

Appraisal of meta-analysis

Apu Chakraborty, James P. Warner and Robert Blizard

Aims and method Prompted by a clinical question, we critically appraised a meta-analysis of neuroimaging in our evidence-based journal club.

Results The results of the meta-analysis suggested differences in ventricular size and sulcal width between controls and people with schizophrenia and mood disorders. However, we were unable to answer the question that prompted this exercise.

Clinical implications Although the evidence-based medicine approach facilitates appraisal of complex articles, some clinical questions are not yet answerable.

Meta-analysis is a quantitative technique for combining the results from several clinical trials to provide an objective overview of their conclusions, usually in the form of one or more global measures. Large randomised, controlled trials are still considered the gold standard in evaluating the efficacy of clinical interventions (Le Lorier *et al.*, 1997). However, large trials are difficult to undertake especially when the disorder being studied is uncommon, and the best available evidence is often found in a series of small trials. Meta-analysis is a technique frequently used to overcome lack of power of small trials. Unlike a review, when an author appraises some of the existing trials and comes to an individual conclusion about the results, meta-analysis involves combining the data from separate trials and reanalysing these data.

With its ability to summarise all information from several trials into a single result, meta-analysis appears attractive to the reader but may lead to false conclusions. Therefore, the results of meta-analyses should not be accepted uncritically. This example of a critique of a meta-analysis, taken from our journal club, shows how the evidence-based approach can simplify and clarify a complex article and provide the appraiser with a framework for assessing the validity and results.

Does brain imaging help distinguish mood disorders from schizophrenia?

Vignette

A 36-year-old musician has been living in a rehabilitation unit for 11 months. He currently complains of 'difficulty in connecting thoughts',

having 'no power to control things' and feeling 'mentally and physically tired'. He has been treated for psychotic illnesses in the past and has responded equally well to antipsychotics and antidepressants. Careful clinical evaluation over time has left us uncertain about whether the diagnosis is schizophrenia or depression. During discussion of this case, the question of whether any investigations, particularly brain imaging, would help to clarify the diagnosis arose.

Question

For patients where the diagnosis is unclear, does computerised tomography (CT) scanning have a role in distinguishing mood disorder from schizophrenia?

Literature search

Knowing that there have been many published studies on brain imaging in schizophrenia, we began by searching the Medline database for the years 1992 to 1996, specifically looking for a meta-analysis. We felt the best way to identify relevant papers was to limit the 'publication type' menu to meta-analysis. Often, the database will interpret the word or phrase you wish to search and decide on the most appropriate subject heading. For example, we tried to search using the phrase 'CT scan' and the database converted this to 'tomography, X-ray computed' (23 350 articles found). Likewise, using the term 'depression' yields fewer articles than 'depressive disorder' (6964 articles found). The 'tomography, X-ray computed' was limited to meta-analysis, and 12 articles were found. Combining these 12 articles with the 'depressive disorder' search produced one article.

The article identified looked relevant. However, in order to check that no important studies had been missed we scanned all the 12 abstracts of the meta-analyses identified. We then scrutinised the 33 articles identified by combining the 'tomography, X-ray computed' and 'depressive disorder' searches. This identified two comprehensive reviews, but no further meta-analysis. Finally, we repeated the search substituting 'depressive disorder' with 'schizophrenia'. No additional relevant articles were identified.

The article selected for appraisal was entitled 'Meta-analysis of studies of ventricular enlargement and cortical sulcal prominence in mood disorders – comparisons with controls or patients with schizophrenia' (Elkis *et al*, 1995).

The paper was obtained from our library.

Brief outline of the article

The paper states that the evidence for higher rates of ventricular enlargement and/or cortical sulcal widening is already well documented in schizophrenia. Similarly conclusive findings are not recorded for mood disorders. This is because a 'vote-counting' method is often used which fails to reveal an effect size if the true effect and the sample size are no more than moderate (often the case). A meta-analytic review is more likely to detect such a difference.

To identify relevant studies to be included in the meta-analysis, the authors conducted a Medline search for 1966 to 1994. To be included, a study had to have compared people with mood disorders with normal controls or with people with schizophrenia on some measure of ventricular enlargement or cortical sulcal prominence. The data had to be unique and provide sufficient information for the extraction of an effect size. Thirty-three studies were selected and four meta-analytic reviews conducted (see Table 1).

Effect size was calculated as the difference between the mood-disordered group and the comparison group divided by the pooled standard deviation. It was deemed 'large, moderate' or 'small', depending on its value. The direction of the effect size was positive if the results supported the hypothesis that those with mood disorders have more ventricular enlargement or more sulcal prominence than the comparison group. Analysis of variance was used to assess the influence of potential moderators of effect size (year, type of control, proportion of mood-disordered subjects who were male, mean age of mood-disordered subjects, proportion of sub-

jects with unipolar depression, imaging modality, interrater reliability, and ease of effect size extraction).

