise them in terms of their public health relevance is increased.

S Afr Med J 1990; **78:** 94-97.

The term 'meta-analysis' has come into popular use in epidemiological publications over the past 3 - 4 years.[1-7] Originally, its use was confined to the educational and social sciences.[8,9] The term refers to the 'epidemiology of results'. Stated more formally, Jenicek[3] has defined meta-analysis as 'the structured and systematic qualitative and quantitative integration of the results of several independent studies'. For epidemiologists, in meta-analysis the unit of observation is the study, as opposed to an individual or community.

A cursory examination of studies published in the *SAMJ* suggests that authors frequently fail to compare their results with those of other investigators but rather report them as if they are unique. In reality, most research involves re-search! Accumulating knowledge needs to be synthesised to facilitate an understanding of truth. This review will highlight the purpose of meta-analyses and outline the methods used in order to encourage the use of such methods locally.

Aims

The purpose of conducting meta-analyses is to: (*i*) increase the power and precision of estimates from numerous smaller studies by pooling the total sample sizes across a number of investigations; (*ii*) explain contradictory results that arise between studies conducted in different settings; and (*iii*) conduct subgroup analyses that could not be conducted, or were not thought of, in the initial studies.

A meta-analysis should complement the usual chit-chat narrative review in the 'discussion' section of an article or in theses. It is particularly useful when summarising results of several studies for health policy-makers.

Centre for Epidemiological Research in Southern Africa, Medical Research Council, Parowvallei, CP
D. YACH, M.B. CH.B., B.SC. HONS (EPIDEMIOL.), M.P.H.

Accepted 30 Nov 1989.
Based on a paper presented at the 8th Epidemiological Conference, Durban, 5 - 7 July 1989.

Methods

Fig. 1 shows a simplified outline of the process of conducting a meta-analysis.[3] A similar approach to that which one would take to conduct an individual epidemiological study needs to be followed in a meta-analysis.

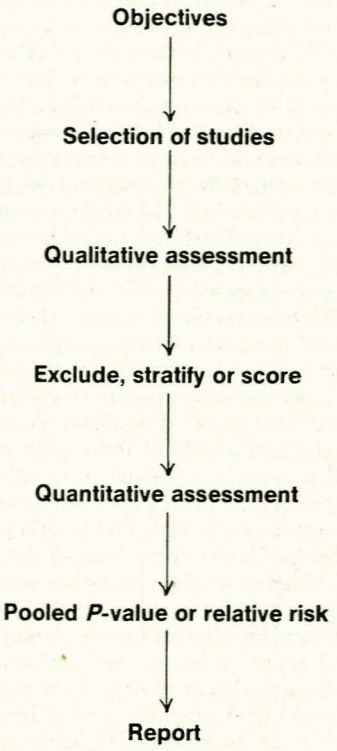

Fig. 1. A simplified approach to meta-analysis (Jenicek[3]).

Objectives

Clearly set objectives are needed. For example, the researchers need to specify whether the major focus is on: (*i*) producing a single pooled result; (*ii*) determining whether there are major subgroup differences; or (*iii*) explaining whether contradictory results may be the result of methodological variability.

Selection of studies

Selection of studies into a meta-analysis is as critical as the selection of individuals in an epidemiological study. In meta-analysis the researcher has several options. Published material can be systematically obtained by first conducting a formal computerised literature search and then continuing by cross-checking key references until no further new published articles are obtained. There are obviously several problems with this approach. In many instances it is difficult to distinguish between independent studies and those that are repeated several times in the literature.

In addition, it is often difficult to obtain foreign journals. The use of only a MEDLINE search (Index Medicus) may exclude a number of articles included under other databases (SCISEARCH or EMBASE, for example). Published articles may also appear in local journals, such as the *South African Journal of Epidemiology and Infection*, which are not yet indexed.

The major problem, however, relates to the large and unknown number of unpublished reports lying in the file-drawers of researchers.[8] Articles may be unpublished for several reasons that could bias the results of a study. For example, individual studies showing negative associations or having inconclusive results are less likely to be published than those showing positive associations. Incomplete studies where the researcher lost interest obviously do not get published. These problems result in publication bias,[1,3,8] which is equivalent to selection bias in an epidemiological study.

Several approaches have been suggested to identify and control for publication bias,[2,8,10] none of which are completely effective. The only way to reduce publication bias is to prevent it! This can be achieved by establishing a register of all research undertaken in a specific area. Already this exists for cancer clinical trials and Medical Research Council-funded research. Follow-up of all trials or studies assembled at the point of funding will allow for a true assessment of the magnitude and direction of publication bias.

