*Indian J. Psychiat., 1998, 40 (2), 135-148*

*REVIEW ARTICLE*

# THE RELEVANCE OF META-ANALYSIS, SYSTEMATIC REVIEWS AND THE COCHRANE COLLABORATION TO CLINICAL PSYCHIATRY

**PRATHAP THARYAN**

## *ABSTRACT*

*Traditional review articles provide clinicians with syntheses of the medical literature but are criticised as being haphazard in their methodology and biased in their conclusions. Systematic reviews use rigorous methods to minimise bias and statistical methods to synthesise results (meta-analysis) that increase power and precision. They permit investigation of generalisability and consistency, improve transparency of methodology, and enhance reproducibility. This article examines the science of systematic reviews and meta-analysis and their relevance to clinical psychiatry. It evaluates the potential errors and sources of bias of meta-analysis, and offers guidelines for evaluation of systematic reviews. It highlights the efforts of the Cochrane Collaboration which is an international organisation involved in preparing, maintaining and disseminating highly structured, frequently updated, and good quality systematic reviews of the effects of interventions in all aspects of health care.*

*Key words : Meta-analysis, systematic reviews, epidemiological methods, bias, research design.*

The relevance of meta-analysis to psychiatry stems from one of the earliest meta-analysis ever undertaken, that evaluated the efficacy of various forms of psychotherapy (Smith & Glass, 1977). Since the 1980's, meta-analysis have increasingly appeared in the medical literature, and scarcely a month now passes without the publication of a meta-analysis of relevance to clinical psychiatry in general medical journals or in mainstream psychiatric literature.

The term 'meta-analysis', or the 'analysis of analyses', was coined by Glass (1976). The term meta-analysis is slowly being replaced by the term 'overviews' or, more recently, 'systematic reviews'. The Potsdam International Consultation on Meta-analysis in 1994 provided the following definitions of terminology: Systematic Review (or Overview): the application of scientific strategies that limit bias to the systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific topic. Meta-analysis (or Quantitative Overview): a systematic review that employs statistical methods to combine and summarise the results of several studies (Cook et al., 1995).

## Why do we need systematic reviews ?

Health care providers, researchers and policy makers are inundated with unmanageable amounts of information and need some method of summarising this for efficient decision making. The traditional review article has provided one way of synthesising the relevant articles into digestible amounts of information.

Often written by experts who are opinion leaders in their fields, these traditional or subjective reviews have been influential in directing research and clinical decision making.

1) Superiority over traditional reviews:

Traditional reviews have, however, been criticised as being haphazard and biased, subject to the idiosyncratic impressions of the individual reviewer (Mulrow, 1987), and rarely as scientific in methodology as the studies they review (Haynes, 1992). An attempt to answer a clinical question on the efficacy of Electroconvulsive Therapy (ECT) in schizophrenia by consulting traditional reviews in journals, specialist and general psychiatric text -books, disconcertingly revealed that there was little agreement on studies selected for review, as well as in crucial aspects of management (Tharyan, 1997).

On the other hand, systematic reviews and meta-analysis apply explicit scientific principles aimed at reducing random and systematic errors (Oxman et al., 1988). This permits replication and enables one to understand why results and conclusions of some reviews differ.

Traditional reviews lag behind and often vary significantly from continuously updated or cumulative meta-analysis. Cumulative meta-analysis refers to the performance of a meta-analysis every time a new trial appears, providing practitioners with up-to-date information on emerging and established advances. Lau et al. (1992) demonstrated the advantage of cumulative meta-analysis when they evaluated 33 trials done on the use of intravenous streptokinase as thrombolytic therapy for myocardial infarction conducted between 1959 and 1988. They found that a consistent, statistically significant reduction in total mortality (odds ratio, 0.74: 95% CI 0.59-0.92) was achieved in 1973, after only 8 trials involving 2432 patients had been completed. The results of the 25 subsequent trials involving 34,542 patients till 1989 (including two very large, expensive trials involving 11,712 patients and 17,187 patients) had little or no effect on the odds ratio establishing efficacy, but simply narrowed down the confidence interval. This suggests that intravenous streptokinase could have been shown to be life saving almost 20 years before it's general adoption to clinical practice.

Similarly, Antman et al. (1992) using cumulative meta-analysis showed that pooled data from 15 randomised trials published before 1990 found no evidence of mortality reduction associated with prophylactic lidocaine for acute myocardial infarction. Despite this evidence, most traditional reviews and expert recommendations in text-books on the topic at that time continued to recommend prophylactic lidocaine and many effective treatments were not being recommended. Cumulative meta-analysis has the potential to provide information to reduce the need for future large trials (Lau et al., 1995)

