# Editoriali/Editorials

# Systematic reviews and the Cochrane Collaboration: improving the evidence-base for all psychiatric treatments

RACHEL CHURCHILL

## INTRODUCTION

Mental health policy and practice requires judgment, balancing experience, knowledge, values and information about available resources. However, over the last thirty years, health professionals in general have been criticized for providing healthcare without adequate evidence of effectiveness, efficiency and safety. Many authors have demonstrated the gap between scientific evidence and clinical practice as well as the diversity of methods used to manage similar health problems, all resulting in variations in clinical practice. These criticisms have highlighted the need to make high quality research-based evidence widely available to inform clinical decision-making, prompting important developments in healthcare research. One methodological development has been to improve the practices used to summarise primary research. This has led to the preparation and maintenance of high quality systematic reviews of the best available evidence, such as those produced by the Cochrane Collaboration.

This paper explains how the science of reviews has advanced, outlining the procedures used to undertake a systematic review and highlighting some of the advantages of this approach. The role of the Cochrane Collaboration in producing and maintaining systematic reviews is explained and, in particular, the work of the groups involved in generating mental health reviews is

described, indicating the breadth of topics covered. Finally, some examples of systematic reviews evaluating psychological interventions are provided.

## THE NEED FOR SYSTEMATIC REVIEWS

Traditional narrative reviews are rarely comprehensive summaries of the available evidence. They are open to subjective judgements and potential biases in the identification, selection and interpretation of included studies. These reviews also typically summarise studies according to whether or not they support one conclusion or another. Where differences between studies in sample size, effect size and quality have been ignored, the conclusions of narrative reviews are likely to be unreliable and misleading. Examples in other areas of medicine have demonstrated how expert opinion, original research and traditional reviews have conflicted with one another, and how misguided were their recommendations when compared with the available RCT evidence, had this been properly summarised at the time (Antman et al, 1992).

A common feature of these reviews is that they include only published studies, leaving the review open to publication bias. This type of bias can occur because research with significant or favourable results is more likely to be published. Publication bias and the practices associated with it (such as the selective reporting of statistically significant results in studies with multiple outcomes) have been shown to prevail for over 30 years in the psychology literature (Sterling, 1959; Sterling et al., 1995) and is a widespread problem across the whole of medical research (Easterbrook et al., 1991). The problem is worsened because statistically significant findings are also more likely to generate multiple publications (which can be difficult to detect), more likely to be published in high-

Address for correspondence: Professor R. Churchill, Department of Health Services Research, Institute of Psychiatry, PO Box 32, De Crespigny Park, London SE5 8AF (UK).
Fax: +44-20-7848.0530
E-mail: r.churchill@iop.kcl.ac.uk
**Declaration of Interest:** the author is employed by the Institute of Psychiatry and receives no funding from any commercial organization.

impact English language journals (Dickersin, 1990), more likely to be cited, and more likely to be published much earlier than non-significant findings (Stern & Simes, 1997). As a result, where only published studies are included in a review, the results are likely to over-estimate the efficacy of treatment.

Systematic reviews are reviews that have been pre-pared using a systematic approach to minimizing biases and random errors, and where these procedures are docu-mented in a materials and methods section (Chalmers & Altman, 1995). In particular, systematic review methods help to limit the potential for selection bias and the effects of publication bias.

## STAGES OF DOING A SYSTEMATIC REVIEW

The steps involved in undertaking a systematic review are summarized in table I. In order to limit the potential for bias, *a priori* decisions about all aspects of the methodology must be carefully laid out in a protocol prior to undertaking the review. Because of the need for transparency and to enable replication of the review, these methods should be described in detail. As with any research, formulating a structured research question for a review is crucial. The objectives of the review need to be stated in terms of the patient group, interventions under study and the outcomes of interest. A more detailed description of eligibility criteria should state what study designs, study participants, interventions, comparisons and outcomes are to be included and excluded from the review.

It is essential that all potentially eligible published and unpublished studies are identified and the search strategy for the review should be very clearly documented to demonstrate this. A comprehensive, unbiased search strategy is one of the key differences between a system-atic review and a traditional review and helps to reduce both the potential for selection bias and the effects of publication bias. Sole dependence on electronic data-bases is not adequate. Electronic databases such as Medline are powerful tools for locating studies, but only between 30 and 80% of all known published randomised controlled trials are identifiable using this database. In mental health, only about half the available RCTs can be identified using an optimal search strategy (Adams *et al.*, 1994). Therefore, searches of electronic databases such as Medline and PsycInfo should always be supplemented by handsearching of specialist journals, searches of spe-cialised registers and grey literature (including disserta-tions & conference abstracts), checking of reference lists

and personal communications.