Qualitative assessment

The next step is to conduct a qualitative assessment of the literature. Ideally, this should be carried out in a standardised manner where the reviewers of articles are blinded to the results.[3,11] Attention to inter- and intra-observer variation should be built into this process. Such a method review requires the use of a checklist similar to that proposed by Feinstein[12] for case-control studies and attention to the key methodological issues associated with varying designs.[13–16] In many instances it may be necessary to contact authors to obtain further details about methodological aspects of the study that were not included in the report.

The qualitative assessment could lead to a decision to exclude those articles that are not acceptable on methodological grounds. Those that are acceptable could be stratified into groups according to the quality of their methods and further analyses conducted within the strata.[3] An alternative approach involves using some form of method scoring system. The score would later be used as an explanatory variable in the final analyses.

An example of a rigorous qualitative assessment of eight published BCG randomised control trials was published in 1983.[11] Table I indicates a summary of the results. The studies were conducted in numerous settings. The authors identified the absence or presence of four key biases. They further made a decision as to whether the precision of the result would be suitable to obtain useful information. The highest vaccine efficacies were found in the three studies with the fewest biases and the greatest potential for showing high precision.

A preliminary meta-analysis by the author, using the eight published case-control studies[17–24] in peer-reviewed journals (up to March 1989), evaluated whether vaccine efficacy was related to certain quality criteria (Table II). The highest vaccine efficacy results for all cases of tuberculosis were found for those studies using what is regarded as the best quality information on the intervention, i.e. a BCG record. When a case-definition (tuberculous meningitis) was used that is least likely to result in misclassification and under-notification, the highest vaccine efficacy result was obtained. This illustrates two points. Firstly, stratification by quality may indicate the effect and direction that misclassification can have on the outcome. Secondly, since the studies are independent and carried out in separate settings, the results themselves may be a true reflection of the different effects of an intervention. In this case, it is likely that BCG provides better protection against tuberculous meningitis than against all cases of tuberculosis.

TABLE I. QUALITATIVE ASSESSMENT OF BCG RANDOMISED CONTROLLED TRIALS*

Study setting	4 key biases†	Adequate level of statistical precision	Vaccine efficacy (%)
North American Indians	**0**	**Yes**	**80**
England	**0**	**Yes**	**76**
Chicago, USA	**½**	**Yes**	**75**
Puerto Rico	**3**	**No**	**29**
India	**3**	**No**	**20**
Georgia-Alabama, USA	**3**	**No**	**6**
Chingleput	**2½**	**Yes**	**−32**
Georgia, USA	**3**	**No**	**−56**

*Clemens, 1983.[11]
†Susceptibility, surveillance, diagnostic testing and diagnostic interpretation. A score of 0 means that all biases were absent; 1 point is allocated for each bias and ½ if it is 'probably present'.

TABLE II. STRATIFICATION BY QUALITY CRITERIA OF BCG VACCINE EFFICACY FROM CASE-CONTROL STUDIES

	Vaccine efficacy (%)*
All TB, BCG	**56**
All TB, BCG scar	**53**
All TB, BCG record	**66**
TB meningitis	**82**

*Vaccine efficacy = (1 — odds ratio) × 100 in a case-control study.
TB = tuberculosis.

Quantitative assessment

Usually, the major interest is some level of overall statistical significance or the pooled measure of effect (a relative risk, odds ratio, risk ratio or attributable risk). In addition, the quantitative assessment may examine heterogeneity between studies or determine whether there are interactive effects between subgroups.

Greenland[25] reviewed a number of quantitative methods that can be used. Often the crude data are presented in published reports. In that case, the direct relative risk and its standard error can be extracted. Sometimes, however, the relative risk and standard error have to be calculated directly or indirectly from the published report. For example, test-based methods are available to calculate the standard error if only the relative risk and the *P*-value are given.[25] Similarly, if only the upper and lower confidence limits of the relative risk are given, approximations can be used to calculate the standard error. Further, several approaches can be used to adjust for common confounders when such adjustments have not been made in the original text.[25]

A single pooled result can be obtained by lumping or collapsing data over all strata and studies. This approach is *not* recommended under any circumstances and can lead to spurious conclusions.[10] For example, one study may produce a highly significant protective effect, a second an equal but opposite effect. Collapsing could result in a relative risk equal

