

Paper plus: Comparing antidepressants in primary care

Systematic reviews can be overwhelming, with too much information to take in, but they need not be. **John Fletcher** takes you through this month's paper step by step and carefully explains what it all means

John Fletcher
primary care editor
BMJ
jfletcher@bmj.com

Abstract

Objective—To compare the efficacy and tolerability of tricyclic antidepressants with selective serotonin reuptake inhibitors in depression in primary care.

Design—Systematic review and meta-analysis of randomised controlled trials.

Data sources—Register of the Cochrane Collaboration's depression, anxiety, and neurosis group. Reference lists of initial studies and other relevant review papers. Selected authors and experts.

Selection of studies—Studies had to meet minimum requirements on adequacy of sample size, adequate allocation concealment, clear description of treatment, representative source of subjects, and use of diagnostic criteria, or clear specification of inclusion criteria, details regarding number and reasons for withdrawal by group, and outcome measures described clearly or use of validated instruments.

Main outcome measures—Standardised mean difference of final mean depression scores and relative risk of response when using the clinical global impression score. Relative risk of withdrawing from treatment at any time and the number withdrawing due to side effects.

Results—11 studies (2951 participants) compared a selective serotonin reuptake inhibitor with a tricyclic antidepressant. Efficacy between selective serotonin reuptake inhibitors and tricyclics did not differ significantly (standardised weighted mean difference, fixed effects 0.07, 95% confidence interval 0.02 to 0.15; $z=1.59$, $P<0.11$). Significantly more patients receiving a tricyclic withdrew from treatment (relative risk 0.78, 95% confidence interval 0.68 to 0.90; $z=3.37$, $P<0.0007$) and withdrew specifically because of side effects (0.73, 0.60 to 0.88; $z=3.24$, $P<0.001$). Most studies included were small and supported by commercial funding. Many studies were of low methodological quality or did not present adequate data for analysis, or both, and were of short duration, typically six to eight weeks.

Conclusion—The evidence on the relative efficacy of selective serotonin reuptake inhibitors and tricyclic antidepressants in primary care is sparse and of variable quality. The study setting is likely to be an important factor in assessing the efficacy and tolerability of treatment with antidepressant drugs.

This month's paper is entitled "Efficacy and tolerability of selective serotonin reuptake inhibitors compared with tricyclic antidepressants in depression treated in primary care: systematic review and meta-analysis" (Steve MacGillivray and colleagues).

To read the paper go to this article on studentbmj.com and click on the link. You do need to refer to the paper to get the most out of the commentary.

Papers in the *BMJ* are published in ELPS (electronic long paper short) format. Clicking the link will give you a choice of which version you read. The abridged version is the short version that appeared in the paper *BMJ*; the full version is the definitive version.

Why do the review?

The investigators wanted to know if selective serotonin uptake inhibitors (SSRIs) are safe and effective for patients in primary care. At first, this seems a strange question to ask. Don't we already know that SSRIs work as well as tricyclics, have a quicker onset, have fewer side effects, and are less toxic in overdose? As is often the case, we thought we knew this but, actually, we did not. Early trials of SSRIs were done with patients recruited from psychiatric clinics. These patients have a different spectrum of depression to those seen in primary care. Depressed patients seen by psychiatrists tend to be more severely depressed and have more somatic symptoms such as weight loss and early morning wakening. Patients in primary care on the other hand are often still functioning socially and might on the surface just seem a little sad or flattened in their mood. None the less, they still need help and general practitioners often prescribe an antidepressant. What is the evidence that these patients will improve and how well do they tolerate SSRIs (box 1)?

Clearly focused subject area

The best way to tell this for most reviews is to look at the title and the last sentence of the introduction. When tackling a clinical problem, it is useful to think of the type of

Box 1: What are systematic reviews and meta-analyses?

Systematic review—A summary and discussion of literature using a systematic approach. This usually involves a clearly focused subject area, a comprehensive search for literature, explicit selection of studies, an assessment of study quality, an appropriate weighting of the articles, and a summary of what has been found

Meta-analysis—A statistical technique that combines the numerical information from individual studies to produce new insights. Most often, meta-analyses are used to simulate a larger study made up of several smaller studies so providing less uncertainty about the final result and narrower confidence intervals

patients; the intervention being studied; what it is being compared to and how the outcome was measured—remember PICO (patients, intervention, comparison, and outcome).

