

The Cochrane Schizophrenia Group Guide Review

This section contains general information on the CSG editorial process and style guidelines. The Guide Review in full follows below.

1.1 The process by which an idea evolves into a review

1.1.2 Titles

After consultation with a Key Editor (the one with primary responsibility for supporting the authors) titles are submitted to the Review Group Co-ordinator and recorded within the Group module - this avoids duplication of effort.

1.1.3 Protocols

One editor provides ongoing support and another supplies peer review in the role of Anonymous Editor. All protocols are also sent for peer review external to the CSG. Reviewers are welcome to seek further expert peer review. Once available on the Cochrane Library comments are welcomed from readers of the Library and these act as a further layer of peer review. Comments should be directed through the Correspondence Manager of the Cochrane Library.

1.1.4 Reviews

All reviews are sent for peer review external to the CSG. Reviewers are welcome to seek further expert peer review. Again, once available on the Cochrane Library comments are welcomed from anyone and are submitted through the Correspondence Manager of the Cochrane Library.

1.1.5 Turn-around time

The CSG hopes to get work processed quickly and efficiently; we aim to have protocols and reviews ready for publication six weeks from the date that the reviewer submits their draft to the Editorial Base. This can take longer or shorter, depending on the volume of work ongoing at the time and the amount of editorial work required for each piece of work.

If you find it useful you are encouraged to cut and paste from this text so your work conforms to the Cochrane style guidelines.

Each section will follow the same structure: explanation/information/background, followed by bulleted style points (if any), followed by examples from reviews as necessary.

General Style Guide

Write clearly: remember at all times that what you write must be accessible to every possible type of reader - clinician, manager, lay, not-English-as-first-language, WWW browser...

Tense: It is Cochrane policy to write protocols in the future tense and reviews in the past tense. However we as a group do not enforce this rule as it inevitably creates more work when writing the review which must be written in the past tense. If you decide to write in the present tense, bear in mind that this will eventually result in more work for you and us at the editorial base!

Please, if at all possible, do not use phrases anywhere in the text such as 'patients', 'subjects', and 'schizophrenics'. Preferred CSG style is 'people with schizophrenia/acute mental illness/other condition'.

References in the text. Really simple - first author - or study tag when referring to an included or excluded trials - and year - even when there are only two authors - and multiple references separated by a comma, in order of year of publication. When from the same year, list alphabetically. EXAMPLE (Smith 1994, Abrams 1997, Patel 1997).

Do not put spaces on either side of the 'stroke' key (/); correct form is risks/benefits not risks / benefits

Hyphenation: please avoid hyphenation if possible. Rather than '**side-effects**' it is correct to say 'side effects' and it is probably more accurate to write 'adverse effects' in any case. The same applies to 'follow-up'. There is no need for hyphenation, 'follow up' does just as well.

Avoid use of the passive voice, ie 'The reviewers/we extracted data' rather than 'Data were extracted by the reviewers.'

Double quotes/single quotes: only use double quotation marks (") when directly quoting what someone has said or written. Use single quotation marks (') at all other times.

Guide Review

This Guide Review will be updated as with any other protocol or review: this is the 2004 updated version. Any suggestions will be welcomed, and the examples provided are here to guide your presentation so that it complies as near as possible to the Cochrane style.

Synopsis

Explanation:

The synopsis is a brief summary of the results of the review in plain language for consumers and non-specialist readers. The synopsis does not replace the abstract but is an additional product. It will be published as part of the Cochrane Review in the Cochrane Database of Systematic Reviews. The synopsis should enhance the accessibility of the review, disseminate its findings to a wide community internationally and act as an aid to browsing in The Cochrane Library.

Style:

According to the official Cochrane format the synopsis should be in two parts: a short, single sentence 'headline' of up to 25 words (in lower case apart from the first letter of the first word); followed by a single paragraph summary of the context and findings of the review (50 to 100 words). These should be separated by a blank line. In order to keep the 'headline' short, some abbreviations and technical terms may be inevitable. This format is not mandatory however and as you will see in the following examples a single paragraph will suffice if this is your preference.

See following link for further information on how to write your synopsis:

<http://www.cochrane.dk/cochrane/handbook/hbook.htm> section 2a

EXAMPLE 1: *Gamma-aminobutyric acid agonists for neuroleptic-induced tardive dyskinesia*

Taking antipsychotic drugs for long periods of time can cause repetitive movements that often occur in the face and mouth. These are disfiguring and do not necessarily cease once medication is reduced or changed. In this review we consider one group of drugs that has been evaluated for treating these movement disorders: the so-called 'GABA agonist' group of drugs. We found no convincing evidence regarding its efficacy in every day practice and there are some data to suggest that adverse effects resulting from its use may in fact prohibit clinical use.

EXAMPLE 2: *Hypnosis for schizophrenia*

Hypnosis is a "procedure during which a health professional or researcher suggests that a client, patient or subject experience changes in sensations, perceptions, thoughts or behaviour". Hypnosis is used by some healthcare professionals to improve symptoms for people with schizophrenia and we identified three small, poorly reported and inconclusive studies on this method of treatment. However, to ascertain whether hypnosis can be beneficial to people with schizophrenia clearly requires better designed, conducted and reported randomised studies.

Abstract

Explanation:

All full reviews must include an abstract of preferably not more than 400 words. It should be kept as brief as possible without sacrificing important content. Abstracts are made freely accessible on the Internet and will often be read as stand-alone documents. They should, therefore, summarise the key methods and content of the review and not contain any material that is not in the review.

Style:

Abstracts should be made as legible as possible without compromising scientific integrity. They should primarily be targeted to healthcare decision makers (clinicians, consumers and policy makers) rather than just researchers. Terminology should be reasonably comprehensible to a general rather than a specialist medical audience.

Background

Explanation:

This should be one or two sentences to explain the context or elaborate on the purpose and rationale of the review.

EXAMPLE 1: *Gamma-aminobutyric acid agonists for neuroleptic-induced tardive dyskinesia*

Chronic antipsychotic drug treatment may cause tardive dyskinesia (TD), a long-term movement disorder. Gamma-aminobutyric acid (GABA) agonist drugs, which have intense sedative properties and may exacerbate psychotic symptoms, have been used to treat TD.

EXAMPLE 2: *Clotiapine*

Acute psychotic illnesses, especially when associated with agitated or violent behaviour, require urgent pharmacological tranquillisation or sedation. Clotiapine, a dibenzothiazepine neuroleptic, is being used for this purpose in several countries.

Objectives

Explanation:

This should be a precise statement of the primary objective of the review, ideally in a single sentence

Style:

Where possible the style should be of the form 'To assess the effects of [*intervention or comparison*] for [*health problem*] for/in [*types of people, disease or problem and setting if specified*]'.

EXAMPLE 1: *Clotiapine for schizophrenia*

To estimate the effects of clotiapine when compared to other 'standard' or 'non-standard' treatments for acute psychotic illnesses in controlling disturbed behaviour and reducing psychotic symptoms.

EXAMPLE 2: *Droperidol for schizophrenia*

To estimate the effects of droperidol, including its cost-effectiveness, when compared to placebo, other 'standard' or 'non-standard' treatments, or other forms of management of psychotic illness, in controlling acutely disturbed behaviour and reducing psychotic symptoms in those with schizophrenia like illnesses.

Search strategy***Explanation:***

This section should list sources and dates of the last search using the active form 'We searched ...' or, if there is only one reviewer, the passive form can be used, e.g. 'Databases X, Y, Z were searched'.

Style:

If the CRG's specialised register was used, this should be listed first in the form 'Cochrane X Group specialised register'. The order for listing any other databases should be the Cochrane Central Register of Controlled Trials, MEDLINE, EMBASE, other databases. The date range of the search for each database should be given. Make sure that what is described here coincides with what is listed in the main Search Strategy!

(* Please remember PsycLIT, not Psychlit, and PsycINFO, not Psychinfo.)

EXAMPLE: *Hypnosis for schizophrenia*

We searched the Cochrane Schizophrenia Group's Register (January 2003), contacted the Cochrane Complementary Medicine Field for additional searching, handsearched references of included or excluded studies and made personal contact with authors of relevant trials.

Selection criteria***Explanation:***

This section should clearly indicate how studies are chosen.

Style:

This should be given as '[*type of study*] of [*type of intervention or comparison*] in [*disease, problem or type of people*]. The outcomes should only be included if the review was restricted to specific outcomes.

EXAMPLE: *Clotiapine for schizophrenia*

The review included randomised clinical trials comparing clotiapine with any other treatment for people with acute psychotic illnesses.

Data collection & analysis

Explanation:

This should be restricted to how data were extracted and assessed, and not include details of what data were extracted. This section should cover whether extraction and quality assessment of studies were done by more than one person. If the reviewers contacted investigators to obtain missing information, this should be noted here. What steps, if any, were taken to identify adverse effects should also be noted.

Style:

The example below should be closely followed and adapted as necessary to the particular needs of the review.

EXAMPLE: *Gamma-aminobutyric acid agonists for neuroleptic-induced tardive dyskinesia*

Working independently, we selected and critically appraised studies, extracted data and analysed on an intention-to-treat basis. Where possible and appropriate we calculated risk ratios (RR) and their 95% confidence intervals (CI) with the number needed to treat (NNT). For continuous data Weighted Mean Differences (WMD) were calculated.

Main results

Explanation:

This section should address the primary objective and be restricted to the main qualitative and quantitative results (generally including not more than six key results). The outcomes included should be selected on the basis of those most likely to help someone making a decision about whether or not to use a particular intervention.

Style:

This section should begin with the total number of trials and participants included in the review, and brief details pertinent to the interpretation of the results (e.g. the quality of the studies overall or a comment on the comparability of the studies, if appropriate).

EXAMPLE: *Gamma-aminobutyric acid agonists for neuroleptic-induced tardive dyskinesia*

We identified eight small poorly reported studies for inclusion. For the outcome of 'no clinically important improvement in tardive dyskinesia' GABA agonist drugs were not clearly better than placebo (n = 108, RR 0.83 CI 0.6 to 1.1). Deterioration in mental state was more likely to occur in people receiving GABA medication (n = 95, RR 2.47 CI 1.1 to 5.4), but this effect was influenced by the decision to assign a negative outcome to those who dropped out before the end of the study. A greater proportion of people allocated GABA medication may fail to complete the trial compared with those allocated placebo (20% versus 9%), but this difference was not statistically significant (n = 136, RR 1.99 CI 0.8 to 4.7). There is a suggestion of an increase in ataxia (loss of power of muscular coordination) for both baclofen and sodium valproate (n = 95, RR 3.26 CI 0.4 to 30.2), and in sedation (n = 113, RR 2.12 CI 0.8 to 5.4) compared with placebo, but this was not significant. Withdrawal of tetrahydroisoxazopyridine (THIP) may cause seizures.

Reviewers' conclusions

Explanation:

This section should present information rather than offer advice. It should be drawn directly from the review's findings and correlate to the main results. Assumptions should not be made about practice circumstances, values, preferences, tradeoffs; and the giving of advice or recommendations should generally be avoided. Any important limitations of data and analyses should be noted. Important conclusions about the implications for research should be included if these are not obvious.

Style:

Please be as concise and to the point as possible.

EXAMPLE 1: *Droperidol for schizophrenia*

This is an important, and surprisingly under-researched, area. To date, use of droperidol for emergency situations has been justified by experience rather than evidence from well conducted and reported randomised trials, but, as world reserves diminish, droperidol will no longer be a treatment option.

EXAMPLE 2: *Clotiapine for schizophrenia*

We found no evidence to support the use of clotiapine in preference to other 'standard' or 'nonstandard' treatments for management of people with acute psychotic illness. All trials in this review have important methodological problems. We do not wish to discourage clinicians from using clotiapine in the psychiatric emergency, but well-designed, conducted and reported trials are needed to properly determine the efficacy of this drug.

Background

Explanation:

The first part of the main Background should summarise your perspective on the care of those with schizophrenia, with descriptions of the intervention on which you mean to focus, for example the drug, psychological intervention, symptom or behaviour.

The next part of the Background should explain why the intervention your review focuses on is important. (Please use good epidemiology for example, how prevalent is the treatment, is it a new experimental intervention or one that is 'standard care' the world over?) Please do not forget the perspective of those from low to middle income countries. (For example, new drugs may be prevalent for those in high income countries, but 80% of people with schizophrenia do not live in such countries.)

Finally, introduce controversies you may wish to investigate in a sensitivity analysis. For example, is there an argument that those in their first episode of schizophrenia will have a differential response to the intervention of interest than those late on in their illness, or that women respond differently than men? If so, this could be included here so that when the sensitivity analysis is outlined later in the text, the reader understands why it is of interest.

Style:

The background should be short and to the point. The Background should not include detailed descriptions of schizophrenia, its causes or prognosis. Such information is contained within the CSG Module and can be referred to, but not repeated without good reason. Please have a look at the description that is in the CSG Module on the Cochrane Library and see if that is satisfactory. If you disagree or have any comments/suggestions do please let us know. You will, however, need to let the reader know in which aspect of schizophrenia/chronic mental illness you are interested and present descriptions and good epidemiology relating to that.

If you are intent on describing the receptor blockade of a drug or the psychological theory of why one of the talking therapies are thought to work but will be unable to do that in the type of background that could readily be read by the educated lay reader, please add a short paragraph at the end, entitled 'Technical background'. This avoids the lay reader being excluded from the main discussion.

EXAMPLE: *Droperidol for schizophrenia*

Violent or acutely disturbed people pose a risk to themselves and to others, as well as a diagnostic dilemma (Thomas 1992). Ideally, in order to ensure a safe and therapeutic environment, attempts should be made to calm the patient either through verbal de-escalation or intensive nursing techniques. Behaviour may frequently be too disturbed or agitated for 'verbal tranquillisation' to be effective, and further action, in the form of rapid tranquillisation, may be necessary.