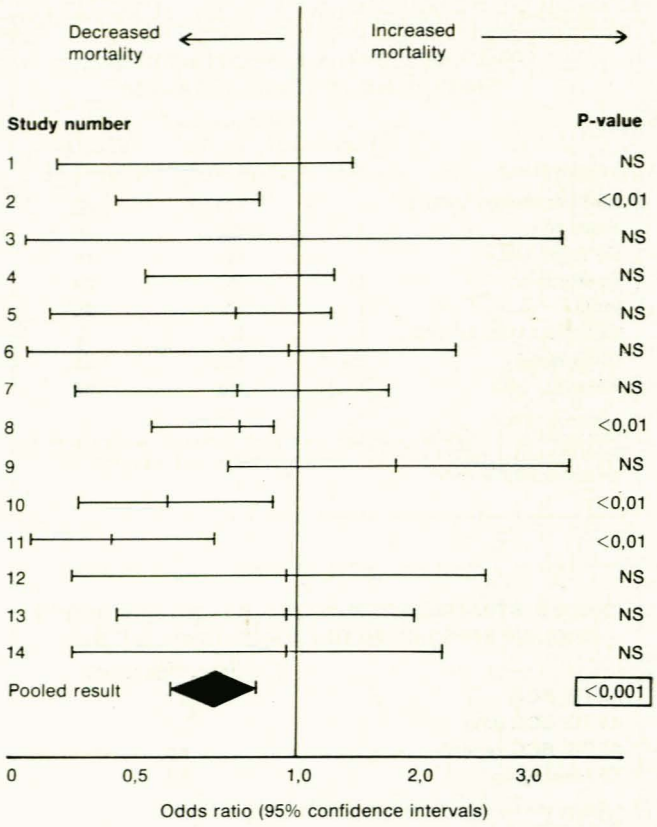

Fig. 2. Relationship between mortality and drug use in 14 published articles.

to 1, indicating no effect. This is clearly a poor summary of the results of two studies.

Graphical methods of display, using the relative risk and 95% confidence intervals for individual studies, are a recommended method of summarising the results from several studies (Fig. 2).[10] This is equivalent to conducting an exploratory data analysis before definitive analyses in individual studies.

In Fig. 2 the results of several studies looking at the relationship between a certain drug and mortality are shown. For many individual studies, non-significant results were obtained. However, the overall pooled result has a high level of precision. It should be noted that the scale used for a graphical display needs to take account of the asymmetrical nature of the confidence intervals of the relative risk. For example, a relative risk of 0,2 for the exposed to the unexposed is equivalent in terms to a relative risk of 5 for the unexposed to the exposed. It is recommended that either the reciprocal or the natural log of the relative risk be used for graphical displays.[26] The graphical display can also be used to identify outliers.[27]

A pooled level of statistical significance can be obtained using several formulae. A simple recommended approach is to calculate $-2 \Sigma \log (P_1)$ where P_1 are the individual P-values. The weakness of using a pooled P-value is that it does not distinguish between protective or non-protective effects, i.e. the direction of association is not indicated.

Several methods are available to calculate the weighted or pooled relative risk, the most recommended weight being the reciprocal of the variance of the relative risk.[2,3,10] This was used in a recently published article on the relationship between stroke and smoking.[2] Peto[28] recommends that for randomised controlled trials the most easily interpretable as well as statistically acceptable method involves a summation of the observed versus the expected number of outcomes (events) for each independent study.[28]

Table III gives an example of the use of the meta-analysis in exploring subgroup analyses for the relationship between smoking and stroke.[2] The overall pooled relative risk of 1,51 provides useful information about the overall risk. Interactive effects are often more important. For example, gender and age differences are highlighted in Table III as well as the dose-response relationship for cigarette smoking and stroke. This would strengthen an assessment of causality.

TABLE III. SUBGROUP ANALYSES TO DETERMINE THE RELATIONSHIP BETWEEN SMOKING AND STROKE IN SEVERAL PUBLISHED STUDIES*

	Relative risk
Overall result	**1,51 (1,45 - 1,58)**
Male	**1,43**
Female	**1,72**
Age (yrs)	
< 55	**2,94**
55 - 74	**1,75**
⩾ 75	**1,11**
Cigarettes (/d)	
< 10	**1,37**
10 - 19	**1,45**
⩾ 20	**1,82**

*Shinton.[2]

Conclusion

Meta-analyses provide a framework for conducting literature reviews. They are also a useful way of comparing the results of a single new study (the researcher's new result) with those from other publications.

An important outcome of the use of meta-analysis has been to provide a firmer basis for policy recommendations on key issues. For example, it is unlikely that a randomised controlled trial will be conducted for BCG. However, several case-control studies reported here, as well as follow-up studies, could be subjected to a meta-analysis to try to determine the likely beneficial effect of continued BCG vaccination.

Further work is required to determine how best to conduct meta-analyses for non-experimental designs. In these studies the biases reduced by the randomisation process of randomised controlled trials are of concern. In addition, there is less standardisation of the intervention compared with a randomised controlled trial where a specified drug dose given at a certain frequency can be examined.