Here the authors have made it clear in both places that they are looking at patients in primary care. They are interested in SSRIs and are comparing them to tricyclics. The outcomes they are interested in are efficacy and tolerability.

This review starts well with a clearly focused question. Some reviews do not, and if you are not sure what is going on having read the title and the introduction of a review it is not a good sign and you should probably stop and read something else.

Comprehensive literature search

A comprehensive search would be one that picked up most of the important studies. The studies that are needed can be found in many ways and a good literature review makes use of a variety of sources. The main source for this review was the register of trials held by the depression, anxiety, and neurosis group of the Cochrane Collaboration. At first sight this may look like a rather limited search. In fact the Cochrane Collaboration do a lot of the legwork and then catalogue the studies in their register of randomised controlled trials (box 2).

Box 2: Aspects of a good search

- Several electronic databases (not just Medline)
- Visits to specialist libraries
- Discussions with experts in the field
- Searches of the references contained in the papers as they are found
- An attempt at finding unpublished literature such as PhD theses
- Consideration of articles published in foreign languages

Electronic searching is an art, and it is possible to miss important studies because the wrong search terms have been used. A good systematic review makes its search strategy available for others to check for MESH terms, explosion, and other technical aspects of literature searching.

This review is probably based on a good search for studies, though we have to trust their searching skills, and they have consulted experts and references. This all matters because it is desirable for a review to be based on a representative sample of all the available literature. The problem is that the best known and most easily obtained studies may be the ones that readers and the editors of journals find the most interesting. This is referred to as publication bias; we are all interested in breakthroughs and bored by damp squibs. When a new treatment is reported everyone wants to talk about it if there seems to be some early promise of improved results: no one does if the treatment does not work. Since large numbers of studies are published every week, some studies will produce seemingly good results even for ineffective treatments, purely on the basis of chance. These studies appear more often in a quick search and can bias the conclusions of a review.

Explicit study selection

The two main considerations in selecting studies are the subject matter and the study design. A good review will define the subject matter in sufficient detail that the reader could apply the definition and repeat the selection proce-

dure. At the start of the methods section, this review defines primary care patients as “patients being treated by a primary care practitioner (family practitioner, general practitioner) in a primary care setting and not a specialist practitioner (psychiatrist) in a secondary or tertiary setting.” They also concentrated on adults by excluding child or elderly participants. They included studies comparing SSRIs to tricyclic antidepressants for the treatment of depression. Lastly, they define their outcomes for efficacy on both a continuous scale and a dichotomous scale (only two possibilities) and their outcomes for side effects on a dichotomous scale.

Depression scores can vary over a wide range and the effect of treatment can be summarised as a change in depression score. Although this can be a sensitive measure of change it is less easy to interpret than a simple percentage of people responding to treatment, which is how a dichotomous measure is summarised. For side effects the researchers counted the number of people withdrawing from treatment. If a study did not cover the subject matter in terms of the patients, intervention, comparison, and outcome measure, then it was not included in the review.

The most appropriate study to answer a question of what works in health care is the randomised controlled trial. Other research designs can give valuable information on such questions as harmful effects (case control studies), prognosis (case series with follow up), or patient experience (qualitative studies).

This review was focused and explicit in its selection of studies. This matters because it is all too easy when conducting a literature search to be distracted by interesting looking articles that are of tangential interest or that are a second best method for addressing the subject. Without explicit methods these articles can creep into the review and bias the authors’ conclusions.

Did they assess the quality of the studies in the review?

A sound research design can still produce flawed results if it has not been carried out properly. Whilst no piece of research is perfect it is unwise to place much reliance on a study that does not meet certain minimum standards. A good review will make some attempt at judging if the research was of a good standard. It is preferable if an explicit scoring method is used. Although researchers argue about which scoring method is preferable they agree that a recognised one should be used or the criteria described clearly enough that the readers could repeat the scoring themselves.

This review uses the method described in the Cochrane Collaboration Handbook, which is a widely respected approach to the conduct of systematic reviews. This matters because poorly conducted research allows the possibility of flawed results. Many of the mistakes that are made lead to an overstatement of the effectiveness of the intervention. This is not universal but it often happens that the poorer studies are more optimistic in their results.