To avoid biases in selecting studies for inclusion, the eligibility of each study for inclusion should be based on the inclusion/exclusion criteria laid out in the review pro-tocol. Study eligibility should be judged by two indepen-dent observers with a third observer to help resolve any disagreements. Again for the purposes of transparency, excluded studies, along with the reasons for exclusion, form part of the review. The quality of studies is of obvi-ous importance to systematic reviews. The best available evidence should always be used, and in treatment studies, the best evidence comes from RCTs. Randomised designs help to limit the effects of bias and these should be used wherever possible. However, even RCTs are not immune to bias. There is a large body of research demon-strating how a number of specific methodological fea-tures can affect the results of trials, and many others are suspected. Suitable methods for assessing methodologi-cal quality are still widely debated and many different quality scales and checklists are available. However, empirical evidence suggests that a more appropriate assessment of study quality involves identifying method-ological characteristics that might be important in the protocol and assessing the effects of these individually (Jüni *et al.*, 1999). To avoid errors, it is important that two independent assessors extract all the data, preferably using a carefully designed and piloted standardized data extraction form.

The characteristics of included studies should be tabu-lated. In many cases a meta-analysis may be undertaken, to provide a synthesis of the numeric outcome data from a group of similar studies making the same comparisons. In a meta-analysis, patients enrolled in the same study are compared directly and a weighted average of the results is calculated, giving larger studies more influence than smaller ones. The results are graphically displayed together with their confidence intervals in a 'forest plot'. Although not very powerful, statistical tests of hetero-geneity should be undertaken to assess whether individ-ual study results are likely to reflect a single underlying effect, as opposed to a distribution of effects.

The direction and magnitude of the average effect is intended to help guide decisions about clinical practice. However, many features of a study can influence out-come. For example, the effects of a given treatment are unlikely to be identical across different groups of patients and the interventions themselves may vary in their imple-mentation and delivery. Where possible, it is important to examine the influence of these factors on outcome, not only to avoid a misleading analysis, but also because these may be important factors in guiding treatment deci-

sions. Factors deemed to be important should be pre-specified in the protocol to avoid the dangers of a post hoc analysis. Finally, key assumptions and decisions made by reviewers should always be explored as part of a sensitivity analysis. Again, the factors deemed to be important, such as study quality, should ideally be pre-specified in the protocol. By excluding studies from the analysis as appropriate, the robustness of the findings to the effects of different assumptions can be examined. The influence of potential publication bias should also be investigated. The interpretation of findings should take account of the limitations of the included studies, the limitations of the review itself, the strength of the evidence available and the applicability of the review to the context in which it will be applied. The review should be published regardless of the direction of its findings.

## ADVANTAGES OF SYSTEMATIC REVIEWS

Whatever the outcome, systematic reviews provide an objective summary of the current state of the evidence on a particular topic. Because the process of doing systematic reviews is transparent, they can be replicated by others to check their reliability and accuracy. Their findings can be used to inform treatment decisions as well as policy and service development. They can also provide the rationale and justification for further primary research and can guide future research by demonstrating the limitations of the available evidence-base. Where a meta-analysis is possible, combining data from a number of smaller studies results in increased power to detect or exclude relatively small effects, offering an attractive alternative to undertaking a new adequately powered large primary study. Meta-analysis is particularly appealing because single studies often lack the statistical power to detect or

Table I. – *Steps in conducting a systematic review\*.*

| | |
|---|---|
| *1  Formulate review question* | treatment allocation, blinding and handling of patient attrition<br>• consider blinding of observers to authors, institutions and journals |
| *2  Define inclusion and exclusion criteria*<br> • participants<br> • interventions and comparisons<br> • outcomes<br> • study designs and methodological quality | |
| | *6  Extract data*<br> • design and pilot data extraction form<br> • consider data extraction by more than one observer<br> • consider blinding of observers to authors, institutions and journals |
| *3  Locate studies*<br>Develop search strategy considering the following sources:<br> • *The Cochrane Central Register of Controlled Trials* (CENTRAL)<br> • electronic databases and trials registers not covered by CENTRAL<br> • checking of reference lists<br> • handsearching of key journals<br> • personal communications with experts in the field | |
| | *7  Analyse and present results*<br> • tabulate results from individual studies<br> • examine forest plot<br> • explore possible sources of heterogeneity<br> • consider meta-anlaysis of all trials or subgroups of trials<br> • perform sensitivity analyses, examine funnel plots<br> • make list of excluded studies available to interested readers |
| *4  Select studies*<br> • have eligibility checked by more than one observer<br> • develop strategy to resolve disagreements<br> • keep log of excluded studies, with reasons for exclusions | |
| | *8  Interpret results*<br> • consider limitations, including publication and related biases<br> • consider strength of evidence<br> • consider applicability<br> • consider numbers-needed-to-treat to benefit / harm<br> • consider economic implications<br> • consider implications for future research |
| *5  Assess study quality*<br> • consider assessment by more than one observer<br> • use simple checklist rather than quality scales<br> • always assess concealment of | |