2. Increased power:

Even a well-conducted randomised control trial may have insufficient power to demonstrate an effect convincingly, unless the effect of the intervention is very strong or the trial is very large. In general, the real difference between two treatments, in many important outcomes, are often not large, but may be clinically important. Many studies with negative results do not have sample sizes large enough to even detect a 50% difference between treatments (Sackett, 1979). By combining results of several trials, meta-analysis provide a 'tower of power' that increases the sample size and thus the power to detect significant differences (Mulrow, 1994). This is readily apparent in the Cochrane Collaboration's logo (Fig. 1). This graphically depicts, as Odds Ratios (OR) and Confidence Intervals (CI), the results of 7 trials that evaluated the effects of a short course of corticosteroids given to women expected to give birth prematurely, with the outcome depicted being neonatal death. In

graphical displays of meta-analysis, sometimes called 'forest plots', if the odds ratios for negative outcomes lie to the left of the midline (OR < 1) the results favour the experimental treatment, while the reverse (OR > 1) favour the control condition. The midline represents the point of no significant difference (OR = 1). In Figure 1, only two trials had clear-cut statistically significant results, graphically depicted as the two trials whose confidence intervals lie entirely to the left of the midline. The other five trial results were not statistically significant as is evident from the confidence intervals straddling the midline. Using the traditional 'vote counting approach', the conclusion drawn would be that there is no evidence to indicate that corticosteroids reduce neonatal mortality, since the number of inconclusive studies outnumber the positive studies. However, when data from all of the studies were pooled, this yielded a definitive significant combined effects estimate (represented by the diamond to the left of the midline demonstrating a positive treatment effect), which indicated strongly that corticosteroids reduce neonatal mortality from complications of immaturity.

FIGURE 1: THE COCHRANE COLLABORATION LOGO

The diamond at the bottom represents pooled odds ratio. Pooling data provides 'power' and reveals significant effects that individual small studies are unable to do.

3. Increased precision:

Yusuf et al. (1985) assessed 24 RCTs that evaluated mortality reduction by the long-term use of beta-blockers after myocardial infarction. Four trials showed results that were in favour of beta-blockers, while twenty trials found no significant benefit, with 14 of them showing an unfavourable result. Combination of the results of the 24 trials based on about 20,000 subjects yielded a pooled odds ratio of 0.77 (95% CI = 0.70-0.85; note the narrow confidence intervals); i.e. mortality was 23% (CI between 15 -30 %) lower in treated patients than in controls. This demonstrates the greater precision with meta-analysis compared to the 'vote co nting' approach.

This is largely due to the differential weighting of studies and evaluation of weighted pooled data in meta-analysis, while with the 'vote counting' approach, equal weights are assigned to all studies irrespective of quality, sample size, magnitude of effect, or precision of results, with inappropriate over-reliance on significant 'p values'.

It is important to remember that results that achieve conventional levels of significance ($p < 0.05$) merely tell us that the results are less likely to be accounted for by chance. The use of odds ratios and confidence intervals (a measure of statistical uncertainty), enables estimation of the size of the effect of intervention as well as the range within which the true effect may plausibly lie (Gardner & Altman, 1986).

4. Generalisability and consistency:

Since different studies use different populations, designs, definitions, measurements and variations in treatment, systematic reviews provide an opportunity to assess the generalisability of findings that is not possible with individual studies.

Consistency of results of whether effects are in the same direction, and of similar magnitudes, can be made and reasons for inconsistency can be explored.

## PITFALLS OF META-ANALYSIS

Meta-analysis have been likened to 'an exercise in mega silliness', with the major contention that including all material - good, bad, and indifferent- in a meta-analysis delegates to

the reader the subjective judgements that meta-analysis was designed to avoid. Over the past decade critiques of meta-analysis have increasingly appeared in medical literature (Goodman, 1991; Shapiro, 1994; Egger & Smith, 1995; Feinstein, 1995; Bailar, 1995 & 1997; Naylor, 1997). These articles raise a number of questions.

1) Can meta-analysis be trusted?

There are many problems in conducting meta-analysis and erroneous results are sometimes produced. A meta-analysis of 7 RCTs provided evidence for the effectiveness of intravenous magnesium, in the treatment of patients with suspected myocardial infarction (Teo et al., 1991).

However, the results of a mega trial, the ISIS 4 trial, have called into question the results of the meta-analysis and the effectiveness of magnesium (The ISIS-4 Collaborative Group, 1995).

The predictive ability of meta-analysis of RCTs have been questioned by Villar et al. (1995). More recently, important discrepancies, mainly in relation to the size of the effect rather than the direction, were detected between published meta-analysis and subsequent large randomised controlled trials on the same topic (LeLorier et al., 1997). Adherence to guidelines for systematic reviews for RCTs (Cook et al., 1995) could reduce such errors, though possibly not completely eliminate them. It therefore behoves readers of systematic reviews to be aware of potential sources of errors and bias.

2) What are the sources of bias in systematic reviews and meta-analysis?

Systematic reviews and meta-analysis are retrospective observational studies intermediate in design between primary research and traditional 'narrative' reviews. The units of observation are the results obtained from other studies, hence the reviewer has many constraints that can lead to biases and subsequent errors. The following are common sources of bias (Felson, 1992; Mulrow & Oxman, 1994; Bailar, 1997; Naylor, 1997).

A) Sampling bias:

One of the challenges, and an important aspect of the validity, of a systematic review is the complete identification of all relevant studies. Biases in sampling could occur due to any of the following:

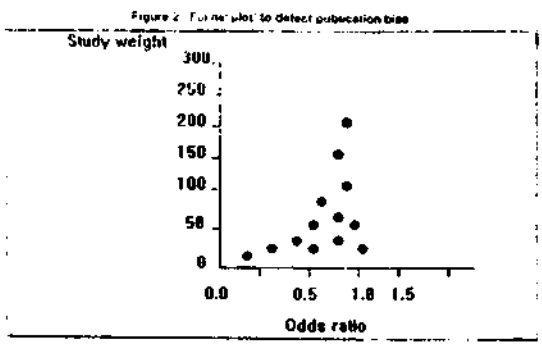
i) Publication bias. This refers to the tendency for studies that report statistically significant results to be published. Unpublished studies are likely to have negative results as well as smaller sample sizes. Reliance only on the results of published trials produces over-estimates of treatment effects and leads to erroneous conclusions that could prove detrimental to patient care (Egger & Smith, 1995).