Various drug regimens are used in such emergency situations, and clinical practice differs. One survey from the USA (Binder 1999) found that the Medical Directors of 20 Emergency Rooms preferred drug management for aggressive people to be a haloperidol-lorazepam mixture (Table 01). In 1993, a similar survey of clinicians' preferences in the UK found that chlorpromazine was the most common choice (Cunnane 1994). Another survey of Emergency Rooms in Rio de Janeiro (Huf 2002) found that a haloperidol-promethazine mixture was

commonly used for emergency intramuscular (im) sedation of severely agitated/aggressive people (70-100 people with suspected psychotic illness/week/3.5millionTable 02). A survey of frequency of emergency prescribing in a General Psychiatric Hospital in South London showed that rapid medical tranquillisation was required 102 times in 160 days (Pilowsky 1992). Eight different drugs were used (Table 03), amongst which diazepam, haloperidol and droperidol were used most often.

Droperidol (marketed as Dehydrobenzperidol, Dridol, Droleptan, Droperidols, Inapsin, Inapsine, Leptanal comp, Leptofen, Paxical or Sintodian) has been widely used in Europe by psychiatrists since the 1960s for treating acute or chronic psychoses (Cocito 1970, Resnick 1984). It inhibits the effects of dopamine. In the United States it is used primarily in conjunction with anaesthetics because of its sedative and antiemetic properties (Resnick 1984). Reported advantages of droperidol over haloperidol (another inhibitor of the effects of dopamine) include: a faster onset of action when given intramuscularly, swifter elimination from the body and fewer adverse effects (Richards 1998). The most commonly reported side effects for droperidol include hypotension (abnormally low blood pressure) and tachycardia (above normal heart rate). Other adverse effects include restlessness, hyperactivity, anxiety and dysphoria (feeling ill at ease). The frequency of adverse effects involving movement disorders is reported to be 20-40% (Cocito 1970). Droperidol has been associated only rarely with serious adverse effects such as neuroleptic malignant syndrome (altered consciousness, muscle rigidity and autonomic instability) and sudden death. The latter has been reported to be associated with high doses of droperidol (25 mg or more) in people at risk for cardiac dysrhythmia, such as those with severe electrolyte disturbances or alcohol withdrawal (RxList 2000). Droperidol should not be given to people with severe depression as it may aggravate their symptoms (Martindale 1982).

Following an extensive risk-benefit assessment requested by the Medicines Control Agency, Janssen-Cilag, the pharmaceutical company who market droperidol, concluded that the oral form of droperidol should be discontinued and that the injectable form would no longer be commercially viable. The Medical Director of Janssen-Cilag told PharmaTimes (<http://www.pharmatimes.co.uk/>) that the decision had been taken because many patients who receive droperidol also receive other medications that extend QT prolongation, and are more likely to have background illnesses that may exacerbate the problem. He added that the company intended to implement a world-wide withdrawal of droperidol, and supplies would stop entering the distribution chain at the end of March 2001.

Technical Background

Droperidol, 1-(1-3-(p-fluorobenzoyl)propyl-1,2,3,6-tetrahydro-4-pyridyl)-2-benzimidazolinone, is a butyrophenone neuroleptic. Butyrophenones inhibit the effects of dopamine and resemble phenothiazines such as trifluoperazine. They have fewer sedative and antimuscarinic effects than other phenothiazine derived anti-psychotic drugs, but exhibit more pronounced side effects upon the extrapyramidal nerve system. As of March 2000, the cost of medication with droperidol (droleptan) in Great Britain was £0.90 for 2 millilitre amp injection, or £0.25 for a 10 mg tablet (BNF 2000). Droperidol may be taken orally (5-20 mg repeated every 4-6 hours, as necessary) or as an im or iv injection (dosages: up to 10 mg repeated every 4-6 hours; and, 5-15 mg repeated every 4-6 hours, respectively). The onset of action from injection is three to 10 minutes, although the peak effect may not be apparent for 30 minutes. The duration of sedation and tranquillisation may last for two to four hours, although alteration of alertness may persist for as long as 12 hours. Children can be given 0.5-1 mg of oral or im dose daily (RxList 2000).

Objectives

Explanation:

This is a precise statement of the review's primary objective, and should include the interventions reviewed and the targeted problem. Sensitivity analyses that you have introduced in your Background should also be listed here. It might also mention why this review was undertaken and how it might relate to a wider review of a general problem.

Style:

This should be a very short statement. Do not say "to review the effectiveness of. . . ." as this sounds as if the focus of the review is thought to be of value. "To review the effect(s) of. . . ." is more appropriate.

EXAMPLE 1: *Droperidol for schizophrenia*

To estimate the effects of droperidol, including its cost-effectiveness, when compared to placebo, other 'standard' or 'non-standard' treatments, or other forms of management of psychotic illness, in controlling acutely disturbed behaviour and reducing psychotic symptoms in those with schizophrenia like illnesses.

EXAMPLE 2: *Hypnosis for schizophrenia*

To investigate the use of hypnosis for people with schizophrenia or schizophrenia-like illnesses compared to standard care and other interventions.

Criteria for considering studies for this review

Types of studies

Explanation:

This section states only the methods used in the studies that you hope to find. Interventions, participants and outcomes have their own sections below.

Style:

See example below. The first sentence may be all that is required.

EXAMPLE: *Gamma-aminobutyric acid agonists for neuroleptic-induced tardive dyskinesia*

We included all relevant randomised controlled trials. We included trials that were described as double-blind, but that did not mention whether the study was randomised, in a sensitivity analysis. If there was no substantive difference within primary outcomes (see 'Types of outcome measures') when these studies were added, then we included them in the final analysis. If there was a substantive difference, we used only clearly randomised trials and described the results of the sensitivity analysis in the text. We excluded quasi-randomised studies, such as those allocating by using alternate days of the week.

Types of participants

Explanation:

This should clearly describe the participant group in which you are interested. Try to be pragmatic. If you say that you will only include data from people with DSM-IV schizophrenia then you will lose lots of valuable data (DSM-III was not that bad, was it?) and make your work difficult to generalise.

Style:

See examples below. If a detailed explanation is required make it reader friendly by paying close attention to the style of listing, capitalising and punctuation indicated in Example 2.

EXAMPLE 1: *Amphetamines for schizophrenia*

Adults (18+ years) with schizophrenia. To account for changing definitions in diagnostic criteria, when study participants were described as suffering from 'severe/chronic mental illness/disorder' they met the inclusion criteria. We did not include studies involving individuals with bipolar disorder in the review.

EXAMPLE 2: *Gamma-aminobutyric acid agonists for neuroleptic-induced tardive dyskinesia*

People with schizophrenia or any other chronic mental illness, diagnosed by any criteria, irrespective of gender, age or nationality who:

1. required the use of antipsychotics for more than three months;
2. developed tardive dyskinesia (diagnosed by any criteria at baseline and at least one other occasion) during antipsychotic treatment;
3. for whom the dose of antipsychotic medication had been stable for one month or more (the same applies for those free of antipsychotics).

Types of interventions

Explanation:

This should list and describe all interventions of interest, including comparison interventions. This may, or may not include details of dose, frequency and mode of administration. In studies on antipsychotic drugs a problem can arise if several doses of the new compound are compared with one dose of the conventional antipsychotic. There are two possible solutions for this problem:

- i) Only one dose of the new antipsychotic (or other intervention) is used in the comparison. The selection of the dose should then be made according to rational considerations. For example, clearly sub-therapeutic doses are sometimes used for the purpose of a quasi-placebo control condition. These should then not be used as the comparator dose. It is also important that the selection of the dose should be undertaken blind, without knowledge of the results.
- ii) All patients on the different doses of the new antipsychotic are considered as one group. This is easy to do for dichotomous outcomes, but more complicated to estimate for continuous outcomes, if individual patient data are not available. A statistician or the contact editor should be contacted for advice.

Style:

These should be listed in order of interest, with the experimental intervention first being compared to the following interventions:

EXAMPLE 1: *Clotiapine for schizophrenia*

1. Clotiapine: any dose, given orally, or by intramuscular injection against one or more of the following:
2. Standard medication: drug treatments that fit with normal 'custom and practice'. This may involve increasing the dose of standard medication or the addition of another 'standard' psychotropic drug, such as an antipsychotic, an anxiolytic (benzodiazepine or other) or a mood stabiliser.
3. Non-standard medication: drug treatments which are being trialled as a new type of intervention.
4. Placebo.

EXAMPLE 2: *Cognitive behavioural therapy for schizophrenia (a far more complex example)*

1. Cognitive behavioural therapy

The label cognitive behavioural therapy has been applied to a variety of interventions, accordingly, is difficult to provide a single, unambiguous definition. Recognising this, the reviewers constructed criteria that were felt to be both workable and to capture the elements of good practice in CBT.

In order to be classified as 'well defined' the intervention must clearly demonstrate the following components:

- i. the intervention involves the recipient establishing links between their thoughts, feelings and actions with respect to the target symptom; and
- ii. the intervention involves the correction of the person's misperceptions, irrational beliefs and reasoning biases related to the target symptom.
- iii. the intervention should involve either or both of the following: i. the recipient monitoring his or her own thoughts, feelings and behaviours with respect to the target symptom; and ii. the promotion of alternative ways of coping with the target symptom.

In addition, all therapies that do not meet these criteria (or that provide insufficient information) but are labelled as 'CBT' or 'Cognitive Therapy' were included as 'less well defined' CBT. A sensitivity analysis on the primary outcomes of this review (see types of outcomes) was conducted in order to investigate if this hierarchy of definition made any difference.

2. Standard care

The care that a person would normally receive had they not been included in the research trial. The category 'standard care' also incorporates 'waiting list control groups' where participants receive drug or other interventions.

3. Specific medication

Where the control group is given a specific drug in comparison to CBT.

4. Non-intervention

Untreated control groups.

5. Additional drug interventions

Where standard care has been supplemented by additional medication.

6. Other psychosocial interventions

Where standard care has been supplemented by additional psychological and/or social interventions, such as non-directive counselling and supportive therapy and other 'talking therapies'.

Types of outcome measures

Explanation:

This should list what you think the outcomes *should be*, rather than what you think is there, in the trials. For example, you may think that it is crucial that economic outcomes are noted in a systematic review of a new treatment. You may know the single study for this new antipsychotic and that there are no data on this outcome. By listing it in 'Types of outcome measures' you highlight the need for it and set a research agenda for the future.

Don't forget outcomes like death, or leaving the study early. If applicable, reviewers usually pre-state what they would consider short, medium and long-term outcomes.

Style:

These should be listed in order of importance. You may also want to designate both primary and secondary outcomes. Each numbered outcome should be treated as one sentence, and should follow the punctuation and capitalisation style laid out below.

EXAMPLE 1: *Crisis intervention for people with severe mental illness*

Five main outcomes were considered.

1. Service utilisation
 - 1.1 Admission to hospital
 - 1.2 Number of days in hospital and
 - 1.3 Number of staff/user contacts.

2. Satisfaction with treatment
 - 2.1 Number of people leaving the study early
 - 2.2 Patient satisfaction
 - 2.3 Staff satisfaction and
 - 2.4 Carer satisfaction.

3. Clinical outcome
 - 3.1 Death/suicide
 - 3.2 Improvement, general or specific
 - 3.3 Compliance with medication
 - 3.4 Antipsychotic medication and
 - 3.5 Relapses.

4. Social outcome
 - 4.1 Social functioning including life skills
 - 4.2 Employed (paid/voluntary/attendance at school/college)
 - 4.3 Able to live independently and
 - 4.4 Number of carers - professional or significant others - needed to maintain stable state.

5. Cost of treatment
 - 5.1 Total, mental health care or medical care costs
 - 5.2 Staff input - hours worked and
 - 5.3 Carer input - change in lifestyle/no change in lifestyle/loss of income.

We selected outcome measures which provided global estimations of functioning. Highly specific outcomes, such as, for example, 'sense of safety' were not reported. Such specific outcomes are rarely reported in more than one study and it is difficult to assess their relevance to the effectiveness of the treatment. Other outcomes not readily falling into these categories were also recorded but were not of pre-stated interest.

We divided outcomes into short term (less than six weeks) medium term (six weeks-three months) and long term (more than three months).

EXAMPLE 2: *Gamma-aminobutyric acid for neuroleptic-induced tardive dyskinesia*

We defined clinical efficacy as an improvement in the symptoms of TD of more than 50%, on any scale.

1. Tardive dyskinesia changes

1.1 Not showing an improvement in the symptoms of individuals of more than 50% on any TD scale

1.2 Not showing any improvement in the symptoms of individuals on any TD scale, as opposed to some improvement

1.3 Deterioration in the symptoms of individuals, defined as any deleterious change on any TD scale

1.4 Any adverse effect, other than deterioration of symptoms of TD, as reported in the trials

1.5 Average change in severity of TD during the trial period

1.6 Average difference in severity of TD at the end of the trial

2. General mental state changes

2.1 Deterioration in general psychiatric symptoms (such as delusions and hallucinations) defined as any deleterious change on any scale

2.2 Average difference in severity of psychiatric symptoms at the end of the trial

3. Acceptability of the treatment

3.1 Acceptability of the intervention to the participant group as measured by numbers of people dropping out during the trial

When appropriate, we grouped the outcomes into time periods - short term (less than six weeks), medium term (between six weeks and six months) and long term (more than six months).