\* *Points 1-7 should be addressed in the review protocol*
Modified version of table taken from Egger *et al* (2001). Reproduced by kind permission of the publishers.

exclude a modest, but clinically important difference between two treatments. There are many examples of such low powered studies in psychiatry. For example, Thornley and Adams demonstrated that of 1941 trials in schizophrenia, only 3% were large enough to detect important effects (Thornley & Adams, 1998). Similarly, Hotopf and others showed that in 122 antidepressant trials, just 10% could have detected a difference of 20% in recovery between TCAs and SSRIs with 80% power and 95% confidence (Hotopf *et al.*, 1997). A great advantage of systematic reviews is the ability to identify potential differences in treatment response by examining clinical heterogeneity. For example, Davey Smith and others examined the effects of clinical heterogeneity by stratifying studies of cholesterol-lowering treatment by rate of death from CHD in the control-arms of each study, and demonstrated that the net benefit from cholesterol-lowering was only evident in those studies where baseline risk of CHD was high. In the low-risk group, these treatments actually had an adverse effect (Davey Smith *et al.*, 1993). Such an analysis would not ordinarily be possible in a single study due to a lack of power. Finally, there will always be different views on the correct method for performing a review, and subjective judgements are unavoidable. However, in a systematic review these judgements are explicit and their effects on the findings and conclusions of the review can be tested objectively in a sensitivity analysis.

## The Cochrane Collaboration

Systematic reviews produced by the Cochrane Collaboration aim to address the evidence-needs of healthcare providers, consumers, researchers and policy-makers (Clarke & Oxman, 2000) and provide a basis for rational decision-making (Mulrow, 1994). These reviews focus on the best available evidence, usually from RCTs. Cochrane reviews are disseminated, maintained and commented upon via the Cochrane Library. The Collaboration aims to ensure that these reviews are updated at least once every two years. The Library also contains review protocols detailing information about ongoing reviews, as well as a range of other databases, including the *Cochrane Central Register of Controlled Trials* (CENTRAL). The Cochrane Library is made available to users on a subscription basis, either on CD ROM or online via http://www.update-software.com/cochrane, although several countries have arranged national provisions or free access. Abstracts of all the reviews and titles of published protocols can be viewed over the internet (http://www.cochrane.org/cochrane/revabstr/).

The Collaboration is organized into a number of different groups. There are about 15 Cochrane Centres around the world, including one located in Italy (http://www.areas.it). These centers support the work of locally based members of the Collaboration. The main work of the Collaboration, that of producing and maintaining reviews and protocols, is done by about 50 Collaborative Review Groups.

## Collaborative Review Groups (CRGs)

CRGs are composed of individuals (researchers, healthcare professionals, consumers and others) from around the world who share an interest in developing and maintaining systematic reviews relevant to a particular health area. CRGs fulfill an essential role in supporting the reviewers in preparing and maintaining high quality systematic reviews. They enable a multidisciplinary approach to producing evidence in a particular field, ensuring the relevance of reviews to a range of professional groups. Due to the diversity of therapies available and the range of practitioners involved, this is especially important in mental health. CRGs are coordinated by an editorial team comprising a Coordinating Editor, a Review Group Coordinator, a Trials Search Coordinator and a board of editors from a number of different countries. The editorial team supplies methodological and content expertise throughout the preparation of a review, helps to prevent duplication, and provides formal training, guidance, and other resources, including hard copies of papers and translations of articles. This team also organizes comprehensive peer-review, before assembling completed protocols and reviews for publication in the Cochrane Library. Once published, CRGs ensure that these manuscripts are kept up to date.

Each CRG has developed and maintains its own specialist register of trials and aims to identify studies within its scope in a way that will ensure comprehensiveness, protect against bias and minimise needless duplication of effort. Cochrane reviewers have the advantage of having access not only to the CRGs register, but also to the information specialist who manages it. Since many reviewers have only limited expertise in using electronic databases, this can be a distinct advantage. To provide a simple example, the reviewer who searches Medline for articles by exploding the term 'Depressive disorders' will identify less than 2% of the studies he would have identified had he searched using the term 'Depression'. CRGs undertake regular searches of all relevant electronic databases and other specialized registers (including registers on ongoing trials), coordinate handsearching of specialist journals, continuously search the grey literature and are in regular contact with the pharmaceutical industry.