Publication bias often arises due to authors of negative trials not submitting manuscripts for publication (Rosenthal, 1979), editorial and referees' policy in not publishing studies with results that challenge conventional beliefs (Dickersin, 1990), and conflicts of interest caused by drug companies discouraging publication of sponsored trials with negative or adverse results (Dickersin, 1990). Non-English-language references are under-represented in electronic databases such as MEDLINE and only published articles are included, so depending only on MEDLINE increases the potential for publication bias (Dickersin et al., 1987), and language bias (Easterbrook et al., 1991). To protect against bias and ensure that all relevant data are included in a review it is important to use multiple sources to identify relevant trials (Dickersin et al., 1994).

The importance of publication bias has been recently recognised by editors of over a 100 journals around the world who have called for authors of unpublished controlled trials (including those trials published only as abstracts) to submit them for registration. The aim is to increase accessibility of these trials to those wishing to do systematic reviews (Smith &

Roberts, 1997).

There are some ways to assess the presence and magnitude of publication bias in systematic reviews, though none of them are fully validated. One method is the computation of the "fail-safe number" which is the number of unpublished studies that would be needed to change the results of a meta-analysis (Rosenthal, 1979). If the 'fail-safe number' is very large, then publication bias is unlikely to have an effect on the results of the meta-analysis. The other is the use of the "funnel plot" (Mulrow & Oxman, 1994). This is a simple graphical test where the effect size or the odds ratios of the studies is plotted against the sample size or the study weight (usually the inverse of the variance of the odds ratio). If there is no publication bias the plot would resemble an inverted funnel with larger studies (with greater weights) at the top in the middle and smaller studies spread at the bottom. If one assumes that odds ratios greater than one favours the control condition, a gap in the bottom right of the funnel (Figure 2) indicates that negative studies with small sample sizes were not identified (the most common scenario). A gap on the bottom left side indicates that smaller positive studies were missed.




The gap in the bottom right indicates that negative studies with small sample sizes were missed (see text for details). Source: Mulrow & Oxman (1994)

The ultimate way to avoid publication bias is to build a comprehensive register of published and unpublished trials preferably for each speciality (Adams & Gelder, 1994), and to prospectively register trials at the stage after research & ethics committee approval is obtained (Naylor, 1997).

ii) Retrieval bias. Electronic databases such as MEDLINE (the electronic form of Index Medicus) and EMBASE (the electronic form of excerpta medica) are powerful tools for locating studies.

However, only 30 - 80% of all known published randomised controlled trials are identifiable using MEDLINE, depending on the area or specific question (Dickersin et al., 1994). In the field of mental health, MEDLINE searches failed to identify 30-50% of RCTs while hand-searching journals identified 95% of the trials (Adams et al., 1994). Many journals are not indexed in MEDLINE and many articles in indexed journals do not provide sufficient information to permit proper indexing. Moreover, the retrieval rate increases with the expertise of the retriever and the casual searcher is unlikely to retrieve a majority of indexed trials (Dickersin et al., 1994).

It is often necessary to search more than one electronic database, as the overlap in journals covered by MEDLINE and EMBASE is only approximately 34% (ranging from 10% to 75%) (Mulrow & Oxman, 1994). Other strategies to improve comprehensiveness of the search for trials are checking cross-references of retrieved trials and consulting references in existing 'narrative' reviews though one needs to guard against 'reference bias' (Gotzsche, 1987) which is the tendency to selectively quote references that agree with the authors' viewpoint. Personal communication with authors of trials, experts in the field, pharmaceutical companies or manufacturers of equipment (e.g. ECT machines), hand-searching journals and conference abstracts, though also subject to 'reference bias', are potentially useful to complete retrieval of relevant studies in the 'grey literature'.

iii) Multiple publication bias. Not infrequently reviewers are faced with the problem of trials being published in different journals as a series of articles, thus leading to the possibility of them

being recorded as separate trials, thereby inflating erroneously the number of participants in the meta-analysis. This is more likely if the articles are published with different authors, as was detected by Huston & Moher (1996), who recorded seven different publications with different authorship of a single trial of Risperidone in chronic schizophrenia.

B. Selection Bias

Once trials are identified they are usually selected according to inclusion criteria predetermined by the reviewers. There are various causes of selection bias: i) Inclusion criteria bias: The use of inclusion criteria can introduce bias if the reviewer consciously or otherwise sets criteria in a manner that excludes trials he/she is aware of. One of the ways of minimising selection bias is to select a comprehensive and clearly formulated protocol a priori and the manner in which the methodological quality of the studies shall be critically appraised for relevance and validity. Selection bias can be further reduced by two (or more) reviewers independently assessing trials for inclusion, and further by blinding the reviewers to the authors, institutional affiliation, as well as to the results of the trials during assessment for inclusion. A combination of a content expert (with pre-formed opinions that can bias assessments of both the relevance and validity of articles) and a non-expert (with no pre-formed and potentially biased views) may confer an advantage in minimising selection bias.

ii) Bias in assessing quality of trials : It is important in a systematic review to exclude poor quality trials as their inclusion tends to exaggerate the overall estimate of treatment effects and may lead to incorrect inferences (Khan et al., 1996). Hence some criterion of quality needs to be applied that is both simple to use as well as valid.

