ORIGINAL ARTICLES

The use and misuse of meta-analysis

F E Lecky, R A Little, P Brennan

Abstract

Objective—To demonstrate how the results of a meta-analysis can confuse rather than clarify therapeutic dilemmas if clinical heterogeneity among trials is ignored. Then to further discuss the qualities emergency physicians should expect from published meta-analyses if they are to affect clinical practice.

Subjects and methods—The data and results were examined from 23 randomised controlled trials of selective decontamination of the digestive tract (SDD), which have been combined in a previous meta-analysis. These were reviewed to take account of clinical heterogeneity, particularly with regard to severity of patient illness.

Results—Severity of patient illness predicts degree of reduction in mortality with SDD in a regression analysis: log odds ratio (OR) of death with SDD = $-0.0074 - (0.0035 \times \text{control group mortality rate})$, $P = 0.017$. This is also true when trials are stratified into more and less severely ill patients: pooled OR (a) for CMR > 41% = 0.69 (0.54 to 0.89), with (b) CMR < 37% = 1.02 (0.86 to 1.21). This difference was not suggested by the original meta-analysis result.

Conclusions—Failure to take account of clinical heterogeneity between trials can mean a meta-analysis result ignores important differences in the effect of a treatment on different groups of patients. The discussion indicates how emergency physicians might guard against basing clinical practice on misleading meta-analysis results.

(J Accid Emerg Med 1996;13:373-378)

Key terms: meta-analysis; randomised controlled trials; selective decontamination

Meta-analysis is a statistical technique for combining the results of published and unpublished data from multiple (medical) research projects. Its use is becoming increasingly common in the medical literature, with 1000 analyses found on MedLine between 1987 and 1993.¹ "Meta" is used in the sense of meaning more comprehensive.

Proponents of meta-analysis feel the technique lends more objectivity to a systematic review, and some argue that it should completely replace the typical review article; however, there are sceptics. In a recent *BMJ*

article Professor H J Eysenck concluded that "If a treatment has an effect so recondite and obscure as to require meta-analysis to establish it I would not be happy to have it used on me."²

In emergency medicine and critical care, large multicentre trials are difficult to perform, but there is an increasing number of small randomised controlled trials which may not individually have the power to detect important treatment benefits. Emergency physicians may therefore feel that meta-analysis is a useful technique, as combining data from many small randomised controlled trials increases the statistical power to detect clinically significant treatment effects. Meta-analyses are also used by the Cochrane Collaboration and others engaged in performing systematic reviews of evidence in clinical medicine.

Consequently, it is important that emergency physicians are able to judge whether or not a meta-analysis result is providing them with a sensible answer to a therapeutic dilemma. Therefore in this paper we aim to show how meta-analysis can be used or misused by investigators attempting to clarify treatment issues in critical care.

Methods

The subject for study is a critical care meta-analysis of selective decontamination of the digestive tract (SDD). This means that enteral and sometimes parenteral antibiotics are given prophylactically to intensive care (ITU) patients in order to prevent oropharyngeal and gastrointestinal microbial carriage or to eradicate early colonisation of the lower airways, skin, and urinary tract by both community and hospital acquired micro-organisms. Therefore one of the hypotheses tested by the meta-analysis was that infection control with SDD would reduce ITU patient mortality. The authors combined data from 23 randomised controlled trials with a range of between 14 and 200 patients in the treatment arm. The data and results of the meta-analysis were published three years ago in the *British Medical Journal*.³

The result of the meta-analysis of the effect of SDD on mortality is shown in fig 1 (figure 3 in the original paper).³ The odds ratios from each of the 23 trials were combined using the Mantel-Haenzel-Peto method to give the resulting white diamond labelled "typical odds ratio". In these trials the odds ratio means the odds of dying in the active treatment group divided by the odds of dying in placebo group.