Search strategy for identification of studies

Explanation:

1. Electronic searching

1.1 Prior to undertaking any searches please liaise with the CSG's Trial Search Co-ordinator (TSC) Judy Wright: jwright@cochrane-sz.org who will discuss databases that can be searched on your behalf and will help you develop search strategies.

1.2 Databases already searched

The Cochrane Schizophrenia Group undertakes searches of a number of databases, conference proceedings and extracted citations of all trials relating to schizophrenia or chronic/serious mental illnesses. Regular searches of Biological Abstracts CINAHL, the Cochrane Library, EMBASE, MEDLINE and PsycLIT are undertaken by the TSC, and relevant citations are added to an in-house register. If the scope of your review includes the treatment of people with schizophrenia or serious mental illness, it is not necessary for you to search these databases. We have also created a CD Rom of all papers searched which can be borrowed on request.

1.3 Additional database searching

In addition to the databases listed above, the Cochrane Schizophrenia Group has access to Sociological Abstracts, LILACS, Social Science Citation Index and Science Citation Index. Where appropriate, the TSC can also search these databases on your behalf.

1.4 Searching the Internet

The World Wide Web can be a useful source of information for systematic reviews. However, caution must be exercised with regard to the authenticity of the sites visited and the data contained within. The TSC can help also help with these searches.

When recording the search you need to include the following:

1.4.1 The website address

1.4.2 The date the website was accessed

1.4.3 The name of the search engine (e.g. Altavista) or subject index (e.g. Yahoo) that was searched to find the data (if one was used)

1.4.4 The exact terms used to search the search engine including all search syntax (eg 'controlled trial AND thioridazine')

1.4.5 Any search limitations e.g. limited to English language material, limited to pages published after 1996.

1.5 Doing it yourself

You are welcome to search other databases, or the ones we have already undertaken, as long as you do so in full consultation with your Contact Editor. It is preferable, however, that searches are undertaken by the TSC. At the very least, all searches should be discussed with her. It is necessary to record the exact date every database is searched and every word or phrase used in the search strategy so it may be reproduced at a later date. If the search has been undertaken by the TSC, this information will be recorded in the correct format (see the example below) and sent to you.

1.6 Update searches

Once a review has been published, its search strategy will updated by the TSC when the review is due for an update and the reviewers will be notified of this.

2. Handsearching

After contacting the TSC, reviewers are encouraged to handsearch literature relevant to their topic. This would usually involve searching the full contents of each issue of a specialist journal relevant to the topic of your review for a specified number of years. The major psychiatric journals (e.g. British Journal of Psychiatry, Archives of General Psychiatry) are already being systematically handsearched by either the Cochrane Schizophrenia Group or another Cochrane entity for the mutual benefit of the Collaboration.

3. Additional search strategies

3.1 Citation searching

This can be undertaken either electronically or manually. The methods by which reports are selected for citation searching are made explicit.

3.2 Personal contact

The methods by which individuals are selected for contact are made explicit and reproducible. For example, all members of a specific interest group could be contacted on a specified date.

3.3 Pharmaceutical industry

The methods by which companies are selected for contact and, in turn, how they select material for the review are made explicit and reproducible. Janssen, for example, has been very helpful, but pharmaceutical companies should not be contacted directly - ONLY through the CSG editorial base. The CSG works to foster good relationships with pharmaceutical companies in an attempt to obtain the largest dataset possible. This has involved face-to-face contact with company officials who are able to provide unpublished information on file. Individual approaches may lead to rejection and to the perception that the CSGs' efforts are illcoordinated.

The following is an example of how the search strategy section might look in your review. It is included to illustrate the level of detail that is required when recording search strategies and also the standard format that should be used.

Style:

It is pretty well impossible to make this readable therefore it must, at the very least, be clear. Please follow the following example carefully. The TSC will help you write this and you can pass drafts back and forth to her.

EXAMPLE: Clotiapine for schizophrenia

1. Electronic searching for update

1.1. We searched The Cochrane Schizophrenia Group's study based register (April 2004) using the phrase:

[[(*clotiapin* OR *clothiapin* OR *etumin* OR *etomin *OR *2159*) in REFERENCE) and (clot* in STUDY)]

This register is compiled by systematic searches of major databases, hand searches and conference proceedings (see Group Module).

2. Electronic searching for the original version

2.1 The Cochrane Controlled Trials Register (CENTRAL, Issue 2, 2000) was searched May 2000, using the phrase:

[clotiapine or clothiapine or entumin]

2.2 We searched The Cochrane Schizophrenia Group Register (updated May 2000) using the phrase:

[clotiapine or clothiapine or clotiapina or etumine or entumine or entumin or etumina or etomine or "HF 2159" or "LW 2159"]

2.3 We searched EMBASE (1980-2000, May 2000) using the phrase:

[clotiapine or clothiapine or clotiapina or etumine or entumine or entumin or etumina or etomine or "HF 2159" or "LW 2159"]

2.4 We searched MEDLINE (1966-2000, May 2000) using the phrase:

[clotiapine or clothiapine or clotiapina or etumine or entumine or entumin or etumina or etomine or "HF 2159" or "LW 2159"]

2.5 We searched PASCAL (<http://services.inist.fr/public/fre/const2.htm>, 1973-2000, May 2000) through the Internet using the phrase:

[clotiapine or clothiapine or clotiapina or etumine or entumine or entumin or etumina or etomine or "HF 2159" or "LW 2159"]

2.6 We searched PsycLIT (1970-2000, May 2000) using the phrase:

[clotiapine or clothiapine or clotiapina or etumine or entumine or entumin or etumina or etomine or "HF 2159" or "LW 2159"]

3. Hand searching

We searched reference lists of included and excluded studies for additional relevant studies.

4. Requests for additional data

4.1 We contacted the Medical Information Centre of NOVARTIS, the company who markets clotiapine, for published and unpublished data on the drug. They sent us a search they had done in their database in April 2000, using the term 'Entumin'.

4.2 We contacted authors of relevant studies to inquire about other sources of relevant information.

Methods of the review

Explanation:

This should clearly describe how the methods were applied to the selection criteria. Every reviewer is encouraged to read the Cochrane Handbook (Clarke 2002), acquire training by attending workshops recommended by the Review Group Co-ordinator (RGC), and talk to the relevant editor.

1. Study selection

Selection of trials is encouraged to be performed independently and in parallel by at least two reviewers. The Group has no policy on whether those selecting studies should be blinded to author, institution and journal. Inter-rater agreement may be calculated using Cohen's kappa. Where disagreement occurs, this may be resolved by discussion or, where this is not possible, by contacting authors for further clarification.

2. Assessment of methodological quality

The CSG encourages the use of the quality rating based on the three quality categories as described in the Cochrane Collaboration Handbook (Clarke 2002). If additional scales are used please describe these.

3. Data collection

Data are usually independently extracted by at least two authors. Again, any disagreement can lead to discussion and all decisions are then documented and, where necessary, the authors of the studies contacted to help resolve issues. Summary data from published or unpublished reports are the most frequent source for this Group's reviews. Where individual patient data are sought every effort is made to ensure completeness.

4. Data synthesis

4.1 Incomplete data

For particular outcomes, where data is felt to be too prone to bias because of poor levels of follow up, reviewers PRE-STATE the proportion of study drop out at which the data is thought to be unacceptable. Reviewers should take into account that they may find studies of differing length, and that **drop-out** rates usually increase with study length. Reviewers may therefore pre-state different limits for acceptable levels of drop-out for studies of differing duration.

4.2 Dichotomous - yes/no - data

4.2.1 Statistics: For dichotomous (binary) outcomes, for example 'admitted' or 'not admitted', the pooled relative risk (RR), with the confidence interval (CI, usually 95% CI) is estimated. Usually the random effects model is applied as this is the most conservative. Your key editor will advise here. Be careful always to enter data on the unfavourable outcome because the statistics are based on calculation of risks. Where possible, the weighted number needed to treat statistic (NNT) is also calculated for a favourable effect and the number needed to harm (NNH), for a harmful effect. The confidence intervals of NNT and NNH should also be reported. A calculator can be downloaded from <http://www.mango3d.cwc.net/vsx.htm> although this does not calculate the confidence interval of the NNT - ask your contact editor for help. Please note that NNT/NNH varies with the underlying risk of the population and that your values will not be generalisable if the risk of your control group differs from the risk of this patient population in general (Smeeth 1999).

4.2.2 Intention to treat:

Where possible, data is presented on a 'once-randomised-always-analyse' basis. For outcomes, such as relapse, the authors assume certain outcomes for those lost to follow up. This is often the negative assumption, for example, those who were lost to follow up all relapsed. This is negative but even-handed. If you have access to survey data that provides more realistic estimates of prognosis that you wish to use, for example, that 50% of people that leave studies experience a relapse within a year, please reference these surveys and then apply their estimates. If you assume that all those who leave have the negative outcome, this may be far from the truth if this outcome is very rare, such as 'death' or 'agranulocytosis'. If you assume that all drop-outs experienced this rare adverse effect, your analysis will produce unrealistic results. Thus, your pre-stated assumption should vary depending on outcome. Many sideeffects are rare outcomes, so PRE-STATED realistic estimates, preferably referenced, should be made, if an assumption is to be added into the calculation of prevalence. For example, it could be assumed that of the people leaving the study of clozapine versus chlorpromazine 2% leaving the clozapine group would have agranulocytosis by 20 weeks and only 0.5% of those on chlorpromazine - and a reference to these numbers should be included. All assumptions are tested in a sensitivity analysis. Deaths are not counted as losses of contact and it would be extremely negative to assume that all losses to follow up were deaths.

4.2.3 Cluster trials:

Studies increasingly employ 'cluster randomisation' (such as randomisation by clinician or practice) but analysis and pooling of clustered data poses problems: Firstly, authors often fail to account for intra class correlation in clustered studies, leading to a 'unit of analysis' error (Divine 1992) whereby p values are spuriously low, confidence intervals unduly narrow and statistical significance overestimated. This causes type I errors (Bland 1997, Gulliford 1999). Secondly, RevMan does not currently support meta-analytic pooling of clustered dichotomous data, even when these are correctly analysed by the authors of primary studies, since the 'design effect' (a statistical correction for clustering) cannot be incorporated.

Where clustering was not accounted for in primary studies, you could present data in a table, with an (*) symbol to indicate the presence of a probable unit of analysis error. Subsequent versions of your review would seek to contact first authors of studies, to seek intra-class correlation co-efficients of their clustered data and to adjust for this by using accepted methods (Gulliford 1999). Where clustering has been incorporated into the analysis of primary studies, these data could also be presented in a table. **No further secondary analysis (including metaanalytic pooling) should be attempted until there is consensus on the best methods of doing so, and until RevMan, or any other software, allows this. A Cochrane Statistical Methods Workgroup is currently addressing this issue. In the interim, individual studies could be very crudely classified as positive or negative, according to whether a statistically significant result ($p < 0.05$) was obtained for the outcome in question, using an analytic method which allowed for clustering. ASK CLIVE**

4.3 Continuous - scale - data

4.3.1 Intention to treat: for continuous data this is difficult for reviewers, unless individual patient data are available. The reviewer has to use what data are reported, and many so called 'intention-to-treat analyses are either 'completer analysis' or involve data with considerable assumptions. For example the 'Last Observation Carried Forward' (LOCF) technique is often used to produce an 'Intention to treat' analysis. LOCF often assumes that the person is stable from that last observation to the end of the study. This is problematic. If you are in any confusion over this please contact your Editor. It is important not to forget that at some point there is such a great loss that it is not worth presenting data at all (see 4.1).

4.3.2 Normal data: Mental health continuous data are often not normally distributed. Most statistics assume a normal distribution. To avoid including non-normally distributed data in the statistical analysis, the following criteria are applied to all data before inclusion:

4.3.2.1 Standard deviations and means were reported or derivable from data in the paper, or were obtainable from the authors.

4.3.2.2 When a scale starts from zero, the standard deviation, when multiplied by two, is less than the mean (as otherwise the mean is unlikely to be an appropriate measure of the centre of the distribution, Altman 1996). Endpoint scores on scales often have a finite start and end point and this rule can be applied.

4.3.2.3 When continuous data are presented on a scale which includes a possibility of negative values (such as change on a scale) it is impossible to tell whether data is non-normally distributed (skewed) or not. It is thus preferable to use scale end point data, which typically cannot have negative values. If end point data is not available, reviewers may choose to use change data as the statistics used in Metaview are quite robust towards skewness.

4.3.2.4 If a scale starts from a positive value (such as PANSS which can have values from 30 to 210) the calculation described above in ii) should be modified to take the scale starting point into account. In these cases skewness is present if $2SD > (S - S_{min})$, where S is the mean score and S_{min} is the minimum score.

4.3.2.4. Data with an important degree of skew, as defined above, should be subject to sensitivity analyses. Skewed data should be excluded and then included to see if they make a substantive difference. If they do not then they should be included and this discussed. If they do make a substantive difference, ask your key editor for advice.

4.3.3 Rating scales: A wide range of instruments is available to measure mental health outcomes. These instruments vary in quality and many are not valid, or even ad hoc. For outcome instruments some minimum standards have to be set. They could be that: (a) the psychometric properties of the instrument should have been described in a peer-reviewed journal; (b) the instrument should either be: (i) a self-report, or (ii) completed by an independent rater or relative (not the therapist). Unpublished instruments are frequently employed and are more likely to report statistically significant findings than those that have been put into print (Marshall 2000). It is generally not advised to pool across different scales in order to produce effect sizes. If authors do feel that they wish to do this then they should talk to their Contact Editor.