An important final reason for meta-analysis is its potential use for planning future studies. Meta-analysis should ideally form the basis of any literature review at the beginning of a research protocol.

REFERENCES

1. Ozminkowski RJ, Wortman PM, Roloff DW. Inborn/outborn status and neonatal survival: a meta-analysis on non-randomised studies. *Stat Med* 1988; **7:** 1207-1221.
2. Shinton R, Beevers G. Meta-analysis of relation between cigarette smoking and stroke. *Br Med J* 1989; **298:** 789-794.
3. Jenicek M. Meta-analysis in medicine: where we are and where we want to go. *J Clin Epidemiol* 1989; **42:** 35-44.
4. Sacks AS, Berrier J, Reitman D, Ancona-Berk BA, Chalmers TS. Meta-analysis of randomised control trials. *N Engl J Med* 1987; **316:** 450-455.
5. Bulpitt CJ. Meta-analysis. *Lancet* 1988; **2:** 93-94.
6. Naylor CD. Two cheers for meta-analysis: problems and opportunities in aggregating results at clinical trials. *Can Med Assoc J* 1988; **138:** 191-195.
7. Saloojee Y, Simon R, Ellenberg S, eds. Proceedings of the workshop on methodologic issues and overviews of randomised clinical trials, May 1986. *Stat Med* 1987; **6:** 217-404.

8. Light RJ. Accumulating evidence from independent studies: what we can win and what we can lose. *Stat Med* 1987; **6:** 221-228.
9. Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976; **5:** 328.
10. Demets DL. Methods for combining randomised clinical trials: strengths and limitations. *Stat Med* 1987; **6:** 341-348.
11. Clemens JD, Chuong JJH, Feinstein AL. The BCG controversy: a methodological and statistical re-appraisal. *JAMA* 1983; **249:** 2362-2369.
12. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research.* Philadelphia: WB Saunders, 1985: 543-547.
13. Botha JL, Yach D. Epidemiological research methods: II. Descriptive studies. *S Afr Med J* 1986; **70:** 766-772.
14. Botha JL, Yach D. Epidemiological research methods: III. Randomised controlled trials (for interventions). *S Afr Med J* 1987; **71:** 657-660.
15. Yach D, Botha JL. Epidemiological research methods: IV. Case-control studies. *S Afr Med J* 1987; **71:** 763-766.
16. Yach D, Botha JL. Epidemiological research methods: V. Follow-up studies. *S Afr Med J* 1987; **72:** 266-269.
17. Miceli I, De Kantor IN, Colaiacoyo D *et al.* Evaluation of the effectiveness of BCG vaccination in using the case-control method in Buenos Aires, Argentina. *Int J Epidemiol* 1988; **17:** 629-634.
18. Myint TT, Win H, Aye HH, Kyaw-Mint TO. Case-control study on evaluation of BCG vaccination of newborn in Rangoon, Burma. *Ann Trop Paediatr* 1987; **7:** 159-166.
19. Packe GE, Innes JA. Protective effect of BCG vaccination in infant Asians: a case-control study. *Arch Dis Child* 1988; **63:** 277-281.
20. Camargos PAM, Guimaraes MDC, Antunes CMF. Risk assessment for acquiring meningitis tuberculosis among children not vaccinated with BCG: a case-control study. *Int J Epidemiol* 1988; **17:** 193-197.
21. Blin P, Delolme HG, Heyraud JD, Charpak Y, Sentilhes L. Evaluation of the protective effect of BCG vaccination by a case-control study in Yaounde, Cameroon. *Tubercle* 1986; **67:** 283-288.
22. Kue Young T, Hershfield ES. A case-control study to evaluate the effectiveness of mass neonatal BCG vaccination among Canadian Indians. *Am J Public Health* 1986; **76:** 783-786.
23. Shapiro C, Cook N, Evans D *et al.* A case-control study of BCG and childhood tuberculosis in Cali, Columbia. *Int J Epidemiol* 1985; **14:** 441-446.
24. Padungchan S, Konjanart S, Kasiratta S, Daramas S, Ten Dam HG. The effectiveness of BCG vaccination of the newborn against childhood tuberculosis in Bangkok. *Bull WHO* 1986; **64:** 247-258.
25. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987; **9:** 1-30.
26. Hebert JR, Miller DR. Plotting and discussion of rate ratios and relative risk estimates (Letter). *J Clin Epidemiol* 1989; **42:** 289-290.
27. Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. *Am J Public Health* 1988; **78:** 961-966.
28. Peto R. Why do we need systematic overviews of randomised trials? *Stat Med* 1987; **6:** 233-240.