University of Manchester,
Oxford Road,
Manchester,
United Kingdom:
North Western Injury
Research Centre
F E Lecky
R A Little

ARC Epidemiology
Unit
P Brennan

Correspondence to:
Dr Fiona Lecky, North
Western Injury Research
Centre, Stopford Building,
University of Manchester,
Oxford Road, Manchester
M13 9PT

Accepted for publication
12 July 1996

Figure 1 Meta-analysis of 23 trials showing overall effect of selective decontamination of the digestive tract (SDD) on mortality.³

Therefore an odds ratio of 1 means the odds of dying in the treatment and placebo arms are equal, an odds ratio of less than 1 suggests SDD treatment reduces the odds of dying compared to placebo. Conversely an odds ratio of greater than 1 suggests that SDD increases the odds of dying compared to placebo. To be 95% certain that treatment is either significantly better or worse than placebo the confidence interval of the odds ratio should not cross 1. This is a typical example of how a meta-analysis, performed to determine the effect of a treatment on mortality from a condition, is commonly presented.

Figure 1 shows that although the odds ratio result (black square) was less than or greater than 1 in many of the trials, the confidence interval (horizontal line) for every trial result crossed 1. No individual trial could therefore confirm that SDD reduced or increased the odds of dying when compared with placebo. The reason for this may be none of the individual trials was big enough to detect a small but important difference between the

effects of SDD and placebo on the odds of dying in ITU patients.

The white diamond in fig 1 is the meta-analysis result or summary statistic. It has narrow confidence intervals due to the large number of patients, but the confidence intervals still cross 1. Therefore there is still apparently no certain difference between the effects of SDD and placebo on the odds of ITU patients dying. The result does not appear to support the hypothesis that SDD will reduce mortality in this patient group.

Further meta-analyses were performed on certain subgroups of the 23 trials. These are shown in fig 2.³ None shows any definite difference between SDD and placebo apart from the top subgroup which represents the 14 trials where SDD meant treatment with parenteral as well as topical (enteral) antibiotics. In this subgroup the odds ratio and confidence intervals were less than 1, suggesting a definite treatment benefit.

The authors in their discussion acknowledged the clinical heterogeneity of the patient groups which they had combined in that there were medical, surgical, and trauma patients, and some were more severely ill than others.³ There was a calculation of how many needed to be treated to save one life in the topical and systemic subgroup, but no cost-benefit analyses and no discussion of how treatment benefit may vary between the more and less severely ill patients. It will be demonstrated why this was an important omission.

The authors concluded that their meta-analysis offered new insights into the effect of SDD on mortality, that the total effect was still uncertain, but that further trials may not be necessary as they were planning to perform an individual patient meta-analysis.³ This involves taking certain patients belonging to a given subgroup from some or each of the individual trials and calculating a new odds ratio for that group.

The objectives of the present paper were to review the SDD data and meta-analysis result, then to use statistical techniques to demonstrate its clinical significance. The hypothesis for study is that the SDD meta-analysis result for effect on mortality has little clinical significance as it obscures important differences in treatment effect for subgroups of patients. The methods used to demonstrate this were a tabular illustration of the heterogeneity of patients included in the 23 trials, a weighted regression analysis using patient severity of illness or risk of death to predict treatment benefit, and stratification to pool odds ratios from groups of trials with higher and lower control group mortality rates.

Results

The table, taken from the data given in the SDD paper,³ gives some indication of how different the patients were in each of the 23 trials. This heterogeneity is evident in the columns showing percentages of patients whose primary problem was medical, surgical, or traumatic, and more importantly in the control group mortality rate/risk of death, which is a marker

Figure 2 Meta-analysis of subgroups showing effect of selective decontamination of the digestive tract (SDD) on mortality.⁵

of the severity of patient illness in each of the trials. This ranges from 18% to 58%. The odds ratios suggest that there may be a trend, in that the lower odds ratios (indicating greater treatment benefit from SDD) tend to occur at the top of the table where the trials include more severely ill patients (higher control group risk of death).

A weighted regression analysis of the patient risk of death (control group mortality rate) against the effect of treatment (natural log of the odds ratio) from each trial confirms the trend suggested by the table. The equation printed at the top of fig 3 shows that patient risk of death/severity of illness (control group mortality rate) is a significant predictor of the degree of SDD treatment benefit, $P = 0.017$, $R^2 = 20.5\%$. The regression has a negative slope with the log

Figure 3 Effect of treatment in relation to risk of death on ITU.

odds ratio decreasing (increasing treatment benefit) as risk of death increases.

The regression analysis was performed using the Stata statistical software package,⁹ which weighted each trial's effect on the regression by the inverse of the standard error of its logged odds ratio result.