5. Heterogeneity

5.1 How to find it: always look carefully at your data in case it is clearly odd. Does one study clearly sit apart from the others? If so, they are heterogeneous data, no matter what the statistics say. As well as inspecting the graphical presentations, differences between the results of each included trial are checked using a test of heterogeneity. The Mantel-Haenszel chisquare test of heterogeneity is automatically calculated by MetaView. The test is rather insensitive, and many reviewers PRE-define $p < 0.10$ as evidence of heterogeneity.

5.2 What to do about it: it is important not to add really heterogeneous studies together. Hopefully this will be avoided by use of your inclusion criteria but sometimes unusual studies slip through the net. If the results of one or two studies are clearly different from the rest, go

back to the original papers and see if an explanation for the heterogeneity can be found. If, on this further inspection a clear reason is found, data should not be pooled. If the reasons for the difference between study results are not a result of differing patients, interventions or outcomes, but seem to reflect a true variation in study results, data may be pooled using the random effects model, which takes into account the variation between studies. If this eliminates the heterogeneity, data should remain pooled and this discussed in the text. If, after using a random effects model, heterogeneity remains, data should not be pooled and the results presented separately and discussed. In any event, if heterogeneity is present, the reviewers undertake a sensitivity analysis to the presence or absence of these data.

6. Addressing publication bias

Data from all identified and selected trials can be entered into a funnel graph (trial effect versus trial size) in an attempt to investigate the likelihood of overt publication bias. The funnel graph is performed by the MetaView software, for inspection by the reviewers, but will not be published in The Cochrane Library.

7. Discussion and conclusions

Reviewers are encouraged to undertake the discussion and conclusions section viewing data from the perspectives of not only clinicians and researchers, but also policy makers, managers and consumers of care.

8. Tables and figures

Where possible, data are entered into RevMan in such a way that the area to the left of the line of no effect indicates a favourable outcome for the intervention of interest. Note that for some scales a high score indicates a bad outcome, but for other scales a high score indicates a good outcome. In order to get the WMD point estimates on the correct side you may have to enter the scale scores as negative values.

Style:

The outline below should be closely followed in preparing this section. We encourage you to 'cut and paste' and adapt the text as necessary to meet the needs and criteria of individual reviews. Please try and avoid lists of numbers that refer to several outcomes and end with the word 'respectively'. This is likely to confuse any reader. If we see this we end up 'derespectivising' and that is a lot of avoidable work. Try not to use absolutes. Instead of '*There are no studies of X versus Y*' it is better to write '*We did not find any studies of X versus Y*'.

When recording results such as a relative risk, please use the following format: (n = 456, RR 4.2 CI 1.2 to 6.7, NNT 5 CI 2 to 7). Note that there is no more than one figure after the decimal point - although we do realise that there are exceptions to this, and that NNTs and the confidence intervals are whole numbers (rounded to the nearest complete person).

EXAMPLE 1: *Clotiapine for schizophrenia*

1. Study selection

Simone Carpenter and Michael Berk independently inspected citations identified in the search. For the update, John Rathbone independently inspected and selected papers. Potentially relevant abstracts were identified and full papers ordered and reassessed for inclusion and methodological quality. Once we had obtained the full reports, we resolved disputes over whether studies met the inclusion criteria by discussion.

2. Assessment of quality

We allocated trials to three quality categories, as described in the Cochrane Reviewers' Handbook (Clarke 2003), by all reviewers working independently. When disputes arose as to which category a trial was allocated, we attempted to resolve them by discussion. When this was not possible and further information was necessary to clarify into which category to allocate the trial, data were not entered and the trial was allocated to the list of those awaiting assessment. We included trials only in Categories A or B in the review.

3. Data management

3.1 Data extraction

We extracted data independently and where further clarification was needed we contacted the authors of trials to provide us with the missing data.

3.2 Intention to treat analysis

We excluded data from studies where more than 50% of participants in any group were lost to follow up (this does not include the outcome of 'leaving the study early'). In studies with less than 50% dropout rate, people leaving early were considered to have had the negative outcome, except for the event of death. Regarding the outcomes of 'aggression', 'self harm' and 'harm to others', as they are major risks of non-treated acute psychotic illness, we considered 5% of the people leaving the study early to have had a negative outcome.

We analysed the impact of including studies with high attrition rates (25-50%) in a sensitivity analysis. If inclusion of data from this latter group did result in a substantive change in the estimate of effect, we did not add their data to trials with less attrition, but presented them separately.

4. Data analysis

4.1 Binary data

For binary outcomes we calculated a standard estimation of the random effects risk ratio (RR) and its 95% confidence interval (CI). We also calculated the weighted number needed to treat statistic (NNT), and its 95% confidence interval (CI). If heterogeneity was found (see section 5) we used a random effects model.

4.2 Continuous data

4.2.1 Skewed data: continuous data on clinical and social outcomes are often not normally distributed. To avoid the pitfall of applying parametric tests to non-parametric data, the following standards are applied to all data before inclusion: (a) standard deviations and means were reported in the paper or were obtainable from the authors; (b) when a scale started from the finite number zero, the standard deviation, when multiplied by two, was less than the mean (as otherwise the mean was unlikely to be an appropriate measure of the centre of the distribution, Altman 1996); (c) if a scale started from a positive value (such as PANSS which

can have values from 30 to 210) the calculation described above was modified to take the scale starting point into account. In these cases skew is present if $2SD > (S - S_{min})$, where S is the mean score and S_{min} is the minimum score. Endpoint scores on scales often have a finite start and end point and these rules can be applied to them. When continuous data are presented on a scale which includes a possibility of negative values (such as change on a scale), there is no way of telling whether data are non-normally distributed (skewed) or not. It is thus preferable to use scale end point data, which typically cannot have negative values. If end point data were not available, the reviewers used change data, but they were not subject to a meta-analysis, and were reported in the 'Additional data' tables.

4.2.2 Summary statistic: for continuous outcomes a weighted mean difference (WMD) between groups was estimated. Again, if heterogeneity was found (see section 5) we used a random effects model.

4.2.3 Valid scales: we included continuous data from rating scales only if the measuring instrument had been described in a peer-reviewed journal and the instrument was either a self report or completed by an independent rater or relative (not the therapist).

4.2.4 Endpoint versus change data: where possible we presented endpoint data. If both endpoint and change data were available for the same outcomes we reported only the former.

4.2.5 Cluster trials: studies increasingly employ 'cluster randomisation' (such as randomisation by clinician or practice) but analysis and pooling of clustered data poses problems. Firstly, authors often fail to account for intra class correlation in clustered studies, leading to a 'unit of analysis' error (Divine 1992) whereby p values are spuriously low, confidence intervals unduly narrow and statistical significance overestimated. This causes type I errors (Bland 1997; Gulliford 1999).

Where clustering was not accounted for in primary studies, we presented the data in a table, with a (*) symbol to indicate the presence of a probable unit of analysis error. In subsequent versions of this review we will seek to contact first authors of studies to obtain intra class correlation co-efficients of their clustered data and to adjust for this by using accepted methods (Gulliford 1999). Where clustering has been incorporated into the analysis of primary studies, we will also present these data as if from a non-cluster randomised study, but adjusted for the clustering effect.

We have sought statistical advice and have been advised that the binary data as presented in a report should be divided by a 'design effect'. This is calculated using the mean number of participants per cluster (m) and the intra-class correlation co-efficient (ICC) Design effect = $1 + (m - 1) * ICC$ (Donner 2002). If the ICC was not reported it was assumed to be 0.1 (Ukoumunne 1999).

If cluster studies had been appropriately analysed taking into account intra-class correlation coefficients and relevant data documented in the report, synthesis with other studies would have been possible using the generic inverse variance technique.

5. Investigation for heterogeneity

Firstly, we considered all the included studies within any comparison to judge clinical heterogeneity. Then we visually inspected graphs to investigate the possibility of statistical heterogeneity. This was supplemented, primarily, by employing the I-squared statistic. This provides an estimate of the percentage of inconsistency thought to be due to chance. Where the

I-squared estimate was greater than or equal to 75%, this was interpreted as evidence of high levels of heterogeneity (Higgins 2003). Data were then re-analysed using a random effects model to see if this made a substantial difference. If it did, and results became more consistent, i.e. falling below 75% in the estimate, the studies were added to the main body of trials. If using the random effects model did not make a difference and inconsistency remained high, data were not summated, but were presented separately and reasons for heterogeneity investigated.

6. Addressing publication bias

We entered data from all included studies into a funnel graph (trial effect against trial size) in an attempt to investigate the likelihood of overt publication bias (Egger 1997).

7. Sensitivity analyses

We analysed the effect of including studies with high attrition rates in a sensitivity analysis. Where a trial was described as 'double-blind' but implied that the study was randomised, it was also included in a sensitivity analysis.

8. General

Where possible, we entered data in such a way that the area to the left of the line of no effect indicated a favourable outcome for clotiapine.

EXAMPLE 2: Amphetamines for schizophrenia

1. Selection of trials

The two reviewers (Sue Nolte - SN, Gary Latchford - GL) independently performed the inspection of the citations identified in the search outlined above. Potentially relevant abstracts were identified and full papers ordered and reassessed for inclusion and methodological quality. Any disagreement was discussed and reported.

2. Assessment of quality

SN, GL and DW, working independently, allocated trials to three quality categories based on allocation concealment, as suggested in the Cochrane Collaboration Handbook (Clarke 2003). When disputes arose as to which category a trial was allocated, again, we attempted resolution by discussion. When this was not possible, and further information was necessary, we did not enter data into the analyses and assigned the study to the list of those awaiting assessment. We only included trials in Category A or B in the review.

We appraised methodological quality by examining the description of randomisation, blinding and participants leaving early as each of these factors reduces the risk of bias (Jadad 1996).

3. Data management

3.1 Data extraction

We independently extracted data from included studies and entered details as to why we chose not to put outcomes from included studies in the review in the 'included studies table.'

3.2 Intention to treat analysis

We excluded data on outcomes where more than 50% of participants in any group were lost to follow up. In studies with less than 50% dropout rate, people leaving early were considered to have had the negative outcome, except for the event of death or unless other reasons unrelated to interventions were clearly reported. The influence of including studies with high attrition

rates (25-50%) was analysed in a sensitivity analysis. If inclusion of data from this latter group did result in a substantive change in the estimate of effect, we did not add their data to trials with fewer dropouts, but presented them separately

4. Data analysis

4.1 Binary data

We calculated a standard fixed effect risk ratio (RR) and its 95% confidence interval (CI) for binary outcomes. Where possible we calculated the number needed to treat/harm statistic (NNT/H). If we found heterogeneity (see section 5) we employed a random effects model.

4.2 Continuous data

4.2.1 Skewed data

Continuous data on mental health outcomes are often skewed and not normally distributed. Review Manager (RevMan) analyses of continuous data are based on the assumption that the data are, at least to a reasonable extent, normally distributed. It would be inappropriate to apply parametric tests to non-parametric data. Consequently to identify non-parametric data and avoid this pitfall, we applied the following standards to all data before inclusion:

(a) Standard deviations and means were reported in the paper or were obtainable from the authors

(b) When a scale started from a finite number (such as zero), the standard deviation, when multiplied by two, was less than the mean (as otherwise the mean was unlikely to be an appropriate measure of the centre of the distribution - Altman 1996). Endpoint scores on scales often have a finite start and end point and this rule can be applied to them.

Change data (endpoint minus baseline) is problematic. With no individual patient data it is impossible, even after applying the above test, to know if change data are skewed. After consulting the ALLSTAT electronic statistics mailing list, we decided that change data could be entered into RevMan. However, in doing this it is assumed that either that data were not skewed or that the analyses could cope with the unknown degree of skew. Unfortunately, with no individual patient data it is impossible to test this assumption. Where both change and endpoint data were available for the same outcome category, only endpoint data are presented as it was believed that this type of data was less open to bias and closer to being interpreted clinically.

4.2.2 Summary statistic

For continuous outcomes we estimated a weighted mean difference (WMD) between groups. Again, if heterogeneity was found we used a random effects model.

4.2.3 Valid scales

A wide range of rating scales is available to measure outcomes in drug trials and these measures differ in validity. It is commonly acknowledged that measuring instruments should be both reliable and valid. Therefore, we included continuous data from rating scales only if the measuring instrument had (a) been described in a peer-reviewed journal (Marshall 2000) and (b) the instrument was either a self-report or completed by an independent rater or relative (not the researcher).

4.2.3.1 Publication

Before publication of an instrument, most scientific journals insist that reliability and validity are demonstrated to the satisfaction of referees. We decided, therefore, as a minimum standard, not to include any data from a measure in this review unless its properties had been published in a peer-reviewed journal (Marshall 2000).

4.2.3.2 Psychometric assessments

A wide range of psychometric assessments is also available. Again, these measures vary in quality and many are questionably validated, or even ad hoc. We evaluated the quality of psychometric tests included in this review using the 1997 revision of the Dutch Rating System for Test Quality employed by the Committee of Test Affairs of the Dutch Association of Psychologists (COTAN) (Evers 2001). This rating system evaluates the quality of a test on seven criteria:

1. Theoretical basis and the soundness of the test development procedure
2. Quality of the testing materials
3. Comprehensiveness of the manual
4. Norms
5. Reliability
6. Construct validity
7. Criterion validity.

Several items are included for each criterion. Some of these items are termed as 'key questions'(*), which ensure minimum conditions of quality are met. Consequently, should a key question be rated negative, the rating criterion will be 'insufficient'. Once all items have been rated, the final grades ('insufficient', 'sufficient' or 'good') for each of the criteria can be determined. For the purposes of this study only tests with a grading of either 'sufficient' or 'good' for all criterion will be included in the review.

4.2.3.3 Subscale versus total scores

To be considered as above.