Publication bias was examined graphically by using 'funnel plots'. Total sample size (ordinate) was plotted against effect size (abscissa). Large studies equate with the true population effect, with random scatter increasing about this central effect as study size decreases, producing an inverted funnel. If the pattern of results does not conform to the inverted funnel shape, there is evidence that certain results are not represented. Often this occurs when small negative studies are not published.

Meta-analysis 1 and 2 yielded highly statistically significant composite effect sizes ($P < 0.001$), supporting the notion that patients with mood disorders do have larger ventricles and more prominent sulci when compared with controls (Table 1). Meta-analysis 3 indicated a small but highly statistically significant effect size ($P = 0.002$), implying that people with schizophrenia have greater ventricular enlargement than mood-disordered subjects. There were too few studies in meta-analysis 4 to permit useful quantitative analysis.

Critical appraisal

The article was appraised using guidelines by Oxman *et al* (1994).

Are the results valid?

Did the overview address a focused question? Yes. The questions addressed in this paper were clearly defined.

Were the criteria used to select articles for inclusion appropriate? Yes. The authors used strict inclusion criteria for selecting the papers. There had to be assessment of ventricular size or sulcal prominence. The studies had to include subjects with mood disorders, either compared

Table 1. Summary of results of neuroimaging studies on patients with depression selected for meta-analysis (Elkis *et al*, 1995)

Meta-analysis	Comparison group	Region	Number of studies	Combined number of subjects		Composite effect ¹ (95% confidence intervals)
				Affective disorder	Comparison	
1	Normal	Ventricular enlargement	29	912	932	+0.437 (0.34 to 0.53)
2	Normal	Sulcal prominence	10	360	323	+0.421 (0.26 to 0.58)
3	Schizophrenia	Ventricular enlargement	11	406	599	-0.201 (-0.33 to -0.07)
4	Schizophrenia	Sulcal prominence	3	-	-	-

1. Effect size is expressed as a fraction: difference between groups/pooled standard deviation.

to normal controls or individuals with schizophrenia, and had to provide adequate statistical information. The data had to be unique, i.e. not a duplicate publication. A drawback to these rigorous criteria is that some studies that did not show significant results were rejected from the analysis because of insufficient statistical information.

Is it likely that relevant studies were missed?

Yes. The authors based their analysis on papers identified by searching one electronic database (Medline). They did not appear to contact researchers in the field for unpublished studies or attempt to search other databases. The failure to identify all studies may be an important source of bias, since small negative studies are often not published, or published in journals not referenced in Medline. However, the funnel plot of the meta-analyses suggested that several studies with a negative effect size were identified, reducing the likelihood of publication bias.

Publication bias refers to an editorial preference for studies with positive results and a meta-analysis that is unable to locate unpublished studies is prone to overestimate effect size. Meta-analyses may reach erroneous conclusions as a result of publication bias (Egger & Davey-Smith, 1998). Rigorous meta-analysis should include a comprehensive trawl for published trials and an exhaustive search for 'grey data', i.e. results of unpublished trials. Although the authors did not do this, the funnel plot suggests few studies were missed.

Was the validity of the included studies appraised? No. The authors did not give explicit information about assessment of the methodological quality (and hence the validity) of the studies in the meta-analysis.

Were the assessments of studies reproducible? No. The authors did not appear to use independent reviewers. Selection of studies for a meta-analysis is subject to bias, and should ideally be done by at least two individuals using the same selection criteria. There should be good concordance between the reviewers.

What are the results?

What are the overall results of the meta-analyses?

(a) Mood disorders compared with normal controls The combined *n* in meta-analyses 1 and 2 were large: they evaluated nearly 40 studies with over 1800 and 600 subjects respectively (see Table 1). Computed effect sizes demonstrated significantly larger ventricles and more prominent sulci in those with mood disorders com-

pared with controls. In both cases, the effect size of around 0.4 corresponds to a moderate difference (Cohen, 1988). In other words, assuming a normal distribution, 50% of the depressed group would score above the mean measurement, compared with 33% of controls. In both, vote-counting alone would not display such an effect. Despite the rigorous statistical methods used, the lack of appraisal of validity of individual studies and the potential for selection bias should be kept in mind before too much significance is attached to the results.

(b) Mood disorders compared with schizophrenia Meta-analysis 3 showed that patients with mood disorders have slightly smaller ventricles than those with schizophrenia. Although the combined effect was statistically significant, each of the nine individual studies reported this comparison as non-significant. Only three studies met the inclusion criteria for meta-analysis 4 and the authors did not analyse the aggregated results.

How precise are the results? The confidence intervals for the effect sizes are given for meta-analyses 1, 2 and 3 (see Table 1). These show significant effect sizes in all three cases.

Will the results help me in patient care?

Can the results be applied to my patient? However statistically significant these structural differences may be, they are not yet robust enough to be employed as reliable diagnostic markers to change management in the clinical setting. The authors did not set out to refine the diagnostic utility of brain imaging, and the sensitivity, specificity and likelihood ratio of this investigation as a diagnostic test are not calculated. Unfortunately, this paper brings us no nearer to a clear-cut diagnosis of our patient.

Were all clinically important outcomes considered?