In assessing the quality of trials, four sources of bias are relevant: Selection bias, performance bias, attrition bias and detection bias. The first two are most likely to result in biased outcomes (Chalmers, 1983; Schultz, 1995). Selection bias can be minimised by randomisation, though not all so called RCTs are really randomised. Empirical research has shown that lack of adequate allocation concealment during randomisation is associated with bias (Chalmers, 1983). Thus trials can be judged on the reported method of allocation concealment. Performance bias refers to systematic differences in care provided to comparison groups other than the intervention of interest. Double blinding is important in protecting against performance bias. There is evidence that participants who are aware of their assignment status report more symptoms, leading to biased results (Sackett, 1979). For these reasons, use of " blinding" is often used as a criterion for validity, though not all supposedly double-blind studies are really double- blind (Oxtoby et al., 1989).

Many scales and checklists have been used to assess the validity and "quality" of randomised controlled trials (Moher et al., 1995). Many of the instruments use complicated scoring systems, with or without weights, that are time consuming, do not improve validity, are liable to confuse the quality of reporting with the validity of the design and conduct of a trial and are not supported by empirical evidence (Mulrow & Oxman, 1994). For example, there is no empirical basis for determining how much weight to assign to different validity criteria. Simple approaches to eliminating highly biased trials are preferred, though this is an area that warrants further evaluation.

A limitation of assessing quality of trials is the poor quality of reporting in many published reports, at least from the point of view of performing systematic reviews. The application of standards for reporting trials (The standards for reporting trials group 1994; Altman, 1996) can facilitate critical appraisal.

C. Data extraction bias.

The accuracy and inter-observer

variability of data extraction from trials as a potential source of bias has been apparent since the earliest meta-analysis (Smith & Glass, 1977). Inter-observer reliability can be maximised and extraction bias minimised by the use of specific and comprehensive pre-tested data-extraction sheets with clear instructions on how to interpret and handle data and by reviewers working independently duplicating the extraction process. Contributing to extraction bias may be 'reporting bias' where inadequate reporting of data for various outcomes, especially unfavourable outcomes, leads to biased results. Published data may often need to be complemented by additional information obtained by personal communication from the authors. There is evidence that both 'data extraction bias' and 'reporting bias' can be minimised by the use of individual patient data in performing meta-analysis (Stewart & Parmar, 1993), though the effort involved in obtaining this may be considerable. This approach may be more feasible if trials are prospectively registered.

D. Bias in analysis of results

The central aim of most systematic reviews is to combine the results (meta-analysis) of included trials using a summary statistic in order to provide a reliable estimate of the effects of an intervention. Typically studies are given weights which are usually the inverse of the variance; i.e. more precise estimates (from larger studies with more events) are given more weight. Each study is summarised using a measure of effect (such as an odds ratio or a relative risk for dichotomous data, using the Mantel -Haenszel technique or the Peto modification (Berlin et al., 1989; Yusuf et al., 1985), or a weighted mean difference for continuous data). This represents the within study comparison of the intervention and control groups. The pooled, weighted measures are combined to produce the summary statistic, usually with confidence intervals that are similar to confidence intervals in single studies. This ensures that participants in each study are only compared to people in the same study, thus preserving the power of randomisation.

Reviews that use absolute measures (Absolute Risk Reduction, Number Needed to Treat, Number Needed to Harm) in addition to relative measures (Relative Risk Reduction) provide useful information for clinicians and health planners. The Number Needed to Treat (NNT) is an estimate, from the meta-analysis, of the number of persons who must be treated with an intervention to result in a positive outcome or to prevent an adverse outcome (Laupacis et al., 1988). It takes into account the baseline risks as well as the magnitude of risk reduction and can be easily calculated from the pooled estimates (Table 1).

**TABLE 1**
**EFFECT MEASURES FOR DICHOTOMOUS DATA**
**(Mulrow & Oxman, 1994)**

| | | Adverse outcome | |
|---|---|---|---|
| | | Present | Absent |
| | Yes | A | B |
| | No | C | D |
| Treatment | | | |

- Odds ratio = (A/B)/(C/D)
- Relative risk (RR) = [A/(A+B)]/[C/(C+D)]
- Relative risk reduction = 1-RR
- Absolute risk reduction (ARR) = A/ (A+B)-C/(C+D)
- Number needed to treat NNT = 1/ARR

No other aspect of systematic reviewing has generated more criticism than this attempt to statistically combine and present a single estimate (Bailar, 1997).

i) Inappropriate statistical methods : In early studies, and even on occasion at present, a meta-analysis is published where an "effect size" is calculated for each study group based on the difference in outcomes before and after treatment, and groups from different studies are directly compared with each other. When this

is done, the power of randomisation is completely lost and the results are often invalid (Mulrow & Oxman, 1994).