4.2.4 Endpoint versus change data

For change data (endpoint minus baseline), the situation is even more problematic than endpoint data. In the absence of individual patient data it is impossible to know if data are skewed, though this is likely. After consulting the ALLSTAT electronic statistics mailing list, we hoped to present change data in Metaview to summarise available information. In doing this, it is assumed either that data were not skewed or that the analyses could cope with the unknown degree of skew. Without individual patient data it is impossible to test this assumption. Where both change and endpoint data were available for the same outcome category we presented endpoint data only. We acknowledge that by doing this, much of the published change data were excluded, but argue that endpoint data is more clinically relevant and that if change data were to be presented along with endpoint data it would be given undeserved equal prominence. Authors of studies reporting only change data were contacted for endpoint figures. We reported non-normally distributed data in the 'Other data types' tables

4.2.5 Cluster trials

Studies are increasingly employing 'cluster randomisation' (such as randomisation by clinician or practice), but analysis and pooling of clustered data poses problems. Firstly, authors often fail to account for intra class correlation in clustered studies, leading to a 'unit of analysis' error (Divine 1992), whereby p values are spuriously low, confidence intervals excessively narrow and statistical significance overestimated, thus causing type I errors (Bland 1997; Gulliford 1999). Secondly, RevMan does not currently support meta-analytic pooling of clustered dichotomous data, even when these are correctly analysed by the authors of primary studies, since the 'design effect' (a statistical correction for clustering) cannot be incorporated.

Where trialists had not accounted for clustering in primary studies, data were presented in a table, with a '*' symbol to indicate the presence of a probable unit of analysis error. Any subsequent versions of this review should seek to contact first authors of studies, to seek intraclass correlation co-efficients of their clustered data and to adjust for this by using accepted methods (Gulliford 1999).

If clustering had been incorporated into the analysis of primary studies, then we would have presented data in a table. No further secondary analysis (including meta-analytic pooling) will be attempted until there is consensus on the best methods of doing so, and until RevMan, or any other software, allows this. A Cochrane Statistical Methods Workgroup is currently addressing this issue. In the interim, individual studies will be very crudely classified as positive or negative, according to whether a statistically significant result ($p < 0.05$) was obtained for the outcome in question, using an analytic method that allowed for clustering.

5. Investigation for heterogeneity

Firstly, in any meta-analysis we were to have considered all the included studies within any comparison to judge clinical heterogeneity. Subsequently we would have made a visual inspection of graphs to investigate the possibility of statistical heterogeneity. This would have been supplemented using, primarily, the I-squared statistic. This provides an estimate of the percentage of variability due to heterogeneity rather than chance alone. Should the I-squared estimate have been greater than or equal to 75%, we would have interpreted this as indicating the presence of high levels of heterogeneity. If inconsistency had been high, data would not have been summated, but we would have presented them separately and investigated reasons for heterogeneity.

6. Addressing publication bias

Should there have been enough studies we would have entered data from all included studies into a funnel graph (trial effect against trial size) in an attempt to investigate the likelihood of overt publication bias (Egger 1997).

7. Sensitivity analyses

It was our intention to investigate the effect of including studies with implied randomisation and high attrition rates by sensitivity analyses. In addition, as the separate forms of amphetamine may have differential effects both centrally and peripherally, we would have also explored potential differences using sensitivity analyses. If there had been no substantive difference within primary outcomes (see types of outcome measures) we would have grouped all included data together in the final analyses. However, if there had been substantive differences, we would have considered groups separately.

8. General

Where possible, we entered data in such a way that the area to the left of the line of no effect indicated a favourable outcome for amphetamine.

Description of studies

Explanation:

This should refer to the information contained in the 'Characteristics of Included Studies' and the 'Characteristics of Excluded Studies' tables. It should briefly describe key characteristics of the study participants, interventions and outcome measures in the included studies and any important differences among the studies. The sex and age range of participants should be stated here unless it is obvious (e.g. if all the participants are pregnant). Reviewers should note any other characteristics of the studies that they feel the reader needs to be aware of.

Style:

This can be very short but usually has a pattern of headings such as: 'Excluded', 'Included - methods', 'Included - participants', 'Included - interventions', 'Included - outcomes', 'Included - notes', etc.

EXAMPLE: *Clotiapine for schizophrenia*

1. Excluded studies

We excluded twenty eight studies. These were mostly uncontrolled clinical trials where outcomes in the same people were compared before and after the use of clotiapine. Block 1974 was a randomised controlled trial, but did not include any participants suffering from acute psychosis. Kammerer 1967 fitted all the inclusion criteria, being a randomised controlled trial comparing clotiapine to 'standard medication' for treatment of 'psychotic' or 'pre-psychotic' people but not all participants were in an acute condition and it was impossible to extract data for these people alone. We therefore had to exclude this study. Rodova 1971 was randomised - according to the scheme A-B-A-B... - using clotiapine versus perphenazine for people with schizophrenia, some of whom were acutely ill. We excluded this study because it was a crossover trial and data were not available for acutely ill people before the cross-over. Kaneko 1969 compared clotiapine with chlorpromazine using matched pairs with treatment allocated by 'the double-blind technique'; randomisation was not mentioned. Moreover, not all participants were acutely ill and it was impossible to ascertain exact numbers of those who were acutely ill from the data presented. Van Wyk 1971 fitted all inclusion criteria, but more than 50% of participants were discharged from hospital before the end of the trial, without follow up, so there was not any outcome data for these people. Reported information did not allow calculation of exactly how many participants were discharged and therefore did not address the issue of 'hospital and service outcome'. For this reason, we also excluded this study.

2. Awaiting assessment

Two studies are currently awaiting assessment. One has proved unobtainable and the other requires translation.

3. Included studies

We identified no new studies for inclusion in the update search of 2004.

3.1 Duration

We included five trials, each lasting between six and sixty days. Only two studies, both of which lasted one week (Subramaney 1998; Uys 1996), measured outcomes at 24 hours or less, thus presenting some data regarding the speed of action of clotiapine. Other studies reported outcomes every few days or weekly (Itoh 1968).

3.2 Trial size

The number of people in the included studies ranged from 30 to 80. We could not always use data from all participants as some did not fit the inclusion criteria. Due to this fact, we were only able to include four participants from the largest trial (Itoh 1968) and the total number of people randomised in relevant studies is currently only 185.

3.3 Participants

All studies included people with schizophrenia. Jacobsson 1974 described how participants in their study suffered from 'psychotic syndromes of a schizophrenic type'. Perales 1974 included only people with 'paranoid schizophrenia'. In Subramaney 1998, people with 'organic' (substance induced) psychosis and bipolar disorder were also included. Uys 1996 also included those with bipolar disorder and people with 'acute paranoid reaction', 'other and unspecific reactive psychosis' and 'unspecific psychosis'. Every article contained the word 'acute' or 'acutely' in the description of the participants' condition, except Itoh 1968, which used the phrase 'in a state of intense excitement'. No definition of 'acute' was given. All trials were conducted on adults. Only Uys 1996 and Jacobsson 1974 gave their reasons for excluding participants from their studies.

3.4 Settings

All trials were conducted on people of both sexes, except Uys 1996, whose participants were all men. All were hospital patients and it seems that no person over the age of 65 was included in any of the trials. We cannot be certain of this last fact as Itoh 1968 only reported the mean age and standard deviation of its participants.

3.5 Interventions

3.5.1 Experimental drug

Clotiapine was given by intramuscular injection (40-160 mg/day) in Subramaney 1998; Uys 1996 and during the five first days in Perales 1974, after which it was administered orally. This treatment routine was also used in Itoh 1968 and Jacobsson 1974, in daily doses ranging from 40-290 mg.

3.5.2 Comparison drugs

Itoh 1968 used perphenazine orally, between 12 mg and 64 mg/day. Jacobsson 1974 used oral chlorpromazine as the control treatment, with doses ranging from 100 mg to 600 mg/day. Perales 1974 compared trifluoperazine to clotiapine. The drug was given IM - 2-8 mg/day - on days one to five and then orally - 10-40 mg/day - on the remaining 40 days. Lorazepam IM, up to 16 mg/day, was compared to clotiapine in Subramaney 1998. Uys 1996 used zuclopenthixol acetate with participants receiving 150 mg of this drug every 72 hours.

In Itoh 1968; Jacobsson 1974 and Perales 1974, clotiapine was compared to 'standard drug treatment', whereas in Uys 1996 it was considered as 'standard' medication when compared to zuclopenthixol acetate. In Subramaney 1998 clotiapine and lorazepam were of equal preference.

3.6 Outcome measures

For clinical outcomes, data were mostly continuous, presented as mean values commonly without standard deviations (SD). Uys 1996 presented data in graphical form only. All trials used some published rating scales (see below), but Subramaney 1998 was the only study to use exclusively published, validated rating scales. Other outcome measurements that were used

(Perales 1974). Jacobsson 1974 measured the number of people needing antiparkinsonian drugs, whereas Perales 1974 was interested in the quantity of such drugs used. Jacobsson 1974 addressed hospital and service outcomes by presenting the number of people substantially improved and discharged before the end of the trial.

3.6.1 Scales used

3.6.1.1 Behaviour

Overt Aggression Scale - OAS (Yudofsky 1986)

This is a scale designed to measure the level of verbal and physical aggression. Sixteen items, all aggressive incidents, are classified according to the degree of aggressiveness implicated. Rating also depends on the duration of each incident as well as the intervention required to control it. High scores indicate a higher level of aggression.

3.6.1.2 Mental state

Brief Psychiatric Rating Scale - BPRS (Overall 1962)

The BPRS is an 18-item scale measuring positive symptoms, general psychopathology and affective symptoms. The original scale has sixteen items, but a revised eighteen-item scale is commonly used. Scores can range from 0-126. Each item is rated on a seven-point scale varying from 'not present' to 'extremely severe', with high scores indicating more severe symptoms.

3.6.1.3 Adverse effects

Simpson-Angus Scale - SAS (Simpson 1970)

The SAS is a 10-item scale that is used to evaluate the presence and severity of drug-induced parkinsonian symptomatology. The ten items focus on rigidity rather than bradykinesia (slow or retarded movement), and do not assess subjective rigidity or slowness. Items are rated for severity on a 0-4 scale, with a scoring system of 0-4 for each item. A low score indicates low levels of parkinsonism.

4. Missing outcomes

Economic outcomes and satisfaction with care were not addressed in any of the trials.

5. Pharmaceutical industry support

At least two studies (Itoh 1968; Uys 1996) were sponsored by a pharmaceutical company.

EXAMPLE 2: Droperidol for schizizophrenia

1. Excluded studies

We excluded ten studies, three of which were not randomised. Girard 1972 was a case control study, and Weiser 1973 a case series. After e-mails from Dr Hooper it was clear that his study (Hooper 1983) also had to be excluded, as allocation to groups had not been random, with participants being alternately allocated to either the treatment or the control intervention. Most of the remaining trials were excluded because participants were not clearly suffering from psychotic illnesses. Foster 1995 included female patients undergoing minor gynaecological surgery. Richards 1998 and Thomas 1992 both included predominantly 'intoxicated' people. Thomas 1992 also included people suffering from trauma, an underlying medical condition, or who were undiagnosed, as did Rosen 1997. Both Richards 1998 and Rosen 1997 included some people with a 'psychiatric' diagnosis but their studies had to be excluded because outcomes for these participants, a small minority of the total, were not separately analysed. Cocito 1970 did include only people with psychosis, but not necessarily with acute illnesses.

Weiser 1975 would have been included, except for the addition of five people to replace those who left the study early. It is not clear to which group(s) those leaving early belonged, so the remaining data were rendered of little value once data from the non-random replacements had been incorporated.

2. Awaiting assessment

There are no studies awaiting assessment.

3. Ongoing

The reviewers know of no ongoing studies.

4. Included studies

We identified only three small trials investigating the effects of droperidol for acutely disturbed people.

4.1 Length of trials

The duration of trials ranged from 30 minutes (Van Leeuwen 1997) to 30 days (Cocchi 1971). This longer study is probably less relevant to the acute management of very disturbed people.

4.2 Participants

All trials included people with psychoses. Resnick 1984 did not specify beyond stating that participants were admitted involuntarily to the emergency department of a psychiatric unit. Van Leeuwen 1997 included people with schizophrenia, manic depression or in a 'confusional state', however ten participants had no specific diagnosis. Cocchi 1971 stated that all participants suffered from schizophrenia.

4.3 Settings

The setting for Resnick 1984 was the emergency department of a psychiatric unit. Van Leeuwen 1997 participants appear to have been hospitalised although no details of setting were stated explicitly. Cocchi 1971 was set in the psychiatric clinic of the University of Milan.

4.4 Study Size

Studies were small. The number of participants ranged from 20 to 40.

4.5 Intervention

The most acute study, Van Leeuwen 1997 administered 10 mg of droperidol iv and compared this with placebo. If participants did not respond to the initial interventions after three minutes, a single 5 mg injection of haloperidol im was given. Furthermore "individually adapted psychotropics" could be administered 30 minutes into the trial should participants fail to respond satisfactorily to either of the previous injections. Resnick 1984 administered 5 mg of droperidol im, with repeat injections of the same dose given at 30, 60 and 90 minutes if necessary. In this study droperidol was compared with haloperidol (5 mg im). Cocchi 1971 administered droperidol (2-10 mg) orally and compared this with 2-10 mg of oral haloperidol.