Many CRGs hold copies of the articles contained in their registers so that these can be made available to reviewers. These registers often record trial details not available in standard databases, providing an excellent source of information about the overall availability of primary evidence for a review in a particular field. For example, the *Cochrane Depression, Anxiety & Neurosis Group* (CCDAN) assembles, maintains and provides reviewer access to two specialised registers of controlled trials. *CCDANCTR-References* currently contains more than 13,800 references (with a quarterly accrual rate of about 400 new records) relating to completed or ongoing trials. *CCDANCTR-Studies* is the product of an ongoing project funded by the EU to convert the registers of all Cochrane mental health groups into study-based registers. CCDAN has coded nearly 10,000 references to around 8,000 trials, identifying multiple publications and indexing the trials for important information including disorder, setting, intervention and outcome. The references from the trials registers of all the Cochrane CRGs are available on the CENTRAL database in the Cochrane Library.

## Cochrane reviews in mental health

In mental health, Cochrane reviews are produced with the guidance of five different *Collaborative Review Groups* (CRGs) who have responsibilities for summarising evidence in different areas. These groups produce reviews intended to inform not only pharmacological treatment decisions, but also decisions about psychological and other non-pharmacological interventions. The contact details for each of these GRCs can be found on the Cochrane Contacts website (http://www.cochrane.org/cochrane/contact.htm) and, additionally, several of the groups have their own website links (listed below). The scope of the *Cochrane Dementia and Cognitive Improvement Review Group* (http://www.jr2.ox.ac.uk/-cdcig) is the prevention, treatment, and management of acquired cognitive impairments, their manifestations and complications, and the care of people affected. The *Cochrane Drug and Alcohol Review Group* (which is based in Rome) is concerned with reviewing trials that describe an active intervention (including prevention, treatment and rehabilitation) aimed at reducing the potential for harm or the actual harm directly related to the use of different dependence producing substances. The *Cochrane Developmental, Psychosocial and Learning Problems Review Group* (www.bris.ac.uk/Depts/CochraneBehav) aims to address a range of medical, social, educational and socio-legal problems, including developmental and psychosocial problems of childhood and adolescence (including juvenile delinquency), learning prob-

lems across the lifespan (including, but not restricted to learning disabilities), personality disorders and adult offending. The *Cochrane Schizophrenia Review Group* is concerned with the evaluation of the prevention, treatment and rehabilitation of people with non-affective, functional, psychotic illness. The *Cochrane Depression, Anxiety and Neurosis Review Group* (CCDAN) (http://www.iop.kcl.ac.uk/IoP/CCDAN/index.htm) is concerned with the evaluation of health care relevant to mood disorders, anxiety disorders, somatoform disorders, chronic fatigue syndrome, dissociative disorders, and eating disorders. CCDAN also includes clinically significant problems such as deliberate self-harm, and suicide attempt. To demonstrate the diversity of Cochrane systematic reviews available, some of the reviews produced by CCDAN are discussed in more detail below.

## Systematic reviews of psychological treatments

To date, members of CCDAN have completed and published around 35 reviews and 50 protocols. More than 50 further protocols are currently under development (these can be viewed on the CCDAN website). Over 40% of the CCDAN reviews have been done on non-pharmacological interventions, including alternative treatments (for example, Kava extract for anxiety (Pittler & Ernst, 2003) and St John's Wort for depression (Linde *et al.*, 2003). It is estimated that around 20,000 trials have been conducted that could inform mental healthcare decisions (Lewis *et al.*, 1997), and a significant proportion of these trials involve non-pharmacological interventions. Reviews of psychological interventions currently account for more than a quarter of CCDAN publications, with the number due to increase in forthcoming issues of the Cochrane Library.

Reviews of psychological treatments present special problems. In the area of psychological treatments, the available trial evidence is far more limited than in the field of pharmacological interventions, as demonstrated by the contents of the CCDAN Controlled Trials Register (figure 1). The quality of the studies included in the review is also likely to have an impact – and the factors associated with quality in psychological treatment trials are not necessarily the same as those in pharmacological treatment trials. Psychological treatments are obviously less straightforward to evaluate using standard trial methodology. For example, although blinding of assessors may still be possible, blinding of the therapist is not, and blinding of the patient is rarely an option. Since rather more than simple participation and compliance is required of patients receiving psychological therapies, the effects of individual differences on treatment

response are likely to be greater. For example, the severity of the disorder may have a considerable impact on a patient's ability to engage in therapy, or their beliefs and expectations about a psychological treatment could have a significant effect on their motivation to become actively involved in therapy. It is also more difficult to ensure fidelity to the psychotherapeutic technique under study, and in many cases, the individual therapists approach may vary considerably, particularly where therapies are not standardized or manualised. In many psychological treatment trials, the investigators have been closely allied to the development of the therapy itself – and potential conflict of interest is no less likely than in a trial undertaken and supported by industry. The mechanism of therapeutic action in psychological treatments is less clear, indeed many factors may be responsible for any observed improvement, making it difficult to identify a suitable comparison arm for a trial. Finally, outcomes that are important following psychological interventions and the time-points at which they are assessed might be quite different from those that are commonly measured in trials. All these difficulties with the primary studies make the need for a systematic approach to review the literature all the greater. The popularity of many psychological interventions with both patients and practitioners increases the need for high quality evidence to guide treatment decisions, ensuring that the best therapies are promoted and, where necessary, challenging clinical practice.