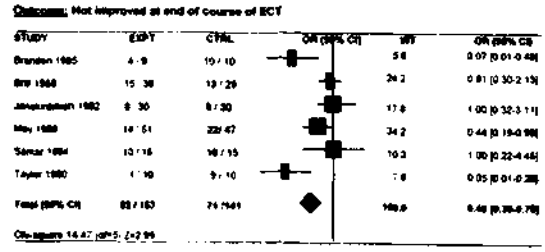
ii) Inappropriate use of statistics : Not all systematic reviews should statistically combine results. The chief reasons for not doing so are inadequate data, or if it does not make sense to combine disparate results or outcome measures, or if there is significant heterogeneity between studies that cannot be explained on methodological grounds. Inappropriate pooling of data can do more harm than good (Naylor, 1997).

iii) Dealing with heterogeneity : There are two types of heterogeneity in relation to systematic reviews. Clinical heterogeneity refers to the differences in studies due to differences in participant characteristics, interventions, and trial methodology. Statistical heterogeneity refers to incompatibility in the statistical results. The former may lead to the latter. There are two ways of evaluating for statistical heterogeneity. One way of doing this is to look at a graphical display of the results. If the confidence intervals for the results of each study do not overlap, it suggests that the differences are likely to be "statistically significant". Figure 3 provides an example of statistical heterogeneity from a comparison of ECT versus sham ECT in schizophrenia with the outcome under evaluation being numbers not improved at the end of the course of ECT (Tharyan, 1998). The pooled data suggests that ECT is superior to Sham ECT in the short term (odds ratio 0.48, confidence intervals 0.30-0.78) with 1 in 5 to 6 patients likely to benefit (NNT = 5.85). In this comparison, however, the confidence intervals of two strongly positive studies (Taylor & Fleminger, 1980; Brandon et al., 1985) barely overlapped with the results of the other studies, indicating that these two 'outliers' differed from the others significantly.

The other method of evaluating statistical heterogeneity is to do a Chi-square test for homogeneity. As a rule of thumb, the Chi-square statistic has a value equal to it's degrees of freedom (one less than the number of trials); values larger than the degrees of freedom (df) would give smaller P values and indicate significant statistical heterogeneity (Thompson, 1995). In Figure 2, the Chi-square value of 14.47 is greater than the degrees of freedom of 5 with a Z score of 2.99 yielding a p value of < 0.005 (from a standard table of cumulative normal distribution) and confirming statistically significant heterogeneity. When there is "statistically significant" heterogeneity, it suggests that the observed differences in results are likely to be caused by factors other than chance.

In the presence of heterogeneity, it is important to attempt a cautious interpretation of likely causes on methodological grounds, such as differences in dose, timing or duration of treatment, participant characteristics, control of bias, and study design. Caution is encouraged in explaining heterogeneity since this is done post hoc. A careful evaluation of study methodology of trials in the meta-analysis in Fig 3 reveals that the two strongly positive trials differed from the others in including a pre-trial stabilisation period with antipsychotics and excluding those who improved with antipsychotics. The other trials either did not use antipsychotics in either limb of the trial or did not have a significant pre-trial stabilisation period. This potentially biased the patients in the two trials against further antipsychotic improvement during the trial. Re-analyses of the data excluding these two trials reduces the strength of the evi-

FIGURE 3: ECT FOR SCHIZOPHRENIA (Tharyan 1998)




EXPT = Experimental CTRL = Control OR = Odds Ratio CI = Confidence Intervals WT = Weight df = degrees of freedom (Source: Cochrane Library 1998,Issue 1)

dence, though not the direction of effect.

Statistical methods (fixed effects versus random effects) to evaluate pooled results and confidence intervals in the face of significant heterogeneity (Berlin et al., 1989; Mulrow & Oxman, 1994) are beyond the scope of this paper.

iv) Sub-group analyses: In any systematic review it is tempting to attempt subgroup analysis of, for example, women, older people, or those with subtypes of schizophrenia. Subgroup analyses are common but can be misleading. They should not be done unless, a) there is indirect evidence suggesting that the difference, if it exists, is plausible, b) the subgroup analysis is one of a small number tested, and stated a priori, in order to prevent spurious positive results by chance, c) the difference is suggested by comparisons within studies rather than between studies, d) the difference is consistent across studies, and e) the magnitude of the difference is practically important as well as statistically significant (Oxman & Guyatt, 1992).

It is important to remember that inferences based on between-study comparisons are based on comparisons between non-comparable groups as even when all the individual studies are RCTs, patients were not randomised to one study or another (Mulrow & Oxman, 1994). The various statistical techniques that can be employed in subgroup analyses (Der Simonian & Laird, 1986) are again beyond the scope of this discussion.

iv) Sensitivity analysis: This refers to evaluations aimed at assessing how sensitive the results of the analysis are to the way the review was conducted. Changing inclusion criteria (excluding studies without operationally defined diagnostic criteria, or excluding very old trials), including or excluding studies of lower quality or with ambiguous inclusion criteria, excluding unpublished studies, reanalysing the data using different approaches to handling data (e.g.: intention to treat analysis versus analysis of completers) or different statistical techniques (random effects versus fixed effects), are commonly used decisions in sensitivity analyses.

If the sensitivity analyses that are done do not materially change the results, it strengthens the confidence that can be placed in them. If the results change in a way that might lead to different conclusions this indicates a need for greater caution in interpreting the results and drawing conclusions, or it may generate hypotheses for further investigations (Mulrow & Oxman, 1994).

E. Interpreting results

A common mistake that occurs in interpreting results when there is inconclusive evidence is to confuse 'no evidence of effect' with 'evidence of no effect'. The former reflects uncertainty while the latter indicates clear evidence exists to discredit a particular intervention. A second mistake sometimes committed is to arrive at conclusions not supported by the evidence in the review, or to make conclusions based on personal convictions inspite of evidence to the contrary presented in the review.