4.6 Outcomes

4.6.1 Global impression

The primary outcome in Van Leeuwen 1997 and Resnick 1984 was whether the patient required further medication after the initial injection. Van Leeuwen 1997 used clinical judgement, rather than a scale, to decide whether or not "adequate psychomotor sedation" had been achieved after the initial injection. For Resnick 1984, the need for further medication was signified when patients scored more than a total of 17 in six categories (anxiety, tension,

mannerisms/posturing, hostility, unco-operativeness and excitement) of the Brief Psychiatric Rating Scale (Overall 1962), although no further details were given regarding the severity of symptoms implied by this score. Cocchi 1971 measured general clinical improvement using a three point unpublished scale. He also used an unpublished 'specific symptoms scale' to measure levels of improvement in individual symptoms/behaviours. As these scales were unpublished (Marshall 2000), and the paper reported only means without standard deviations, these data are not presented.

4.6.2 Mental state

Cocchi 1971 measured changes in mental state at three intervals over a 30-day period using the Rating Scale for the Quantification of Psychotic Symptoms (Goodrich 1953).

4.6.3 Adverse effects

All trials attempted to report on adverse effects. Van Leeuwen 1997 did so for the first three minutes after the initial injection of droperidol or placebo, and stated only that, "no side effects related to the injection of the double blind medication were seen before the administration of haloperidol". Resnick 1984 made observations for adverse reactions, especially movement disorders, for 24 hours after the initial injection. This researcher also measured vital signs for three hours following the initial injection, although no data were given and the outcomes are unusable.

4.6.4 Outcome scales

4.6.4.1 Mental state

Brief Psychiatric Rating Scale - BPRS (Overall 1962)

This is a brief rating scale used to assess the severity of a range of psychiatric symptoms, including psychotic symptoms. The original scale has 16 items, although a revised 18-item scale is commonly used. Each item is defined on a seven-point scale varying from 'not present' to 'extremely severe', scoring from 0-6 or 1-7. Total scores can range from 0-126, with high scores indicating more severe symptoms.

Rating Scale for the Quantification of Psychotic Symptoms (Goodrich 1953)

At this time (2003) we have not been able to find a description of the properties of this scale.

5. Missing outcomes

Hospital and service outcomes, satisfaction with care, and economic outcomes were not addressed by any of the trials.

6. Pharmaceutical industry support

Though the trials do not mention specific sponsorship by a pharmaceutical company, Van Leeuwen 1997 thanks Janssen Pharmaceutical for "the supply of double-blind medication".

Methodological quality of included studies

Explanation:

This should describe the general quality of the included studies. The quality of the description of randomisation, blinding and follow up are usually commented upon.

Style:

This is usually quite succinct.

EXAMPLE: *Droperidol for schizophrenia*

1. Randomisation

Van Leeuwen 1997 specified that treatment was "randomly assigned" with participants listed in chronological order and assigned individually numbered vials. It is therefore unclear whether those randomising could have ascertained the order of prescribing. Resnick 1984 did not specify the explicit means of allocation, although he stated that patients received treatment on a 'randomised basis', and that the codes identifying the packages of medication were "kept in the pharmacy until the conclusion of the study". Cocchi 1971 specified only that the study was randomised, with no details regarding the means of allocation. As the description of randomisation was never explicit, and there is no mention of blinding during testing (see below), all studies were assigned to category 'B', i.e. moderate potential for inclusion of bias.

2. Blinding at outcome

All reports specified blindness at outcome. Van Leeuwen 1997 stated that the trial was "strictly double blind". Resnick 1984 said that the haloperidol and droperidol solutions were held in "identical appearing vials". Cocchi 1971 gave no details of the method of blinding.

3. Loss to follow-up

All participants that left the Van Leeuwen 1997 and Resnick 1984 studies early did so because the treatment worked and they did not need further medication. All patients were included in the analysis while they needed medication. Cocchi 1971 had no people leaving the study early and included all participants in the analysis.

Results section

The text of the results section is discussed later. However, as this is a guide to a complete review within the Module of the CSG, and there is no place to formally discuss the structure of the Table of Comparisons, this is undertaken at this point.

Table of Comparisons

Explanation:

It is easy to have this part of your review untidy. With just a little time and effort, however, you can present data extremely clearly within the Table of Comparisons. Have a look at already completed CSG reviews on the Cochrane Library. They are your best guide. At the first level are the comparisons, that is treatment X versus treatment Y. At the second level the outcomes are set up. Group outcomes into what they are reporting, that is, global measures of effect, mental state, behavioural, leaving the study early, death, adverse effects, etc. Finally at the third level can be the subgroups. There is wide variation in how these levels are used and the following is just one of several examples that can be followed. Please look at other CSG reviews for additional ideas.

Style:

In the Table of Comparisons, style is very important. Please be careful as it will save us all time later. It looks best to put the interventions being compared all in capitals. Please have a look at already completed reviews on the Cochrane Library such as CLOZAPINE or RISPERIDONE reviews. So if you are looking at a new drug versus an old one put something like:

EXAMPLE 1:

01.00.00 CLOZAPINE versus TYPICAL ANTIPSYCHOTICS

Do not put full stops at the end and please spell out 'versus'.

Put 'versus' in lowercase as this makes the interventions stand out in all the graphs.

This is followed by the primary outcomes at level 01.01, 01.02 etc.

For these, put the capital only at the start of the statement.

Example 2: DICHOTOMOUS OUTCOME

01.00.00 CLOZAPINE versus TYPICAL ANTIPSYCHOTIC

01.01.00 Death

When setting up the outcome you will be asked if the data is dichotomous, continuous or other. Death is dichotomous! The other two will be dealt with later. As you set up the outcome you can set default statistics. As mentioned in 'Methods' above, the settings should be set to Relative Risk, fixed effect, 95% CI. Then you will be asked if you need subtotals or totals, or none. Again, this will be explained in a short time. . . . Forget about the graph settings until you see your graph. In Metaview you can then adjust the scale to make things look clear and if this setting is different from the default you can re-enter the outcome settings and change the default.

Lastly we have the subgroups. There is always much temptation to put lots of these (males, females; points in follow up - 0-6 weeks, seven weeks - six months. . .) but you can only display one set. Choose the one you want to display BEFORE looking at the data.

These should be all in lower case, please.

01.00.00 CLOZAPINE versus TYPICAL ANTIPSYCHOTIC

01.01.00 Death

01.01.01 any cause

01.01.02 physical ill health

01.01.03 suicide

Please note the subgroups are all lowercase and ordered - in this case alphabetically, in others it may be by time period.

This brings us to the totals or subtotals problem. In the above example the reviewers wanted to display an 'any cause' subgroup. Let's presume this is simply the addition of the 'suicide' and 'physical ill health' subgroups. Where subgroups are not mutually exclusive, as in this case, use the 'subtotals only' command as you set up the outcome in the comparisons table. In the data example in this Guide Review the Death outcome is only given two sub-groups that are mutually exclusive and so the result is totalled.

This will set up a self explanatory table in the data entry section of your RevMan. Remember, try and undertake an intention-to-treat analysis as you enter the data (see Methods section).

Example 3: CONTINUOUS OUTCOME

This is not much different.

If it is scale data, please do not use the name of the scale as the outcome. A scale is a means by which an outcome is measured. To continue the above example:

01.00.00 CLOZAPINE versus TYPICAL ANTIPSYCHOTIC

01.01.00 Death

01.01.01 any cause

01.01.02 suicide

01.01.03 physical ill health

01.02.00 Mental state: 1. Average score (BPRS endpoint, low=poor)

01.02.01 by 6 weeks

01.02.02 by 6 months

01.02.03 by 1 year

The outcome is presented with its scale, whether it is a CHANGE or ENDPOINT score, and some indication of what the scores mean. As you set up the outcome this time you are prompted to choose the statistic of 'Weighed mean difference' or 'Standardised mean difference'. Again, it is unlikely that the time subgroups are mutually exclusive so the 'subtotals only' button is pressed. The example that has been included in the 'Analysis' in this Guide Review is rather dramatic and best viewed on a different scale of graph as is provided by the default RevMan settings. Again, these defaults can be changed as you set up your outcomes.

Once you come to enter data you will get a slightly different set of boxes into which to enter data. Please see examples provided.

Example 4: OTHER DATA

Most frequently within CSG reviews this is used for continuous data that are too skewed to be presented (see Methods section). If the standard deviation of endpoint scores multiplied by two is greater than the mean then the data probably shouldn't be presented as a graph. RevMan will produce a graph but the statistics become suspect the greater the skewness. Ask your Contact Editor for advice.

In the 'Comparisons Table' there is little difference apart from pressing the 'Other' button when asked within the outcome set up. By doing this a set of text boxes is provided. Please see worked examples.

Please note that all data in the review are MADE UP for the purposes of demonstration. References do not even relate to the included and excluded trials table. Having discussed the Table of Comparisons this Guide will now proceed to discuss the text of the Results section of your review.

Results

Explanation:

This summarises the main findings of the review and any sensitivity analyses. Do not forget that the search is an important part of your work that should be reported. The results section should simply state in text what the results are. Comments on what the results actually mean are reserved for later in the review (Discussion section). As MetaView does not calculate NNT or NNH this is where it may be presented for the first time.

Reviewers should avoid making inferences in this section. A common mistake to avoid (both in describing the results and in drawing conclusions) is the confusion of 'no evidence of an effect' with 'evidence of no effect'. When there is inconclusive evidence, it is wrong to claim that it shows that an intervention has 'no effect' or is 'no different' from the control intervention. It is safer to report the data, with a confidence interval, as being compatible with either a reduction or an increase in the outcome.

Style:

Subheadings should be used to make reading easier, and titled and enumerated as for the comparisons and data tables.

EXAMPLE: *Gamma-aminobutyric acid agonists for neuroleptic induced tardive dyskinesia*

1. The search

The first electronic search identified 47 reports and the update search of 2003 identified a further 43 reports. Agreement regarding reports that may have been randomised controlled trials was 100%. Thirty-one trials were identified and eight randomised trials were included in this review.

2. Comparison: GABA drugs versus placebo 2.1 Tardive dyskinesia outcomes

Only three studies reported outcomes (medium term at 6 weeks) which could be used to assess clinically important improvements (reduction of more than 50% on any scale - Burner 1989;

Fisk 1987; Stewart 1982). Together (total n = 108) they show no significant increase in clinical improvement with GABA drugs (RR 0.83 CI 0.6 to 1.1). Similar outcomes, measured in different ways, showed a tendency for reduced TD symptoms but no finding was statistically significant. Synthesis of homogeneous data for the outcome of 'not any improvement' gave equivocal results (n = 152, six randomised controlled trials (RCTs), RR 0.76 CI 0.5 to 1.17). We did not combine data on symptom scores, both endpoint scores and change ratings, as trials used different scales to assess TD that are not directly comparable. None, however, suggested a significant effect for the GABA drugs. The number of people whose symptoms of TD were deteriorating was not reduced by the GABA drugs (n = 123, four RCTs, RR 1.83 CI 0.57 to 5.92).

2.2 Adverse effects

Six of the eight trials reported adverse effects that are included in the analysis. Ananth 1987 reported that none of the ten included people suffered adverse effects (baclofen versus placebo study). Unfortunately the reporting of adverse effects in Linnoila 1976 was ambiguous and we could not present data (sodium valproate versus placebo study). Compared with placebo, GABA drugs do not clearly cause ataxia, dizziness/confusion, loss of muscle tone, nausea/vomiting, restlessness/akathisia, sedation or seizures. There is a suggestion of an increase in ataxia for both baclofen and sodium valproate (n = 95, two RCTs, RR 3.26 CI 0.4 to 30.2), and in sedation (n = 113, three RCTs, RR 2.12 CI 0.8 to 5.4) compared with placebo, but these findings are not statistically significant. Withdrawal of THIP may cause seizures. In one trial (Thaker 1987) the presence of tonic-clonic seizures after withdrawal of THIP led to the termination of the study (n = 2).

2.3 Mental state

Four trials reported 'deterioration in mental state' as an outcome. Such deterioration was more likely to occur for people receiving GABA medication (n = 95, four RCTs, RR 2.47 CI 1.1 to 5.4), but this effect was influenced by the decision of assigning a negative outcome for those who dropped out before the end of study and we are not confident of the findings.

2.4 Leaving the study early

A greater proportion of people allocated GABA medication failed to complete the trial compared with those allocated placebo (20% versus 9%) but this difference was not statistically significant (n = 136, five RCTs, RR 1.99 CI 0.8 to 4.7).

2.5 Sensitivity analyses

The above results are based on data as presented in the original study reports, with the assumption that those who dropped out before the end of the trial had not improved. When the sensitivity of the results to this assumption was tested, no appreciable differences in the results were found. Additionally, insufficient information was provided to test the hypothesis of a different treatment response for people less than 40 years old, or a differential effectiveness among the various compounds. It was also not possible to ascertain the medium or long-term effects of these drugs.

3. Additional data

Review of the additional second period data from the five cross-over trials included in this review adds little. A significant improvement was reported with baclofen (Gerlach 1978; Ananth 1987), sodium valproate (Linnoila 1976) and THIP (Thaker 1987), but these trials also report a range of moderate to severe adverse effects. Nair 1978 described neither improvement nor adverse effects with baclofen during the follow-up period.

Discussion

Explanation:

This should include brief comments on any methodological limitations of the included studies and the review that are important for decisions about practice or future research. Comments on how the included studies fit into the context of other evidence might be included here, stating clearly whether the other evidence was systematically reviewed. Comments on how the results of the review fit into the context of current clinical practice might be included here, although reviewers should bear in mind that current clinical practice might vary internationally.

Style:

Make sure that this section follows the same order as 'Results' but avoid repetition! It may well be possible to discuss results under more collective headings. For example, the 'Results' section may have had the headings 'Mental state: 1. Average change in BPRS' and 'Mental state 2. Average change in PANSS' and so on. In the 'Discussion' these all could be subsumed under the collective title 'Mental state'.