Weighting and summarising the research

What to look for here is a Forrest plot. Figs 2 to 5 are Forrest plots. If the review presents its findings this way it is likely they have considered the statistical issues appropriately, and most readers will rely on the journal’s statistical reviewers to do a good job. Each individual study is represented in the plot by one row. On the left of the plot are the numbers representing the summary results. The plot contains a central rectangle whose size is related to the number of patients in the study: bigger means more patients. The rectangle is centred on the overall result. The horizontal line shows the 95% confidence intervals



TINA MERANDONIPHOTONICA

for the study. A longer line means less certainty about the result or wide confidence intervals. The central vertical line of the Forrest plot represents “no difference” between treatments and distance along the horizontal axis represents increasing differences between the two treatments. The horizontal axis should have a label telling you which treatment is favoured to the left and which to the right. At the end of several rows for individual studies is a diamond shape on the plot representing the meta-analysis summary of all the studies above it, again centred on the final result and showing the 95% confidence interval.

Something else to look at on a Forrest plot is whether all the studies are consistent with each other. Whether the lines overlap each other is one way to assess this. A more formal method is to look for the test for heterogeneity statistic. If there is positive evidence that the studies are reporting different (heterogeneous) results, the P value will be significant. Hence a large P value, say more than 0.1, reassures us that the studies are likely to be all measuring the same thing. This review shows all the signs of having done the right things.

What were the findings?

Fig 1 is typical of the selection process for a systematic review. The researchers started with 284 promising titles and after applying selection criteria, reading them, and excluding irrelevant or poorly conducted studies were left with 11 studies to summarise. They presented some summary information on the included studies in the paper and reasons for exclusion of some studies in table 2. Most of the studies were of six weeks' duration. It is said that tricyclics can take this long to work and that side effects may subside with longer term treatment. If this is so then SSRIs will be at an advantage in these shorter studies.

The main results are presented in Forrest plots. Figs 2 and 3 show efficacy of SSRIs compared with tricyclics. Fig 2 uses the mean depression score. The depression scores are based on questionnaires where the answers to questions receive a numerical score and the questions are added to produce an overall depression score. To make different scales comparable they have all been standardised by centring them on the mean score and dividing all the scores by the standard deviation of the scores. Providing all the questionnaires are measuring the same thing (depression symptoms) this works but is difficult to follow. Fortunately, what we are looking for is evidence that the two treatments

are the same, and you can see that the diamonds are close to the central line of no difference on the plot.

Fig 3 is easier to interpret. The numbers show how many patients were not improved on treatment. For example in line two, Doogan and colleagues' 21/83 (25%) on SSRIs were not improved at the end of the study compared with 32/96 (33%) who did not improve on tricyclics, so the relative risk of not improving is 25/33 or 0.76 (strictly speaking the statistics are a little more involved but this is a pretty good approximation).

So the main result for efficacy shows that tricyclic antidepressants are probably slightly better than SSRIs but one end of the 95% confidence interval crosses the line of no difference meaning we cannot be sure of this difference. We can be 95% sure that the relative risk of treatment failure lies between 0.86 and 1.43 (fig 3). That is, tricyclics might be 1.43 times as good as SSRIs or only 0.86 as good—that is, worse—but neither difference in effectiveness is huge and the most likely thing is they are nearly the same.

The results for tolerability are presented in figs 4 and 5. Fig 4 shows that fewer people tended to stop their treatment when taking SSRIs than when taking tricyclics, and the summary result is that dropouts on SSRIs are 76% of those on tricyclics. This 95% confidence interval does not overlap the line of no difference so we can be confident that there is a difference. Patients may drop out for many reasons and the findings are strengthened by the results in fig 5, which give the same message for dropouts due to side effects. The authors have used the absolute difference in dropout rates, just over 7%, to calculate the number needed to treat as 14. That is to say for every 14 patients given an SSRI instead of a tricyclic, one person is prevented from dropping out of treatment.

Is this study useful?

It probably is, though your answer to that will depend on how worried you were about applying the results of studies done in a hospital setting to community patients. There are more than 100 trials comparing SSRIs with tricyclics in secondary care settings, and if you believe this evidence is good enough to apply to primary care patients then this review will not be of much interest. For those who think that depression in the community is different and that patients may have different thresholds for tolerating drugs this review of studies in primary care provides valuable evidence that SSRIs are as effective and are better tolerated than tricyclics for treating depression.