Table 2 provides a summary checklist to aid readers to evaluate systematic reviews for possible biases.

The science of meta-analysis is growing and attempts are constantly being made to improve scientific quality. One such endeavour is the efforts of the Cochrane Collaboration.

## WHAT IS THE COCHRANE COLLABORATION?

The Cochrane Collaboration is an international network of individuals and institutions formed in 1993, and committed to preparing, maintaining and disseminating systematic reviews of the effects of health care. Cochrane reviews are highly structured systematic reviews prepared by a group of collaborating authors, called a Cochrane review group, using explicitly refined methods to reduce bias. These reviews are published in the Cochrane Library,

**TABLE 2**
**HOW TO EVALUATE A SYSTEMATIC REVIEW**

1. Are the results of the review valid ?

- Does the review address a focused clinical question ?
- Were inclusion criteria used to select articles *appropriate* ?
- Was the search for relevant studies through ? Is there a likelihood of publication bias?
- Was the validity of included studies adequately assessed ?
- Were benefits of the intervention as well as harmful effects assessed ?
- Was the review based entirely on the results of small sample RCTs ?

2. Was the data handled appropriately ?

- Was data extraction from trials free from bias ?
- Was there an attempt to obtain missing information from authors ?
- Was an intention to treat analysis used ?
- What is the magnitude of the effect ? How precise are the overall results of the review ?
- Is it logical to combine results to produce a summary statistic ? Is there an absolute measure ?
- Is there significant heterogeneity ? If so, can it be explained on methodological or clinical grounds ?
- How sensitive are the results to changes in the way the analysis was done ?

3. Are the conclusions generalisable ?

- Are participants & interventions of included studies relevant to your practice ?
- Are the conclusions justified by the evidence ?
- Is 'no evidence of effect' interpreted as 'evidence of no effect' ?
- Are subgroup analysis interpreted cautiously ?
- Is the clinical significance of the results sufficient to warrant a change in our practice ?

an electronic journal on CD and computer discs (and now showing on the Internet) released quarterly, which contains the Cochrane Database of Systematic Reviews (CDSR). These are up-to-date reviews that are periodically updated in response to comments and criticisms and when new data becomes available. The January 1998 release contains 326 completed reviews and 342 protocols of reviews underway covering a wide range of health problems. The eventual aims are to provide a database for clinicians, policy makers and consumers on the effects of all treatments in all aspects of health care.

The Cochrane Library also contains the Data Base of Abstracts of Reviews of Effectiveness (DARE) which has at present 1554 abstracts of reviews published elsewhere but adjudged to fulfil criteria for systematic reviews. Perhaps the most innovative component of the Library is the largest register of controlled trials available anywhere, the Cochrane Controlled Trials Register (CCTR & CENTRAL). , This contains, in the January 1998 release, abstracts of 1,58,065 randomised or controlled clinical trials, obtained from MEDLINE as well as from hand-searching scores of general and specialist health care journals. The Library also contains 646 references in its Review Methodology Database pertaining to the science of systematic reviews and evidence based medicine. The Collaboration is an enterprise that has been liked to the Human Genome Project in its scope and its potential implications for modern medicine (Naylor, 1997).

The Collaboration has 15 centres world wide currently, with 47 active or proposed collaborative review groups of which 8 currently prepare reviews of interest to mental health professionals. The scope and relevance of the Collaboration's activities to mental health care is steadily increasing as new groups are formed and more people volunteer their services to the Collaboration.

## WHAT IS THE GOLD STANDARD - META-ANALYSIS OR RANDOMISED CONTROLLED TRIALS?

This is an area of current controversy with RCTs being considered as the gold standard of efficacy in some quarters (Bailar, 1997) and meta-analysis considered the method providing the strongest evidence in others (Geddes &

Harrison, 1997). Randomised controlled trials (provided randomisation is proper) guarantee avoidance of moderate biases that non-randomised studies cannot avoid. Meta-analysis based on non-randomised evidence are certainly open to bias. Similarly, though meta-analysis are conducted on non-experimental studies (Brown, 1997; Harris & Barraclough, 1997), the science of meta-analysis of observational studies is in it's infancy and doubts have been raised on the validity of this approach due to the difficulty in controlling biases and confounders (Feinstein, 1995; Shapiro, 1997).

There is incontrovertible evidence, however, that large scale randomised trials - multicentered mega-trials involving tens of thousands of subjects such as the ISIS-4 (1995) - avoid random errors and yield definitive findings that improve the treatments of millions of patients. Alternatively, systematic reviews of large RCTs (> 1000 patients each) are a complementary strategy to mega-trials (e.g. : Fibrinolytic Therapy Trialists' Collaborative Group, 1994) and are to be relied on as the best available evidence in the absence of a mega-trial. On the other hand, meta-analysis based entirely on small randomised controlled trials are likely to give erroneous results, even if the magnitude of effect of the pooled results is large and statistically significant, and should be treated with caution (Egger & Smith, 1995). Thus numbers as well as quality seem to determine the gold standard (Table 3).