Figure 1. – *Interventions studied in trials on CCDANCTR-Studies.*



**Interventions studies in trials on CCDANCTR-Studies**
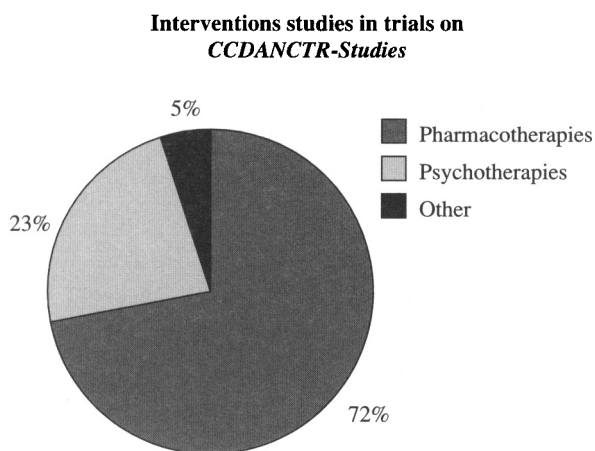
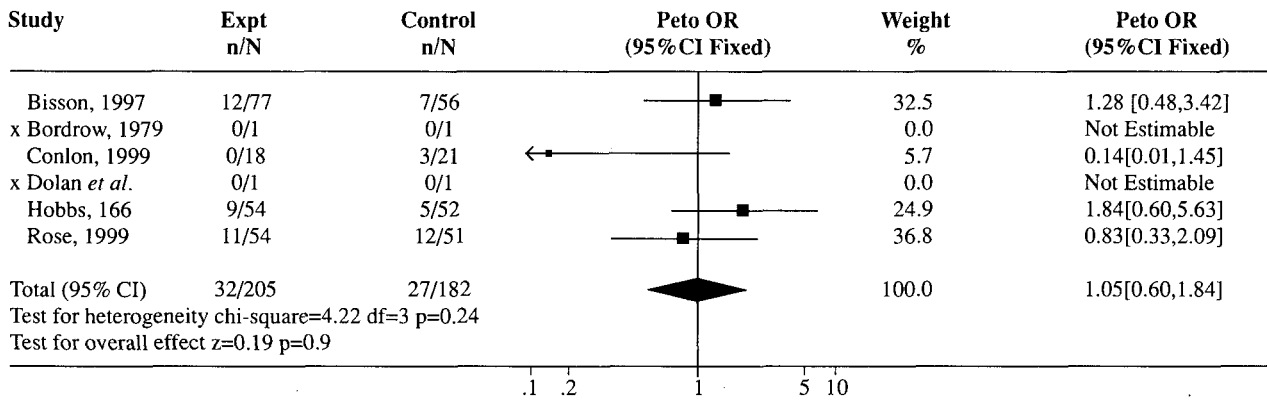One of the most important benefits of systematic reviews has been to demonstrate where there might be some doubt about current practices. Potential adverse consequences are commonly ignored in trials of psychological interventions (information on negative effects is often not considered), but occasionally, the effect of the intervention on the primary outcome provides evidence that the therapy may be harmful. One example of this comes from a Cochrane review which could eventually have a significant impact on policy and practice. Over the last fifteen years, early psychological interventions such as psychological 'debriefing' have been increasingly used to treat psychological trauma. While these interventions have become popular and their use widespread, their efficacy has not been empirically demonstrated. A systematic review first published in 1997 and subsequently updated, assessed the effectiveness of brief psychological debriefing for the management of psychological distress after trauma, and the prevention of post traumatic stress disorder (Rose *et al.*, 2003). Included in the review were trials involving persons recently (one month or less) exposed to a traumatic event who were receiving a single session only of psychological debriefing. The debriefing involved some form of emotional processing/ventilation by encouraging recollection/reworking of the traumatic event, accompanied by normalisation of emotional reaction to the event. Eleven trials fulfilled the inclusion criteria.

Where data synthesis was possible, the results demonstrated that single session individual debriefing neither reduced psychological distress nor prevented the onset of post traumatic stress disorder (PTSD) (see figure 2). Those who received the intervention showed no significant reduction in the short-term risk (3-5 months) of PTSD (odds ratio 1.05 (95% CI 0.60 to 1.84)). At one year, one trial reported that there was a significantly increased risk of PTSD in those receiving debriefing (odds ratio 2.88 (95% CI 1.11 to 7.53)). There was also no evidence that debriefing reduced general psychological morbidity, depression or anxiety. The reviewers concluded that there was currently no evidence that psychological debriefing is a useful treatment for the prevention of post traumatic stress disorder after traumatic incidents, and suggested that compulsory debriefing of victims of trauma should cease. The review is currently undergoing updating following the publication of new trials that also appear to support these findings.

