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 × 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.


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Meta-analysis is a statistical technique for combining the results of published and unpublished data from multiple (medical) research projects.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).4 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.
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. 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. 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. 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, 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
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

<table>
<thead>
<tr>
<th>Trial</th>
<th>Control mortality %</th>
<th>Odds ratio</th>
<th>% Trauma patients</th>
<th>% Medical patients</th>
<th>% Surgical patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ulrich</td>
<td>58</td>
<td>0.49</td>
<td>14</td>
<td>30</td>
<td>56</td>
</tr>
<tr>
<td>Rocha</td>
<td>52</td>
<td>0.54</td>
<td>79</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>Jacobs</td>
<td>50</td>
<td>0.46</td>
<td>18</td>
<td>25</td>
<td>57</td>
</tr>
<tr>
<td>Gaulsorts</td>
<td>49</td>
<td>1.0</td>
<td>0</td>
<td>84</td>
<td>16</td>
</tr>
<tr>
<td>Sanchez-Garcia</td>
<td>46</td>
<td>0.74</td>
<td>17</td>
<td>71</td>
<td>12</td>
</tr>
<tr>
<td>Winter</td>
<td>43</td>
<td>0.74</td>
<td>12</td>
<td>39</td>
<td>49</td>
</tr>
<tr>
<td>Cerra</td>
<td>43</td>
<td>1.4</td>
<td>4</td>
<td>0</td>
<td>96</td>
</tr>
<tr>
<td>Rodriguez Rodan</td>
<td>41</td>
<td>0.48</td>
<td>48</td>
<td>36</td>
<td>16</td>
</tr>
<tr>
<td>Gastmann</td>
<td>36</td>
<td>1.16</td>
<td>15</td>
<td>72</td>
<td>13</td>
</tr>
<tr>
<td>Kerver</td>
<td>32</td>
<td>0.85</td>
<td>38</td>
<td>17</td>
<td>60</td>
</tr>
<tr>
<td>Unerdt</td>
<td>30</td>
<td>0.84</td>
<td>33</td>
<td>52</td>
<td>15</td>
</tr>
<tr>
<td>Palomar</td>
<td>29</td>
<td>1.03</td>
<td>45</td>
<td>38</td>
<td>17</td>
</tr>
<tr>
<td>Ferrer</td>
<td>28</td>
<td>1.07</td>
<td>23</td>
<td>64</td>
<td>14</td>
</tr>
<tr>
<td>Pugin</td>
<td>27</td>
<td>0.97</td>
<td>52</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>Brun Buison</td>
<td>22</td>
<td>0.97</td>
<td>92</td>
<td>79</td>
<td>19</td>
</tr>
<tr>
<td>Cock'n</td>
<td>21</td>
<td>0.63</td>
<td>15</td>
<td>17</td>
<td>48</td>
</tr>
<tr>
<td>Aerdts</td>
<td>20</td>
<td>0.68</td>
<td>34</td>
<td>41</td>
<td>25</td>
</tr>
<tr>
<td>Hammond</td>
<td>19</td>
<td>1.1</td>
<td>27</td>
<td>59</td>
<td>14</td>
</tr>
<tr>
<td>Blair</td>
<td>19</td>
<td>0.76</td>
<td>40</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>Verhaag 1</td>
<td>18</td>
<td>1.15</td>
<td>24</td>
<td>10</td>
<td>66</td>
</tr>
<tr>
<td>Verhaag 2</td>
<td>18</td>
<td>1.22</td>
<td>22</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td>Godard</td>
<td>18</td>
<td>0.62</td>
<td>41</td>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td>Korinek</td>
<td>18</td>
<td>1.36</td>
<td>50</td>
<td>0</td>
<td>50</td>
</tr>
</tbody>
</table>
The effect of control 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).

The effect of cholesterol lowering treatment on total mortality stratified by control group mortality rate.

> 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). 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.5).

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-
alysis 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.7 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).2 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 can-
celled 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 vari-
ous 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 $\chi^2$
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 differ-
ences between the trials included, in terms of
methods used or patients included. A signif-
ificant $\chi^2$ statistic should lead the reader to regard
the published summary statistic with scepti-
cism 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 $\chi^2$
statistic was calculated but not published.

In this situation there is no substitute for care-
fully 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 dif-
ferences 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 so-
phisticated analysis a clinician should immedi-
ately 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 lit-
erature for those wanting to perform a
metaanalysis.1 Various investigators have sug-
gested 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 find-
ings are less likely to be published), and there
are statistical tests for indicating this.4 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 mis-
leading if no medium size trials (500-1000
patients) are included.4 This highlights another
point which is that when meta-analysis at-
tempts to determine a treatment effect using
large patient numbers from many small trials it
is often a poor substitute for a large ran-
domised controlled trial,2 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 qual-
ity 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 neces-
sary in order to provide good quality evidence
to clinicians5 and to prevent meta-analysis fall-
ing further into disrepute as "an exercise in mindless agglomeration of data."

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

The discussion has indicated that when used
appropriately meta-analysis can show how treat-
ment 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.


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