**TABLE 3**
**HIERARCY OF THE STRENGTH OF THE EVIDENCE BASED ON RESEARCH METHODS**

| | |
|---|---|
| 1a. | Large scale ("mega-trial) multicentred RCT |
| 1b. | Systematic review of large RCTs |
| 2. | Systematic review of moderately large RCTs |
| 3. | One moderately large RCT |
| 4. | Systematic review of small RCTs |
| 5. | One small RCT |
| 6. | Non-Randomized controlled trial |
| 7. | Cohort study |
| 8. | Case-control or correlational study |
| 9. | Case series |

## EVIDENCE BASED PSYCHIATRY

We would all like to believe that our practice of psychiatry is based on sound scientific evidence. In some centres psychiatric practice is based on evidence to the same extent as in other medical disciplines (Geddes & Harrison, 1997). To find the best available evidence in order to answer clinical questions, easily accessible syntheses of updated information is a requisite and the Cochrane Library is a potential source of this information with much to offer mental health professionals. However, many completed reviews on mental health interventions in the Library, and elsewhere in the published literature, are based on small sample RCTs and should be interpreted with caution. For example, the review of ECT in schizophrenia (Tharyan, 1998) is based on only 11 trials that fulfilled selection criteria, and the largest comparison includes only 6 trials with 153 ECT treated patients and 141 controls. It appears that large multi-centered randomised trials are required in some areas of mental health care, pending which the results of systematic reviews, however suspect, are our 'best available evidence'.

Caution is urged in generalising the results of randomised controlled trials to clinical practice.

There often are significant differences between characterestics of included and excluded patients in RCTs (Licht et al., 1997). Whether this differentially affects outcome needs further enquiry.

In the final analysis, good clinical medicine (and psychiatry) will always be both an art as well as a science; obtaining the 'best available evidence' could increase the amount of science in the blend.

## REFERENCES

**Adams, C.E. & Gelder, M. (1994)** The case for establishing a register of randomised controlled trials of mental heath care. *British Journal of Psychiatry,* 164, 433-436.

**Adams, C.E., Power, A., Frederick, K. &**

**Lefebvre, C. (1994)** An investigation of the adequacy of Medline searches for randomised controlled trials (RCTs) of the effects of mental health care. *Psychological Medicine*, 24, 741-748.

**Altman, D.G. (1996)** Better reporting of randomised controlled trials: the CONSORT statement. *British Medical Journal*, 313, 570-571.

**Antman, E.M., Lau, J., Kupelnick, B., Mosteller, F. & Chalmers, T.C. (1992)** A comparison of results of meta-analysis of randomised control trials and recommendations of clinical experts: Treatments for myocardial infarction. *Journal of the American Medical Association*, 268, 240-248.

**Bailar, J.C. (1995)** The practice of meta-analysis. *Journal of Clinical Epidemiology*, 48, 149-157.

**Bailar, J.C. (1997)** The promise and problems of meta-analysis. *New England Journal of Medicine*, 337, 559-561.

**Berlin, J.A., Laird, N.M., Sacks, H.S. & Chalmers, T.C. (1989)** A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine*, 8, 141-151.

**Brown, S. (1997)** Excess mortality of schizophrenia. A meta-analysis. *British Journal of Psychiatry*, 171, 502-508.

**Chalmers, T.C., Celano, P., Sacks, H.S. & Smith, H. Jr. (1983)** Bias in treatment assignment in controlled clinical trials. *New England Journal of Medicine*, 309,1358-1361.

**Cook, D.J., Sackett, D.L. & Spitzer, W.O. (1995)** Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *Journal of Clinical Epidemiology*, 48, 161-171.

**DerSimonian, R. & Laird, N. (1986)** Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-188.

**Dickersin, K., Chan, S., Chalmers, T.C., Sacks, H.S. & Smith, H. (1987)** Publication bias and clinical trials. *Controlled Clinical Trials*, 8, 343-353.

**Dickersin, K. (1990)** The existence of publication bias and risk factors for it's occurrence. *Journal of the American Medical Association*, 263, 1385-1389.

**Dickersin, K., Scherer, R. & Lefebvre, C. (1994)** Identifying relevant studies for systematic reviews. *British Medical Journal*, 309, 1286-1291.

**Easterbrook, P.J., Berlin, J.A., Gopalan, R. & Matthews, D.R. (1991)** Publication bias in clinical research. *Lancet*, 337, 867-872.

**Egger, M. & Smith, G.D. (1995)** Misleading meta-analysis. *British Journal of Psychiatry*, 310, 752-754.

**Feinstein, A.R. (1995)** Meta-analysis: statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, 48, 71-79.

**Felson, D.T. (1992)** Bias in meta-analytic research. *Journal of Clinical Epidemiology*, 45, 885-892.

**Gardner, M.J. & Altman, D.G. (1986)** Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal*, 292, 746-750.

**Geddes, J.R. & Harrison, P.J. (1997)** Closing the gap between research and practice. *British Journal of Psychiatry*, 171, 220-225.

**Glass, G.V. (1976)** Primary, secondary, and meta-analysis of research. *Education Research*, 5, 3-8.

**Goodman, S.N. (1991)** Have you ever meta-analysis you didn't like? *Annals of Internal Medicine*, 114, 244-246.

**Gotzsche, P.C. (1987)** Reference bias in reports of drug trials. *British Medical Journal*, 295, 654-656.

**Harris, E.C. & Barraclough, B. (1997)** Suicide as an outcome for mental disorders. A meta-analysis. *British Journal of Psychiatry*, 170, 205-228.