EXAMPLE 1: *Clotiapine for schizophrenia*

(Please note how the order reflects that of the Results section, in turn reflecting the order in the Comparisons table):

1. Few studies

Acute psychosis is difficult to study and co-operation from the study population is rare. This may be one of the reasons for the scarcity of controlled clinical trials using clotiapine solely for this indication. Few relevant studies have taken place since the mid 1970s, which probably correlates with its disappearance from the market in many countries around this time (USA, United Kingdom and France amongst others). Novartis states that the main reason for its disappearance was financial, as it was not economically feasible to market an off-patent compound. More recently, despite some interest in its use for chronic treatment-resistant schizophrenia (Lokshin 1998), clotiapine has mostly been associated with management of alcoholism and drug abuse. More studies are needed for proper analysis of clotiapines efficacy.

2. Methodological quality

2.1 Randomisation

Selection bias can be minimised through good quality blinding at allocation. Improper allocation concealment can affect estimation of treatment effects (Chalmers 1983). Only Itoh 1968 and Perales 1974 reported adequate methods of random sequence generation. Allocation concealment was also well described in Itoh 1968, but was not mentioned in any other study. No trial included in this review would have rated highly with respect to the CONSORT statement (Moher 2001).

2.2 Blinding

Rigorous blinding procedures are considered important to minimise performance and attrition bias in controlled trials (Jadad 1996). In psychopharmacological studies, it could be assumed that blinding is particularly important, as outcome measures may be subjectively-rated items. If outcomes are more concrete, such as death, hospital discharge and leaving the study early, blinding may not be quite so important. An adequate blinding mechanism (identical-looking medication) was described in Itoh 1968 and Jacobsson 1974. Such a description was absent in Subramaney 1998. Perales 1974 and Uys 1996 could not be considered double-blind trials as

The compared treatments differed (two tablets versus one tablet, no placebo injections to match the number of injections in both groups), and were administered by an un-blinded nurse. Moreover, no trial tested the blindness (by asking participants and research personnel which drug they thought was being administered) although there has been evidence that physicians can, more often than would be allowed by chance, differentiate a new drug from the older one used as a control (Fisher 1993). This can affect estimated efficacy of the tested drugs. Indeed, in a meta-analysis of 22 'double-blind' studies, the effect sizes of older drugs were only one-half to one-quarter the size of those reported in earlier studies in which the old drugs were the only agent appraised (Greenberg 1992). This bias is difficult to eliminate since what is measured in a controlled trial is precisely the difference in efficacy and safety between several tested treatments, and this difference is often what allows participants or raters to guess which medication is being administered.

2.3 Description of losses to follow up

Attrition rates were low in all studies which is a common factor in short trials conducted in hospital. The highest number of drop-outs was in Jacobsson 1974. This was the only trial to mention drop-outs apart from Uys 1996, who listed 'protocol violation' as a reason for leaving the study early.

2.4 Data reporting

Overall, data reporting was poor. Results were most often presented in mean values with no SD, making it impossible to re-analyse. Unlike the other studies, Subramaney 1998 did present variances with its results and Uys 1996 might have included this information if it been a fully published trial.

3. Publication bias

The data included in this review were insufficient to enter into a funnel plot so we could not address the possibility of publication bias.

4. Generalisability of findings

4.1 Setting

All trials were set in hospital. Many people suffering from acute psychosis do require hospital admission so the results are relevant to everyday practice. Nevertheless, it would be of interest to assess the use of clotiapine for acute psychosis on an outpatient basis, thereby preventing the need for hospitalisation.

4.2 Duration

The difference in duration between trials - six days to 60 days - influences the outcomes measured. Rapid tranquillisation is the first outcome of interest when dealing with acute psychotic people and should be assessed by measuring change in behaviour less than 24 hours after onset of medication. Only the shorter studies (Subramaney 1998; Uys 1996) addressed this time period, but we could not use their data. The longer studies had the potential to address the outcome of antipsychotic effect (Itoh 1968; Jacobsson 1974; Perales 1974), but again, very few data are presented.

4.3 Participants

We modified the Inclusion criteria in the protocol during the reviewing process to include people suffering from acute psychosis due to substance abuse. This change was made because we felt there was no reason to exclude this category of participants as, in an emergency situation, it is not always possible to distinguish between drug-induced and functional

psychosis - and there is a substantial co-morbidity - so the same treatment is applicable in both cases. This modification allowed us to include Subramaney 1998.

Groups of participants differed among the various included trials. Itoh 1968 included only people with schizophrenia, Jacobsson 1974 included those suffering from 'psychotic syndromes of a schizophrenic type', Perales 1974 limited inclusion criteria to paranoid schizophrenia and Uys 1996 specifically excluded people with organic psychosis. Only two trials, Subramaney 1998; Uys 1996 mentioned which diagnostic criteria they used to include people in their trial (DSM-III-R in the former, ICD 9 in the latter). It seems reasonable to suppose, with a few exceptions, that participants in Itoh 1968; Jacobsson 1974 and Perales 1974 were all 'psychotic', as in our current definition of the concept. No clear definition of 'acute' was reported, although all trials used this concept. We are unsure as to the exact meaning of 'acute' in these trials, but again, think it reasonable to assume that such a pragmatic use of this term would not limit the applicability of the results.

4.4 Control drug

A different control drug was used in each included trial, making results difficult to compare, although this is somewhat obviated by each of the antipsychotic controls being a phenothiazine (Itoh 1968 - perphenazine, Jacobsson 1974 - chlorpromazine, Perales 1974 - trifluoperazine). In Uys 1996 (comparison drug - zuclopenthixol acetate), clotiapine was not the experimental drug but the control. This could have reduced its measured effect if the blinding was not rigorous (Greenberg 1992).

4.5 Outcomes

Among the seven groups of pre-defined outcomes, only five were addressed. We found no information regarding important outcomes such as tranquillisation, injury to others, compulsory administration of treatment, measured acceptance of treatment, change in hospital status, economic outcomes and satisfaction with care. Except for measurements of global improvement/deterioration, outcomes were always assessed using published rating scales. Where scales were employed, however, no criteria for response were defined and degree of improvement was proportional to the change in number of points recorded on each scale. This does not necessarily reflect a satisfactory degree of improvement as would be defined by carers, patients or their families. Along with the poor reporting of data, the limited outcomes recorded greatly lessens not so much the applicability, but the value of this review.

5. COMPARISON 1. CLOTIAPINE versus OTHER ANTIPSYCHOTICS

5.1 General improvement

Four studies addressed this outcome (Itoh 1968; Jacobsson 1974; Perales 1974; Uys 1996), but Uys 1996 did not report usable data. Data are quite heterogeneous (I-squared 58%). Itoh 1968 and Jacobsson 1974 were administering drugs orally and used a low potency neuroleptic as a comparison, whereas Perales 1974 compared clotiapine IM to trifluoperazine IM - a highpotency neuroleptic. For this general outcome there is no trial-based evidence that clotiapine is either better or worse than other drugs.

5.2 Hospital and service outcomes

Only one small (n = 49) trial addressed the issue of hospital discharge due to substantial improvement (Jacobsson 1974). The results were equivocal but it is problematic to draw conclusions from such limited data.

5.3 Leaving the study early

Few people left these studies early and clotiapine was no different to the other control drugs.

5.4 Adverse effects

Again data are limited. Two studies reported on the use of antiparkinsonian medication to control neuroleptic-induced movement disorders, and there was significant heterogeneity between their results. Jacobsson 1974 used approximately the same amount of anticholinergic drugs in both treatment groups. This was probably because clotiapine was being compared to chlorpromazine, a similar drug as regards propensity to cause parkinsonism, whereas when clotiapine was compared with zuclopenthixol (Uys 1996), the latter group needed anticholinergic drugs more frequently than their counterparts receiving clotiapine.

Overall, the number of people suffering adverse effects was very small so it is impossible to determine the side effect profile of clotiapine. If new trials were to be conducted, however, the incidences of rashes and seizures resulting from use of clotiapine would be of particular interest.

5.5 Additional outcomes presented in the trials but with no useful data for analysis

5.5.1 Behaviour

No definitive conclusion can be drawn regarding clotiapine's effectiveness in the management of behavioural problems. Data are poorly reported, although Jacobsson 1974 seemed to indicate that clotiapine has a positive effect on behaviour, comparable to that of chlorpromazine - a low-potency neuroleptic with strong sedative properties. Behaviour may have been evaluated in the general improvement ratings. If so, separate information should have been given regarding this very relevant outcome.

5.5.2 Sedation

Surprisingly, only one study, Uys 1996, addressed this outcome, when in many cases sedation proves the only way to manage acutely psychotic people with severely disturbed behaviour. Unfortunately, the finding that clotiapine has rapid sedative properties equal to those of zuclopenthixol acetate, was not supported by data. If this study had been fully published it would have been of great interest.

5.5.3 Mental state

There is little evidence to support the superiority (or inferiority) of clotiapine's psychotropic properties when compared to other compounds. Globally, clotiapine was shown to be as good as other 'standard' and 'non-standard' treatments of acute psychosis. Perales 1974 found the drug superior to trifluoperazine but presented only inexact 'p' values. As presentation of results on psychopathological measurements was generally poor, most data could not be used. However, although mental state as an outcome is of interest in the very short-term follow up of these studies, it is not usually as important as immediate tranquillisation or sedation.

5.6 Outcomes absent from these studies

Satisfaction with care and economic outcomes were not reported. This may be because these have only recently been recognised as important outcomes.

6. COMPARISON 2. CLOTIAPINE versus BENZODIAZEPINES

Only one small study (Subramaney 1998) used a benzodiazepine, lorazepam, as the drug of comparison. If clotiapine is still a drug of choice in the emergency situation, as it appears to be in some countries, further comparative trials are needed.

6.1 Mental state

No substantial difference in improvement between clotiapine and lorazepam was reported, although the latter is not known for its antipsychotic properties. This is not particularly surprising, as all participants were also receiving haloperidol as baseline therapy, with clotiapine and lorazepam being administered only for aggressive/violent outbursts.

6.2 Leaving the study early

Only one participant in each treatment group dropped out before the end of the trial. It is not possible to draw conclusions except that study attrition was appropriately low.

6.3 Behaviour

Data from the use of the Overt Aggression Scale (OAS) were skewed and are therefore presented in tables. Behaviour improved with both treatments. It is possible that each drug is equally potent at quelling aggression, or is equally useless, and that other methods used together with time, settled the situation.

6.4 Adverse effects

Subramaney 1998 found fewer movement disorder adverse effects with lorazepam. This is an important advantage for this drug. However, in the absence of good data on tranquillisation (such as time from administering to tranquillisation), the primary reason for drug use in this trial, it is difficult to know how to consider the adverse effect profile. Should clotiapine be more tranquilising than lorazepam then its ability to quell the dangerous situation may outweigh the problem of clotiapines association with movement disorders.

EXAMPLE 2: Droperidol for schizoprenia

1. Droperidol discontinued

Janssen-Cilag Limited produced droperidol (<http://www.janssen-cilag.co.uk/index.asp>) until March 2001, when production of all formulations of its branded form (Droleptan) were discontinued. The Medicines Control Agency in the UK (<http://www.mca.gov.uk/>) had raised concerns about the potential effects of long-term droperidol on the electrical conduction of the heart (cardiac Q-T interval prolongation) and requested a risk-benefit assessment. The company concluded that the oral formulations should be discontinued to prevent use. The authors have asked Janssen-Cilag UK whether or not the injectable form for rapid tranquillisation was withdrawn because of concerns about safety. The company have informed us that the injectable form of droperidol also carries significant risks (Lawrence 1997) and has been associated with prolonged Q-T intervals (Guy 1991; Lischke 1994) as well as the rare, but potentially fatal, cardiac arrhythmia, 'Torsade de pointes' (Guy 1991; Kosai 1995; Michalets 1998). Although the benefits of its continued use in the acute situation may have outweighed these risks, careful monitoring - including ECGs and electrolyte assays - would have been necessary. This would have compromised the cost effective production of injectable droperidol, and so it was discontinued along with the oral preparations.

2. Small number of studies

Acute psychosis is difficult to study and co-operation from the study population is rare. This may be one of the reasons for the scarcity of controlled clinical trials using droperidol solely for this indication. Droperidol appears to have been widely used in emergency room situations for people who are agitated or acutely disturbed but who have not, at point of medication, been

diagnosed (Binder 1999; Pilowsky 1992). Several papers identified by our searches concerned the use of droperidol for people who were later diagnosed as having either trauma, an underlying organic condition, or who were intoxicated. As these diagnoses fell outside our remit for types of participants, the trials were excluded from this review. Nevertheless, acute disturbance due to suspected mental illness is so common (Huf 2002), and management of such situations so important, that there is little excuse not to have good evidence for the use of droperidol. However, a total of only 69 people seem to have been randomised into trials of droperidol versus placebo or haloperidol that are relevant to the emergency control of disturbance thought to be due to mental illnesses.

3. Quality

No trial reported adequate methods of random sequence generation, and only Van Leeuwen 1997 included any description of the method of randomisation. No trial included in this review would have rated highly with respect to the CONSORT statement (Begg 1996; Moher 2001) and the inclusion of bias is likely.

4. Publication bias

With such small studies publication bias is also likely, and for both truly relevant trials (Resnick 1984, Van Leeuwen 1997) droperidol was the experimental intervention, so any publication bias would favour droperidol.