Given the potential for variation in response to psychological treatments, the ability to examine clinical heterogeneity in trials of these interventions can be a considerable advantage. An example of the use of this type of analysis in the assessment of psychological treatment studies comes from a comprehensive review of brief psychological treatments for depression in primary care

Figure 2. – *Meta-analytic presentation of psychological debriefing versus control for the prevention of PTSD (taken from Rose et al., 2003).*

**Comparision: 01 Debriefing versus control**
**Outcome:      03 PTSD (short term)**

| Study | Expt n/N | Control n/N | Peto OR (95%CI Fixed) | Weight % | Peto OR (95%CI Fixed) |
|---|---|---|---|---|---|
| Bisson, 1997 | 12/77 | 7/56 | | 32.5 | 1.28 [0.48,3.42] |
| x Bordrow, 1979 | 0/1 | 0/1 | | 0.0 | Not Estimable |
| Conlon, 1999 | 0/18 | 3/21 | | 5.7 | 0.14[0.01,1.45] |
| x Dolan et al. | 0/1 | 0/1 | | 0.0 | Not Estimable |
| Hobbs, 166 | 9/54 | 5/52 | | 24.9 | 1.84[0.60,5.63] |
| Rose, 1999 | 11/54 | 12/51 | | 36.8 | 0.83[0.33,2.09] |
| Total (95% CI) | 32/205 | 27/182 | | 100.0 | 1.05[0.60,1.84] |

Test for heterogeneity chi-square=4.22 df=3 p=0.24
Test for overall effect z=0.19 p=0.9



(Churchill *et al.*, 2001; 2003). One of the comparisons made was between CBT and treatment as usual and 12 trials provided sufficient data for inclusion in the pooled analysis for post-treatment recovery on a total of 654 patients. The individual odds ratio (OR) for post-treatment recovery was statistically significant in favour of variants of CBT in seven of the trials, whilst the remaining five demonstrated no significant difference from treatment as usual. The pooled OR for recovery with CBT compared with treatment as usual was 3.42 (95 % CI, 1.98 to 5.93) suggesting that the odds of recovery for those receiving variants of CBT were more than three times greater than for those receiving usual treatment. The result was highly significant (z = 4.39, p = 0.00001). Because of statistical heterogeneity, a random effects model (which allows for both within-study sampling error and between-study variation) was used to pool the data from these trials, and potential causes for the observed heterogeneity were explored.

In this review, the baseline severity of participants was thought likely to be an important predictor of outcome. Stratifying the trials according to the degree of baseline severity resulted in a slightly reduced but still highly significant pooled estimate for recovery in favour of variants of CBT for trials in which participants had more severe depression (OR 2.64; 95% CI, 1.37 to 5.09). However, in those trials where baseline severity was unspecified, a much greater OR was observed, suggesting that these trials might be overestimating the effect of CBT (see figure 3).

These trials had used a number of different methods for recruiting patients into the trial. The criteria for inclu-

sion in this review specified the setting in which therapy was provided, but not the source of patient recruitment. Many of the trials involved volunteers and since the motivation of patients is likely to be an important factor in the outcome of therapy, studies were stratified by method of recruitment. This analysis revealed a lower, although more precise OR for those recruited via outpatients clinics and referrals, compared with trials that recruited volunteers or responders to advertisements (see figure 4). This analysis again demonstrated distinct differences in the observed efficacy of CBT for patients recruited from different sources and suggested that inclusion of trials involving volunteers may lead to an overestimate of the effect size.

## CONCLUSIONS

Improvements in methods for reviewing the results of research have improved considerably over the last two decades. The science of systematic reviews has largely replaced the process of writing more traditional narrative reviews - which are prone to bias and likely to be unreliable. In particular, the Cochrane Collaboration is now generating and keeping up-to-date a substantial body of reliable research-based evidence from systematic reviews of trials. These reviews are key resources for practitioners, patients, and health policy-makers, enabling treatment across all areas of health care, including mental health, to become evidence-based.