It can be argued that a negative association could occur by chance between control group mortality rate and the logged odds ratio. However, this is unlikely to explain the strength of association found between them in this instance, as a similar result is obtained by performing the same analysis using overall trial mortality rates rather than just those in the control group.

Stratifying the trials into more and less severely ill patients and pooling the odds ratios also shows this effect (fig 4). The pooled odds ratio for the bottom 15 (two thirds of) trials where patients were less ill (less risk of death, control mortality rate CMR <37%) was 1.02 with a confidence interval extending from 0.86 to 1.21. As the confidence interval crosses 1 this suggests that for this group of patients treatment with SDD did not reduce the odds of dying when compared with placebo. The pooled odds ratio of the eight trials with more severely ill patients (higher risk of death/CMR

Heterogeneity of patients between SDD trials by disease, severity of illness (control group mortality rate) and treatment benefit (odds ratio)

Trial	Control mortality %	Odds ratio	% Trauma patients	% Medical patients	% Surgical patients
Ulrich	58	0.49	14	30	56
Rocha	52	0.54	79	21	0
Jacobs	50	0.46	18	25	57
Gaussorges	49	1.0	0	84	16
Sanchez-Garcia	46	0.74	17	71	12
Winter	43	0.74	12	39	49
Cerra	43	1.4	4	0	96
Rodrigz Rodan	41	0.8	48	36	16
Gastinn	36	1.16	15	72	13
Kerver	32	0.85	38	17	60
Unertl	30	0.84	33	52	15
Palomar	29	1.03	45	38	17
Ferrer	28	1.07	23	64	14
Pugin	27	0.97	52	36	12
Brun Buisson	22	0.97	2	79	19
Cock'rl	21	0.63	15	17	48
Aerdt	20	0.68	34	41	25
Hammond	19	1.1	27	59	14
Blair	19	0.76	40	15	45
Verhaag 1	18	1.15	24	10	66
Verhaag 2	18	1.22	22	10	68
Godard	18	0.62	41	31	28
Korinek	18	1.36	50	0	50

Figure 4 Effect of selective decontamination of the digestive tract on ITU deaths stratified by control group mortality rate (all trials).

Figure 5 Effect of selective decontamination of the digestive tract on ITU deaths stratified by control group mortality rate (topical and systemic trials).

Figure 6 Effect of cholesterol lowering treatment on total mortality stratified by control group mortality rate.⁴

> 41%) was 0.69. This falls outside the confidence interval of the odds ratio from the less ill group, suggesting a significant difference between the two groups. Also the confidence interval for this more severely ill group ranged from 0.54 to 0.89, suggesting that treatment with SDD in this group confers real benefit compared to placebo. This difference in effect was apparently not detected by the authors of the original analysis.

The same basis for stratification was also used in the 14 trials where topical and systemic antibiotics were used to decontaminate the gut. The original paper shows an overall benefit for this group as the confidence interval of the odds ratio is less than 1 (OR = 0.8, confidence

interval 0.67 to 0.97).³ However, the analysis shown in fig 5 emphasises the marked difference in effect for those with a greater and lesser risk of death. The less severely ill group of nine trials with a lower risk of death (CMR < 31%) again appears to derive no benefit from SDD treatment, the OR being 1.04 (confidence interval 0.81 to 1.25). However, those ITU patients with a greater risk of death in the upper five (one third of) trials derive significant benefit, with OR = 0.62 (confidence interval 0.47 to 0.82). As 0.62 is outside the confidence interval of the odds ratio from the group of trials with a lower risk of death, this suggests that combined topical and systemic SDD also has significantly different treatment effects for these two groups.

Figures 4 and 5 are not altered if the overall mortality rates from the SDD trials are used instead of the control group mortality rates as a basis for stratification. This is because the same trials appear in the upper third of both control and overall mortality rates. This makes it unlikely that patients are more severely ill by chance in the control arms of trials where SDD confers the greatest benefit.

Discussion

Our analysis shows how treatment effects can be significantly different for more and less severely ill groups of patients. This means that a single summary statistic, such as that given by the SDD authors in fig 1 (which combines effects across groups) may therefore not be helpful to the clinician.