5. Applicability of findings

The included trials were set in a psychiatric hospital, an emergency department, and a psychiatric crisis unit. The truly relevant studies (Resnick 1984, Van Leeuwen 1997) reported outcomes for the very short term, i.e. those of value in the crisis situation. Van Leeuwen 1997 included people with schizophrenia, manic-depression and confusional states, with seven of the 41 participants having no diagnosis recorded. Resnick 1984 included participants who were involuntarily hospitalised and had 'underlying psychoses'. This trial specifically excluded patients who were intoxicated (the only trial to specifically mention exclusion criteria). These inclusion criteria should make any findings applicable to the acute management of disturbed people thought to suffer from serious mental illnesses. However, although need for repeat injection is of importance, it would also have been desirable to have had outcomes such as further aggressive episodes, tranquillisation, sedation and carer satisfaction.

5. COMPARISON 1. DROPERIDOL versus PLACEBO

The small amount of data available ($n = 41$) suggests that droperidol is superior to placebo three minutes after injection (fewer repeat injections needed). However, the design of this study precluded good information about adverse effects, and, while such limited data could generate hypotheses, it does not provide conclusive evidence.

6. COMPARISON 2. DROPERIDOL versus HALOPERIDOL

Although the search identified two trials, only Resnick 1984 ($n = 27$) was clearly relevant to the acute management of disturbed people. Cocchi 1971 did not assess the immediate effects of droperidol given as emergency medication, but compared it to haloperidol in a study of 30 days' duration. Droperidol was not statistically different to haloperidol for the proportion of participants needing additional injections at 30 minutes, but this difference would be clinically relevant if sustained in larger studies (~40% droperidol versus ~60% haloperidol). Haloperidol did cause a mild dystonic reaction in one person. The results of Resnick 1984 indicate that droperidol is a valuable drug in the acute situation, but, were it still being used, all findings would need to be replicated.

7. Heterogeneity

This review is a re-presentation of the findings of trials rather than a meta-analysis in which heterogeneity could operate.

8. Sensitivity analysis

It had been hoped to conduct sensitivity analyses comparing results when trials with high attrition with those with only competitor data. All three trials reported on loss to follow up and this was not undertaken.

Reviewers' conclusions

The primary purpose of the review should be to present information, rather than to offer advice. Implications for practice and Implications for research are subheadings in this section.

Implications for practice

Explanation:

The implications for practice should be as practical and unambiguous as possible. They should not go beyond the evidence that was reviewed.

Style:

Be honest but tactful.

EXAMPLE 1: *Droperidol for schizophrenia*

1. For people with psychotic illness

Acute psychotic illness, especially when associated with agitated, aggressive or violent behaviours, may require urgent pharmacological tranquillisation or sedation. Droperidol, a butyrophenone neuroleptic that is no longer manufactured, may still be used for this purpose where supplies have not run out. Use of droperidol is founded on clinical experience rather than evidence from randomised trials, and is now limited as world supplies diminish.

2. For clinicians

Intramuscular droperidol was a popular choice for the acute management of very disturbed people. The evidence presented in this review allows no firm conclusions to be drawn about its comparative efficacy to haloperidol or other drugs used in this situation. It is quite possible that one drug did have advantages over another, and clinicians could have helped evaluate different approaches by supporting clinically relevant randomised controlled trials. However, this is no longer possible as droperidol is increasingly of historical interest only.

3. For managers/policy makers

Currently, those wishing to produce guidelines for management of the acutely disturbed person have so little data to work with that it is understandable that the finished results differ (Wing 1998, McEvoy 1999), and are unsatisfactory and of limited value. Droperidol is unlikely to feature as a viable option any longer.

EXAMPLE 2: *Clotiapine for schizophrenia*

1. For people with psychotic illness

The use of clotiapine in preference to other methods of managing acute psychotic illness is not supported by research. There is no evidence supporting claims made by open clinical trials that it has a faster onset of action and stronger sedative properties than other compounds. There is, however, no evidence refuting its efficacy. Clotiapine may cause fewer movement disorders than zuclopenthixol acetate but does cause more than lorazepam

2. For clinicians

Recommendations on the use of clotiapine for people suffering from acute psychosis, rather than other 'standard' or 'non-standard' treatments, is to be viewed with caution. This review

found only a few, poorly reported controlled trials. Wide confidence intervals on all findings preclude statistically significant findings. Taking clotiapine results in less need for anticholinergic medication for extrapyramidal side effects compared to 150 mg zuclopenthixol acetate IM. Orally, clotiapine seems to produce a degree of clinical improvement and quantity of movement disorders comparable to that of chlorpromazine. No trial addressed the major clinical issue of 'rapid tranquillisation'. Clinical practice is often based on clinicians' judgement in the light of their experience. Nevertheless, this judgement is sometimes not up-to-date, or is based on a few isolated cases and not on objective issues. We strongly recommend that clinicians who value the use of clotiapine for acutely psychotic people should consider undertaking a randomised controlled trial of this drug, including a large study population and focusing on outcomes such as tranquillisation, aggression, sedation and satisfaction with care.

3. For managers/policy makers

People who include clotiapine in their protocols for the management of acute psychosis should be aware that they could be criticised for this choice and would have no good evidence to support their preference for this drug.

4. For funders

There is a need for funders to sponsor well designed, conducted and reported controlled trials on clotiapine for the management of acute psychotic people.

Implications for research

Explanation:

The implications for research should not include vague statements such as 'more research is needed'. Reviewers should state exactly what research is needed, why and how urgently. Opinions on how the review might be improved with additional data or resources might also be included here.

Style:

EXAMPLE: Clotiapine for schizophrenia

1. General

As with all similar studies, public registration of a study before anyone is randomised would ensure that participants could be confident that people would know that the study had at least taken place. Unique study numbers would help researchers to identify single studies from multiple publications and reduce the risk of duplicating the reporting of data. Compliance with CONSORT (Moher 2001), both on the part of authors and editors, would help to clarify methodology and many outcomes.

Intention-to-treat analysis should be performed on all outcomes and all trial data should be made easily accessible. A minimal requirement should be that all data should, at least, be presented as numbers. In addition, continuous data should be presented with means, standard deviations (or standard errors) and the number of participants. Data from graphs, 'p' values of differences and statements of significant or non-significant differences are of limited value. Unfortunately, in spite of the large numbers of participants randomised, we were unable to use

most of the data in the trials included in this review due to the high drop-out rates observed and poor data reporting.

2. Specific

2.1 More studies

This review highlighted the need for good quality controlled trials to assess the efficacy of clotiapine for the management of acute psychosis, in particular addressing outcomes of major importance such as tranquillisation, aggression, sedation and satisfaction of carers. There should also be controlled trials undertaken in the community to assess the drug's efficacy and safety in preventing hospitalisation.

2.2 Protocol adherence

Acutely psychotic people are a difficult study population. Nevertheless, adherence to a simple, pragmatic trial protocol will avoid many problems and there are examples of such studies (Huf 2002b; TREC 2003).

2.3 Randomisation

To reassure readers that selection bias has been properly addressed, explicit description of treatment allocation must be reported. Trialists should not only use, but also describe an adequate method by which they generated a random sequence and the methods used to conceal treatment allocation until the beginning of the trial.

2.4 Blinding

Blinding participants, carers, outcome evaluators and possibly sponsors to the treatment is important to eliminate performance and detection bias. Blinding should, when possible, not only be attempted but tested.

2.5 Withdrawals

Authors should avoid withdrawing people from analysis after randomisation as a consequence of protocol non-compliance. Researchers should perform an intention-to-treat analysis and describe from which group withdrawals came, in order to evaluate exclusion bias.

2.6 Setting

The effectiveness of clotiapine in reducing acute psychotic symptomatology should be evaluated in the community as well as in a hospital setting.

2.7 Outcome measures

Primary outcomes of interest for the measurement of a drug's efficacy in handling acute psychosis, such as tranquillisation, sedation, level of aggression and injury to others, should be assessed within minutes or hours after the administration of the studied medication. Indeed, a treatment cannot be recommended for this indication if it has no immediate or short term clinical effect. Thereafter, outcomes such as 'compulsory administration of treatment', 'mental state' and 'acceptance of treatment' are also of relevant interest to mentally ill people and their carers, as are economic outcomes and satisfaction with care.

2.8 Presentation of data

Authors should present measures of association between intervention and outcome, for example, relative risks, odds-ratios, risk or mean difference, and the raw numbers. Binary outcomes should be reported as well as, or in preference to, continuous results as they are easier to interpret. It is strongly recommended that authors report confidence intervals and statistical power for comparisons presented in the papers. If p-values are used, the exact value should be reported.

Acknowledgements

Explanation:

This section should be used to acknowledge any individuals or organisations who have helped with the review but who have not made a sufficient contribution to the review to be included in the Contributions section.

Style:

Follow the form indicated below, beginning with 'We would like to thank. . . .'

EXAMPLE: *Clotiapine for schizophrenia*

We would like to thank Clive Adams, Drew Davey, Nicola Howson, Nancy Owens and Tessa Grant from the Cochrane Schizophrenia Group for all their assistance with this review.

We would also like to thank Dr Toshiaki Furukawa of The Evidence-Based Psychiatry Centre, Department of Psychiatry, Nagoya City University Medical School, for his help with translation. With thanks also to Mr Werner Hoffmann of Novartis Pharma, Berne, Switzerland, who readily provided information for this review.

Conflict of interest

Explanation:

Any conflict of interest capable of influencing the judgements of any of the reviewers should be reported, including financial, personal, political or academic conflicts. If there are no conflicts of interest, this should be stated explicitly, by reporting 'None known'.

Style:

Beginning with 'One of the reviewers (ABC). . . .' is appropriate, as in Example 1. You can also list multiple conflicts, as indicated in Example 2.

EXAMPLE 1: *from a tardive dyskinesia review*

One of the reviewers (JJM) is a member of the following advisory boards: Janssen-Cilag Australia, Eli Lilly Australia, Lundbeck Australia. In addition JJM has been a co-investigator on studies of neuroleptic medications produced by the following companies: Astra, Janssen-Cilag, Eli Lilly, Zeneca (ICI), Sandoz and Pfizer. The same companies have provided travel and accommodation expenses for JJM to attend relevant investigator meetings and scientific symposia. No funds have been paid directly to JJM. Payments related to participation in drug trials and board attendances have been paid to a Government-audited trust account to support schizophrenia research.

EXAMPLE 2: *Olanzapine for schizophrenia*

Lorna Duggan - has attended functions sponsored by Lundbeck, Janssen, Pfizer and Zeneca and has accepted sponsorship from Eli Lilly for internal flights in the United States.

Mark Fenton - has led Janssen, Lilly and Zeneca sponsored workshops for clinicians.

Ahmed El-Dosoky - has participated in Eli Lilly sponsored research (Loza 1999).

2a.6 References

References to studies are organised under four standard headings: **included studies**, **excluded studies**, **studies awaiting assessment**, and **ongoing studies**. Other references include **additional references** that are cited in the review and **other published versions of the review**; e.g. if the review has been published in a journal. Reviewers should check their references for accuracy ([Dickersin 1986](#), [Eichorn 1987](#)).

Studies awaiting assessment: Potentially relevant studies that have been identified, but cannot be assessed for inclusion until additional data or information are obtained, should be listed here. These need not be cited in the text of the review.

Ongoing studies: Studies which are ongoing but meet the inclusion criteria should be listed here.

Additional references: Other references cited in the text should be listed here. If a report of a study is cited in the text for some reason other than referring to the study (e.g. because of some background or methodological information in the report), it should be listed here as well as under the relevant study.

Other published versions: References to other published versions of the review in a journal, textbook or *CDSR* should be listed here.

2a.7 Tables and figures

Characteristics of included studies: This is a standard table with seven columns: study ID, methods, participants, interventions, outcomes, notes and allocation concealment. Reviewers must decide what characteristics of the included studies are likely to interest users of the review. It is possible to use codes so that each column can include several subcategories of information; e.g. a reviewer could include country, setting and sex under 'participants'. Information on the funding of a study could be included under 'notes'. Footnotes should be used for explanations of any abbreviations used (these will be published in the *CDSR*). Reviewers must also include information about the 'Data source' for all included studies to indicate whether published data only, unpublished data only or a mixture were used, or if unpublished data were sought but have not been used (e.g. because they have not been obtained).

Characteristics of excluded studies: Any studies meeting the inclusion criteria, or appearing to meet the inclusion criteria, that were excluded should be identified and the reason for exclusion should be given (e.g. inappropriate control group).

Characteristics of ongoing studies: This is a standard table with seven columns: Study ID, Trial name or title, Participants, Interventions, Outcomes, Starting date, Contact information and Notes. Footnotes should be used for explanations of any abbreviations used in the table (these will be published in the *CDSR*).

Comparisons and data: A review can include more than one comparison and a study can be included in more than one of these. The comparisons should correspond to the questions or hypotheses under 'Objectives'. Data for each comparison must be entered in a standardised

format from which tables and figures for each comparison can be generated. Reviewers should try to avoid listing many comparisons or outcomes for which there are no data in the review since each comparison generates a graph even if it contains no data and analysis. Instead, reviewers should note these comparisons in the text of their review. Five types of tables are possible: dichotomous data, continuous data, individual patient data, generic inverse variance and other data.

Additional figures: From RevMan 4.2 onwards, Cochrane reviews can contain additional figures. Figures showing statistical analyses should follow the relevant guidance prepared by the Statistical Methods Group. Decisions about the suitability of pictures for inclusion in Cochrane reviews are the responsibility of the reviewers and the editors of their CRG. It is the responsibility of the reviewer(s) to obtain permission to include any figures or pictures for which the copyright is owned by someone else.

2a.8 Comments and criticisms

Summary, Reply and Contributors are subheadings in this section. The summary should be prepared by the criticisms editor for the CRG in consultation, if necessary, with the person submitting the comment. A reply to this should then be prepared by the reviewer(s). Details of the people who contributed to this process should be given. Further information on the comments and criticisms and the updating of reviews is given in section 10.11.