Clinical practitioners undertake most high quality systematic reviews. Their motive is always to find the best

Figure 3. – *CBT versus treatment as usual for depression, stratified by baseline severity. The overall pooled stimate is provided for comparison purpose (taken from Churchill et al., 2001).*

**Comparision: 05 Cognitive behavioural (CBT)+Cognitive (CT)+Behavioural (BT) therapies versus Treatment as usual (TAU)**
**Outcome:     08 Heterogeneity test - recovery - severity**

| Study | CBT+CT+BT n/N | Treatment as usual n/N | OR (95%CI Random) | Weight % | OR (95%CI Random) |
|---|---|---|---|---|---|
| **01 Major** | | | | | |
| Katon 1996 (S25) | 53/77 | 36/76 | | 15.6 | 2.45 [1.27,4.75] |
| Nezu 1989 (S34) | 22/29 | 1/13 | | 4.7 | 37.71[4.14,343.87] |
| Ross 1985 (S40) | 26/44 | 10/23 | | 11.9 | 1.88[0.68,5.21] |
| ScottA 1992 (S46) | 12/30 | 14/30 | | 11.8 | 0.76[0.27,2.12] |
| ScottC 1997 (S46) | 15/24 | 8/24 | | 10.4 | 3.33[1.02,10.90] |
| Teasdale 1984 (S54) | 14/24 | 4/20 | | 9.0 | 5.60[1.43,21.89] |
| Usaf 1990 (S58) | 4/31 | 2/29 | | 6.5 | 2.00[0.34,11.85] |
| Subtotal (95%CI) | 146/259 | 75/215 | | 69.9 | 2.64[1.37,5.09] |
| Test for heterogeneity chi-square=12.92 df=6 p=0.044 | | | | | |
| Test for overall effect z=2.90 p=0.004 | | | | | |
| | | | | | |
| **02 Mild/Moderate** | 25/44 | 14/45 | | 13.4 | 2.91[1.22,6.94] |
| Pace 1993 (S35) | 25/44 | 14/45 | | 13.4 | 2.91[1.22,6.94] |
| Subtotal (95%CI) | | | | | |
| Test for heterogeneity chi-square=0.0 df=0 | | | | | |
| Test for overall effect z=2.41 p=0.02 | | | | | |
| | | | | | |
| **03 Unspecified** | | | | | |
| Fuchs 1977 (S16) | 8/12 | 1/10 | | 4.2 | 18.00[1.65,196.32] |
| Nezu 1986 (S33) | 10/12 | 1/9 | | 3.7 | 40.00[3.05,524.86] |
| Selmi 1990 (S48) | 5/12 | 2/12 | | 5.9 | 3.57[0.53,23.95] |
| Shaw 1977 (S50) | 7/16 | 0/8 | | 2.9 | 13.42[0.66,271.90] |
| Subtotal (95%CI) | 30/52 | 4/39 | | 16.7 | 10.88[3.33,35.39] |
| Test for heterogeneity chi-square=2.49 df=3 p=0.48 | | | | | |
| Test for overall effect z=3.95 p=0.00008 | | | | | |
| | | | | | |
| Total (95%CI) | 201/355 | | | 100.0 | 3.42[1.98,5.93] |
| Test for heterogeneity chi-square=21.20 df=11 p=0.031 | | | | | |
| Test for overall effect z=4.39 p=0.00001 | | | | | |

.01   .1   1   10   100
Favours TAU        Favours CBT+CT+BT

evidence from research to guide practice where uncertainty about the best treatment exists. The Cochrane Collaboration operates globally in supporting the efforts of all those wishing to undertake such reviews.

# REFERENCES

Adams C.E., Power A., Frederick K. & Lefebvre C. (1994). An investigation of the adequacy of MEDLINE searches for randomized controlled trials (RCTs) of the effects of mental health care. *Psychological Medicine* 24, 741-748.

Antman E.M., Lau J., Kupelnick B., Mosteller F. & Chalmers T.C. (1992). A comparison of results of meta-analyses of randomized controlled trials and recommendations of clinical experts. *Journal of the American Medical Association* 268, 240-248.

Chalmers I. & Altman D.G. (1995). *Systematic Reviews.* BMJ: London.

Churchill R., Hunot V., Corney R., Knapp M., McGuire H., Tylee A. & Wessely S. (2001). A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. *Health Technology Assessment Monograph* 5 (35), 1-173.

Churchill R., Hunot V., Corney R., Knapp M., McGuire H., Tylee A. & Wessely S. (2003). Brief psychological therapies versus treatment as usual for depression (Protocol for a Cochrane Review). *Cochrane Library* Issue 1.

Clarke M. & Oxman A.D. (eds.) (2000). Cochrane Reviewers' Handbook 4.1 [updated June 2000]. In *Review Manager (RevMan)* Version 4.1. [Computer program]. The Cochrane Collaboration: Oxford.

Davey Smith G., Song F. & Sheldon T.A. (1993). Cholesterol lowering and mortality: the importance of considering initial levels of risk.