OTHER META-ANALYSES

This difference in treatment effects for differing risk of death from the target disorder (condition being treated) has been demonstrated elsewhere. A meta-analysis of randomised controlled trials of the effect of cholesterol lowering treatments on mortality from coronary heart disease (CHD) indicates a trend for the odds ratio to be lower (treatments being most effective in reducing odds of CHD death when compared with placebo) for the trials where the patients were most at risk of CHD death (fig 6, from a paper by Davey Smith *et al*).⁴

Rather than providing a summary statistic, the authors explored the possibility that there were differing treatment effects for different severities of CHD, and showed that those with the greatest risk of death from CHD benefited from the intervention, whereas those with low risk do not benefit and may even be harmed. The authors presented the data from the cholesterol lowering trials through a simple regression and stratification for severity of illness (fig 6) in a pattern reminiscent of the earlier review of SDD trials in figs 3-5.

Even when mortality from non-CHD causes was taken into account, the high risk group was still found to have a significantly better odds ratio for cholesterol reducing overall mortality than the low risk group. This difference was present whether the trials were of drug or non-drug therapy.⁴

SDD ENDPOINTS

It is clear that in some aspects the SDD meta-analysis is not comparable to that performed in the cholesterol paper. For example the SDD mortality data use any death as an outcome measure whether or not it was a death from infection. There is no biological plausibility for SDD reducing ITU deaths that are not caused by infection; however, the authors point out the difficulty of determining post mortem whether infection was a primary cause of death.³ The SDD outcome measure is therefore less clearly defined than in the cholesterol paper, where the authors were able to comment on the effect of treatment on CHD and non-CHD deaths.

In the SDD paper the authors did look at the effect of treatment on respiratory tract infection rates, and they found a highly significant reduction: OR 0.37 (0.31 to 0.43).³ However, as demonstrated earlier, this did not translate into a reduction in mortality except for the more severely ill group. Even then we are not sure if it was a reduction in infection related deaths. It may be that in the less severely ill group the benefits of reducing respiratory tract infections were cancelled by antibiotic side effects.

RECOGNISING UNACCEPTABLE CLINICAL HETEROGENEITY IN A META-ANALYSIS SUMMARY STATISTIC

The quandary, then, for a clinician reading a meta-analysis which claims that its summary statistic is the answer to a therapeutic dilemma, rests on two decisions. Firstly whether or not the statistic is meaningful, in that it appears to apply to a defined group of patients which can be recognised in clinical practice; and secondly whether it is reasonable to give an average effect statistic given the degree of patient and method heterogeneity present across the various trials. Most clinicians will not have time to perform the kind of analysis done in this paper but there are other indicators.

One clue is often given in the Methods or Results section of the paper. This is a χ^2 statistic, which tests for statistical heterogeneity in the results of the trials included in the meta-analysis. If significant it implies that there is more statistical heterogeneity in the trial results than would be expected by chance, and therefore that there may be important differences between the trials included, in terms of methods used or patients included. A significant χ^2 statistic should lead the reader to regard the published summary statistic with scepticism in terms of clinical significance. However, some meta-analyses do not publish this statistic, and even if published and not "significant", this does not mean that all the trials included were carried out in the same way or on the same patient group.⁵ In the SDD paper the χ^2 statistic was calculated but not published.

In this situation there is no substitute for carefully comparing the patient inclusion criteria, the methods, and the control group mortality rates (an indicator of severity of illness) of the different trials and making a judgement as to whether it is sensible to group the trials, despite whatever differences are found between them. For example,

in the SDD paper the inclusion criteria ranged from "all ITU patients" to only those patients expected to be ventilated for more than five days,³ and hence the great variation in control group mortality rates. Without performing more sophisticated analysis a clinician should immediately be very sceptical that a summary statistic of treatment effect could apply equally to these patient groups or indeed reflect true treatment effect in any of the different subgroups.

QUALITIES OF A GOOD META-ANALYSIS

Guidelines exist in the emergency medicine literature for those wanting to perform a metaanalysis.¹ Various investigators have suggested some markers of a good meta-analysis. As the present paper has demonstrated, it is important to define the patient group and the type of trial to be considered before making a search for relevant studies.¹ It is worth determining whether or not the authors of the meta-analysis have searched for unpublished data to avoid publication bias (negative findings are less likely to be published), and there are statistical tests for indicating this.⁶ A recent *BMJ* review of meta-analyses of magnesium and streptokinase treatments in acute myocardial infarction that predated large multicentre randomised controlled trials indicated that meta-analysis results are more likely to be misleading if no medium size trials (500-1000 patients) are included.⁶ This highlights another point which is that when meta-analysis attempts to determine a treatment effect using large patient numbers from many small trials it is often a poor substitute for a large randomised controlled trial,⁷ though of course it is much easier and cheaper to perform.