**Haynes, R.B. (1992)** Clinical review articles: should be as scientific as the articles they review. *British Medical Journal*, 304, 330-331.

**Khan, K.S., Daya, S. & Jadad, A. (1996)** The importance of quality of primary studies in producing unbiased systematic reviews. *Archives of Internal Medicine*, 156, 661-666.

**Huston, P. & Moher, D. (1996)** Redundancy, disaggregation, and the integrity of medical research. *Lancet*, 347, 1024-1026.

**Lau, J., Antman, E.M., Silva, J.J.,**

**Kupelnick, B., Mosteller, F. & Chalmers, T.C. (1992)** Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine*, 327, 248-254.

**Lau, J., Schmid, C.H. & Chalmers, T.C. (1995)** Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology*, 48, 45-57.

**Laupacis, A., Sackett, D.L. & Roberts, R.S. (1988)** An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine*, 318, 1728-1733.

**LeLorier, J., Gregoire, G., Benhaddad, A., Lapierre. J. & Derderian, F. (1997)** Discrepancies between meta-analysis and subsequent large randomized, controlled trials. *New England Journal of Medicine*, 337, 536-542.

**Licht, R.W., Gouliaev, G., Vestergaard, P. & Frydenberg, M. (1997)** Generalisability of results from randomised drug trials. A trial on antimanic treatment. *British Journal of Psychiatry*, 170, 264-267.

**Moher, D., Jadad, A., Nichol, G., Penman, M., Tugwell, T. & Walsh, S. (1995)** Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16, 62-73.

**Mulrow, C.D. (1987)** The medical review article : state of the science. *Annals of Internal Medicine*, 106, 485-488.

**Mulrow, C.D. (1994)** Rationale for systematic reviews. *British Medical Journal*, 309, 597-599.

**Mulrow, C.D. & Oxman, A.D. (Eds.) (1994)** Cochrane Collaboration Handbook updated September 1997, In : *The Cochrane Library database on disc and CDROM*, issue 4., the Cochrane Collaboration. Oxford : Update Software

**Naylor, C,D. (1997)** Meta-analysis and meta-epidemiology of clinical research. *British Medical Journal*, 315, 617-619.

**Oxman, A.D. & Guyatt, G.H. (1988)** Guidelines for reading literature reviews. *Canadian Medical Association Journal* , 138, 697-703.

**Oxman, A.D. & Guyatt, G.H. (1992)** A consumer's guide to subgroup analyses. *Annals of Internal Medicine*, 116, 78-84.

**Oxtoby, A., Jones, A. & Robinson, M. (1989)** Is your 'double-blind' design truly double-blind? *British Journal of Psychiatry*, 155, 700-701.

**Rosenthal, R. (1979)** The "file-drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.

**Sackett, D.L. (1979)** Bias in analytic research. *Journal of Chronic Diseases*, 32, 51-63.

**Schulz, K.F., Chalmers, I., Hayes, R.J. & Altman, D.G. (1995)** Empirical evidence of bias : dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association*, 273, 408-412.

**Shapiro, S. (1994)** Meta-analysis shmeta-analysis. *American Journal of Epidemiology*, 140, 771-778.

**Shapiro, S. ( 1997)** Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *Journal of Clinical Epidemiology*, 50, 223-229.

**Smith, M.L. & Glass, G.V. (1977)** Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.

**Smith, R. & Roberts, I. (1997)** An amnesty for unpublished trials. *British Medical Journal*, 315, 622.

**Stewart, L.A., & Parmar, M.K.B. (1993)** Meta-analysis of the literature or of individual patient data : is there a difference? *Lancet*, 341, 418-422.

**Teo, K., Yusuf, S., Collins, R., Held, P.H. & Peto, R. (1991)** Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *British Medical Journal*, 303, 1499-1503.

**Tharyan, P. (1997)** Systematic versus traditional reviews in psychiatry : the case of electroconvulsive therapy for people with schizophrenia. Paper presented at the 49th Annual National Conference of the Indian Psychiatric Society held at Kanpur , January, 1997.

**Tharyan, P. (1998)** Electroconvulsive therapy for schizophrenia, In : *Schizophrenia Module of The Cochrane Database of Systematic Reviews, updated 01 December 1997*, (Eds.)

Adams, C.E., Duggan, L., Wahlbeck, K. & White, P., The Cochrane Library (database on disk and CDROM). Issue 1. Updated quarterly. The Cochrane Collaboration, Oxford : Update Software.

**The Standards of Reporting Trials Group (1994)** A proposal for structured reporting of randomized controlled trials. *Journal of the American Medical Association*, 272, 1926-1931.

**The ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group (1995)** ISIS-4 : a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet*, 345, 669-685.

**Thompson, S.G. (1995)** Why sources of heterogeneity should be investigated, In : Systematic reviews, (Eds.) Chalmers, I. & Altman, D.G., pp 64-74. London : BMJ Publishing Group.

**Villar, J., Carroi, G. & Belizan, J.M. (1995)** Predictive ability of meta-analysis of randomized controlled trials. *Lancet*, 345, 772-776.

**Yusuf, S., Peto, R., Lewis, J., Collins, R. & Sleight, P. (1985)** Beta blockade during and after myocardial infarction : an overview of the randomised trials. *Progress in Cardiovascular Diseases*, 27, 335-371.

---

*PRATHAP THARYAN, M.D., MRCPsych, Professor of Psychiatry, Christian Medical College, Vellore 632002.*