Figure 4. – *CBT versus treatment as usual for depression, stratified by source of recruitment. Again, the overall pooled estimate is provided for comparison purposes (taken from Churchill et al. 2001).*

**Comparision: 05 Cognitive behavioural (CBT)+Cognitive (CT)+Behavioural (BT) therapies versus Treatment as usual (TAU)**
**Outcome:       11 Sensitivity - recovery - recruitment**

| Study | CBT+CT+BT n/N | Treatment as usual n/N | OR (95%CI Random) | Weight % | OR (95%CI Random) |
|---|---|---|---|---|---|
| **01 Out-patients/referrals/clinic attenders** | | | | | |
| Katon 1996 (S25) | 53/77 | 36/76 | | 15.6 | 2.45 [1.27,4.75] |
| Ross 1989 (S40) | 26/44 | 10/23 | | 11.9 | 1.88 [0.68,5.21] |
| ScottA 1992 (S45) | 12/30 | 14/30 | | 11.8 | 0.76[0.27,2.12] |
| ScottC 1997 (S46) | 15/24 | 8/24 | | 10.4 | 3.33[1.02,10.90] |
| Shaw 1977 (S50) | 7/16 | 0/8 | | 2.9 | 13.42[0.66,271.90] |
| Teasdale 1984 (S54) | 14/24 | 4/20 | | 9.0 | 5.60[1.43,21.89] |
| Subtotal (95%CI) | 127/215 | 72/181 | | 61.6 | 2.31[1.29,4.14] |
| Test for heterogeneity chi-square=7.99 df=5 p=0.16 | | | | | |
| Test for overall effect z=2.83 p=0.005 | | | | | |
| | | | | | |
| **02 Ad responders/volunteers/combination** | | | | | |
| Fuchs 1977 (S16) | 8/12 | 1/10 | | 4.2 | 18.00[1.65,196.32] |
| Nezu 1986 (S33) | 10/12 | 1/9 | | 3.7 | 40.00[3.05,524.86] |
| Nezu 1989 (S34) | 22/29 | 1/13 | | 4.7 | 37.71[4.14,343.87] |
| Pace 1993 (S35) | 25/44 | 14/45 | | 13.4 | 2.91[1.22,6.94] |
| Selmi 1990 (S48) | 5/12 | 2/12 | | 5.9 | 3.57[0.53,23.95] |
| Usaf 1900 (S58) | 4/31 | 2/29 | | 6.5 | 2.00[0.34,11.85] |
| Subtotal (95%CI) | 74/140 | 21/118 | | 38.4 | 6.67[2.40,18.52] |
| Test for heterogeneity chi-square=9.54 df=5 p=0.09 | | | | | |
| Test for overall effect z=3.64 p=0.0003 | | | | | |
| | | | | | |
| | | | | | 3.42[1.98,5.93] |
| Total (95%CI) | 201/235 | 93/299 | | 100.0 | |
| Test for heterogeneity chi-square=21.20 df=11 p=0.031 | | | | | |
| Test for overall effect z=4.39 p=0.00001 | | | | | |

.01            .1            1            10            100

Favours TAU            Favours CBT+CT+BT

*British Medical Journal* 306, 1367-1373.

Dickersin K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, 263, 1385-1339.

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

Egger M., Davey Smith G. & Altman D.G. (2001). *Systematic Reviews in Healthcare: Meta-analysis in Context.* BMJ: London.

Hotopf M., Lewis G. & Normand C. (1997). Putting trials on trial: the costs and consequences of small trials in depression: a systematic review of methodology. *Journal of Epidemiology and Community Health* 51, 354-358.

Jüni P., Witschi A., Bloch R. & Egger M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association* 282,1054-1060.

Lewis G., Churchill R. & Hotopf M. (1997). Systematic reviews and meta-analysis. *Psychological Medicine* 27, 3-7.

Linde K. & Mulrow C.D. (2003). St John's wort for depression (Cochrane Review). *Cochrane Library* Issue 1.

Mulrow C.D. (1994). Rationale for systematic reviews. *British Medical Journal* 309 (6954), 597-599.

Pittler M.H. & Ernst E. (2003). Kava extract for treating anxienty. (Cochrane Review). *Cochrane Library* Issue 1.

Rose S. & Bisson J., Wessely S. (2003). Psychological debriefing for preventing post traumatic stress disorder (PTSD) (Cochrane Review). *Cochrane Library* Issue 1.

Sterling T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association* 54, 30-34.

Sterling T.D., Rosenbaum W.L. & Weinkam J.J. (1995). Publication decisions revisited: the effect of outcome of statistical tests on the decision to publish and vice versa. *American Statistician* 49, 108-112.

Stern M.J. & Simes R.J. (1997). Publications bias: evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal* 315, 640-645.

Thornley B. & Adams C. (1998). Content and quality of 2000 controlled trials in schizophrenia over 50 years. *British Medical Journal* 317, 1181-1184.