The most important point is perhaps a philosophical one. When performing a meta-analysis there is an inherent tension between achieving objectivity and larger numbers by using all the available evidence, and being apparently subjective and rejecting poor quality trials or those on widely different patient groups. However, it has been suggested that even in meta-analysis, as in other aspects of medical research, such judgements are necessary in order to provide good quality evidence to clinicians⁵ and to prevent meta-analysis falling further into disrepute as "an exercise in mindless agglomeration of data."⁸

CONCLUSIONS

The results of this analysis support the hypothesis that the original summary statistic for the effect of SDD on mortality has little clinical relevance because it obscures important differences in treatment effect for more and less severely ill patient groups.

The discussion has indicated that when used appropriately meta-analysis can show how treatment benefit varies for different patient groups and that net mortality reductions tend to increase with increasing severity of illness or risk of death from the target condition or disease.

A meta-analysis result is not a magic number though it does have a seductive numeracy which clinicians should be wary of. Before bas-

ing clinical practice on a meta-analysis result it is important to read beyond the statistic and make sure it is not the result of combining incomparable trials.

We would like to acknowledge the support of the Wellcome Trust who have funded Dr Lecky's research fellowship, and Mrs C Cooke for secretarial support. Permission was kindly given by the *BMJ* to reproduce the figs 1, 2, and 6, which were previously published in that journal, by the SDD trialists for figs 1 and 2, and George Davey Smith *et al* for fig 6.

- 1 Olson CM. Understanding and evaluating a meta-analysis. *Acad Emerg Med* 1994;1:392-8.

- 2 Eysenck HJ. Meta-analysis and its problems. *BMJ* 1994; 309:789-92.
- 3 SDD Trialist's collaborative group. Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract. *BMJ* 1993;307:525-31.
- 4 Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *BMJ* 1993;306:1367-73.
- 5 Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994;309:1351-5.
- 6 Egger M, Davey Smith G. Misleading meta-analysis. *BMJ* 1995;310:752-4.
- 7 Yusuf S, Flather M. Magnesium in acute myocardial infarction. *BMJ* 1995;310:751-2.
- 8 Shapiro S. Meta-analysis/shmeta-analysis. *Am J Epidemiol* 1994;140:771-8.
- 9 StataCorp Stata Statistical Software. Release 4.0 TX: Stata Corporation, 1995.

CONSULTANT APPOINTMENTS APRIL-SEPTEMBER 1996

Name	Post	Hospital
Mr H B Aslam	Consultant in A & E	Princess Margaret Hospital, Swindon
Mr A Bleetman	Consultant in A & E	Birmingham Heartlands Hospital
Dr P Cutting	Consultant in A & E	Gwynedd Hospital
Miss Denning	Consultant in A & E	Winchester and Eastleigh Hospital
Mr D W Hodgkinson	Consultant in A & E	Ipswich Hospital
Mr Ieuan Jones	Consultant in A & E	Gwynedd Hospital
Mr N Kumar	Consultant in A & E	Kidderminster General Hospital
Dr W Kurivan	Consultant in A & E	Edgware Hospital
Mr M Mahmoud	Consultant in A & E	Canderdale Healthcare, Halifax
Mr A Mariathas	Consultant in A & E	Broomfield Hospital, Chelmsford
Mr C Merrill	Consultant in A & E	Hillingdon Hospital
Dr S Mullett	Consultant in A & E	Queen Alexander Hospital, Cosham
Mr Brian McNicholl	Consultant in A & E	Royal Groups of Hospital, Belfast
Dr P Reid	Consultant in A & E	Lister Hospital, Stevenage
Mr H Rokan	Consultant in A & E	Southend Hospital
Mr J Ryan	Senior Lecturer in A & E	Brighton General Hospital
Dr B Shravat	Consultant in A & E	Edgware Hospital
Mr M Zahir	Consultant in A & E	Coventry and Warwickshire Hospital