This question is more relevant to meta-analyses on therapies. Decisions about whether to adopt a particular intervention depend on an appraisal of safety as well as efficacy. The article by Elkis *et al* (1995) does not consider potential drawbacks to scanning, although in reality these are likely to be minimal.

Are the benefits worth the harms and costs?

There was no explicit analysis of the costs involved. Again, this question is more applicable when a meta-analysis suggests an intervention such as a diagnostic test or a therapy.

Discussion

Accurate searching of electronic databases is a cornerstone of evidence-based medicine. It is important to use the correct medical subject headings (MeSH) and syntax (Greenhalgh, 1997). For example, searching Medline using the word 'depression', will miss a vast amount of literature logged on the database as 'depressive disorder'. Our search strategy could have been improved by using terms such as 'systematic review' rather than limiting to meta-analysis. Also, in retrospect we should have searched the Cochrane database of systematic reviews (Bero & Rennie, 1995), which is a source of reviews carried out to specified standards.

A common method of reviewing cumulated studies addressing a single question is to employ vote-counting. In this method, a reviewer comes to a conclusion based on the number of studies showing a positive effect, a negative effect or no effect. This is unsound as it ignores differences in research design, sample size and effect size. Meta-analysis circumvents some of these difficulties, by enabling several small comparable studies to be considered in a way that identifies relatively small effect sizes (as shown here). Despite this rigorous statistical approach, a systematic comparison with randomised clinical trials shows that some meta-analyses may have poor predictive ability (Le Lorier *et al.* 1997).

The problem with pooling results in this way is, from a clinical standpoint, production of an 'average' population effect may not be relevant to the clinician's particular patient, although the same criticism can be levelled at single trials. Databases may be biased towards publications published in English. Citation bias may occur in that trials that are supportive of a positive effect are more easily published than negative trials. The inclusion of duplicate data may lead to an overestimation of treatment effects. Occasionally, more than one meta-analysis exists on a particular topic, and concordance of conclusions from meta-analyses may be reassuring.

Although the evidence-based approach highlighted possible weaknesses of this article, notably identification of all relevant studies, possible selection bias and lack of appraisal of the validity of the studies analysed, this meta-analysis should not be rejected out of hand. It probably represents the best available evidence regarding CT scanning and mood disorders. An evidence-based approach certainly helped us to assess the usefulness of this study, and provided a structured method of appraising what was at first sight a complex paper. We have not been able to answer our initial question, probably because, in retrospect, the question was too ambitious, and is not yet answerable. Further-

more, the fact that the article we identified did not address the question of diagnostic utility in terms of validity or a likelihood ratio further hampered our ability to answer our question.

Most meta-analyses are of controlled trials of a therapeutic intervention and we felt that choosing one of these would have made this exercise easier. However, meta-analyses of observational studies, including aetiological associations and diagnostic tests are becoming increasingly common (Egger *et al.* 1998), and the principle of pooling the data remains the same. The system of appraisal in terms of validity, results and applicability is also unchanged. However, Egger *et al.* (1998) warn that meta-analyses of observational studies may produce spurious results and should be treated with caution.

This journal club exercise was not a good example of the evidence-based medicine process at work. However, this is probably a reflection of an over-ambitious question and an absence of relevant articles than a flaw in the process itself. Our understanding of the appraisal of meta-analytic studies improved.

References

- BERO, L. & RENNIE, D. (1995) The Cochrane collaboration. Preparing, maintaining and disseminating systematic reviews of the effects of health care. *Journal of the American Medical Association*, **274**, 1935-1938.
- COHEN, J. (1988) *Statistical Power Analysis for the Behavioural Sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.
- EGGER, M. & DAVEY SMITH, G. (1998) Meta-analysis: bias in location and selection of studies. *British Medical Journal*, **316**, 61-66.
- , SCHNEIDER, M. & DAVEY SMITH, G. (1998) Spurious precision? Meta-analysis of observational studies. *British Medical Journal*, **316**, 140-144.
- ELKIS, H., FRIEDMAN, L., WISE, A., *et al.* (1995) Meta-analyses of studies of ventricular enlargement and cortical sulcal prominence in mood disorders - comparisons with controls or patients with schizophrenia. *Archives of General Psychiatry*, **52**, 735-746.
- GREENHALGH, T. (1997) The Medline database. *British Medical Journal*, **315**, 180-183.
- LE LORIER, J., GREGOIRE, G., BENHADDAD, A., *et al.* (1997) Discrepancies between meta-analyses and subsequent large randomised controlled trials. *New England Journal of Medicine*, **337**, 536-542.
- OXMAN, A. D., COOK, D. J. & GUYATT, G. H. (1994) Users' guide to the medical literature VI. How to use an overview. *Journal of the American Medical Association*, **272**, 1367-1371.

Apu Chakraborty, *Senior House Officer*, *James P. Warner, *Lecturer*, and Robert Blizard, *Medical Statistician, University Department of Psychiatry, Royal Free Hospital, Pond Street, London NW3 2QG; e-mail: warner@rfhsm.ac.uk*

*Correspondence