Do antidepressants work? A commentary on “Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration” by Kirsch et al

R H McAllister-Williams

The publication of this meta-analysis received a vast amount of coverage in the UK. This is despite the “bottom line” that the review does not report any novel findings—antidepressants work and their effectiveness increases with baseline severity of depression. This was not the picture painted in the media. Rather the conclusions drawn by the authors took an extreme viewpoint and the review’s publication was sensationalised both by the journal editor and the media. It is difficult to fully understand the reasons for this. Antidepressants have had a bad press in recent years over a number of issues (for example, discontinuation/withdrawal and suicidality) and the authors’ conclusions were in a similar “anti-antidepressant” vein—that they don’t work.

This is not the first time that Irvine Kirsch, Professor of Psychology at the University of Hull has caused a storm about antidepressants. A previous paper in the *BMJ* argued that antidepressants have little or no clinically significant effects. This was a similar conclusion to that of the current paper although drawn for different reasons. To understand how the findings have been sensationalised, and I believe misinterpreted, it is important to separate out the findings in the meta-analysis from the conclusions drawn by the authors.

For antidepressants, the key issue is effectiveness. There is overwhelming evidence from many studies that antidepressants are effective and that they work. The claim that antidepressants don’t work is based on a single meta-analysis of data submitted to the Food and Drug Administration (FDA) for six newer antidepressants—fluoxetine, venlafaxine, nefazadone, paroxetine, sertraline and citalopram. The FDA has a requirement that all controlled studies of a particular drug relating to a particular indication have to be submitted to them, irrespective of whether the data are positive and/or published or not.

The meta-analysis in the *BMJ* has been robustly carried out using standard methodology. Any meta-analysis of published data runs the risk of suffering from publication bias (studies with positive findings being more likely to be submitted by authors and accepted by editors for publication). In an attempt to avoid this bias, Kirsch and colleagues obtained all trial data submitted to the United States Food and Drug Administration (FDA) for six newer antidepressants—fluoxetine, venlafaxine, nefazadone, paroxetine, sertraline and citalopram. The FDA has a requirement that all controlled studies of a particular drug relating to a particular indication have to be submitted to them, irrespective of whether the data are positive and/or published or not. Data from 47 trials were obtained. However it was not possible to obtain mean endpoint depression rating scale scores (the outcome measure being used in the analysis) for five of these studies. Rather than just exclude these studies, the authors decided to exclude antidepressants for which they did not have a full data set in order to minimise the risk of bias. Thus the analysis was actually of 35 trials (5 of fluoxetine, 6 of venlafaxine, 8 of nefazadone and 16 of paroxetine).

Studies included in the analysis were all acute treatment studies of 4-8 weeks’ duration (the majority were for six weeks). It is important to note that most of the studies were done in outpatients. It is stated that two were conducted in inpatients and three in a mixture of in- and outpatients, while 39 were in outpatients. It is not clear how these numbers relate to the 35 studies included in the final meta-analysis.
It is important to understand what the primary outcomes of the meta-analysis were and how they were chosen. This relates to the authors' previous arguments in this area based around the use of categorical outcomes (for example, response vs non-response). It was argued that if the distribution of severity ratings of patients is shifted by just a small extent (for example, reduced by one point on the Hamilton Depression Rating Scale (HDRS) by the active treatment) but the cut-off for the categorical outcome is chosen carefully, then it can appear that the treatment has led to a large increase in the percentage of responders. Moncrieff and Kirsch's paper led to a deluge of correspondence challenging their conclusions (see http://www.bmj.com/cgi/eletters/331/7509/155). However, the argument is statistically correct, and is also the rationale behind the National Institute for Health and Clinical Excellence (NICE) in their depression clinical guideline deciding a priori to assess the clinical effectiveness of antidepressants by comparing mean end-point rating scale scores rather than categorical outcomes. For the difference from placebo to be deemed clinically significant by NICE it needs to be a minimum of three points on the HDRS. Alternatively, NICE said that the standardised mean difference, \( d \) (the difference between two means divided by their pooled standard deviation) had to be 0.5 or greater (a standard definition of a "moderate" effect size). These were the cut-offs used for this review although it should be remembered that they are entirely arbitrary by NICE it.

RESULTS OF THE META-ANALYSIS
In the 35 trials analysed 3292 patients had been randomised to active treatment and 1341 to placebo. This analysis showed that those treated with antidepressants had a mean decrease in HDRS of 9.60, while those treated with placebo had a decrease of 7.80 points, a difference of 1.80. This difference was statistically significant, but did not meet the NICE-defined criteria of three points for clinical significance. Likewise the standardised difference, \( d \), was 1.24 for antidepressants and 0.92 for placebo, an effect size of 0.32 and so below the cut-off of 0.5 set by NICE.

Secondary analysis demonstrated a lack of an effect of duration of treatment study or the different antidepressants, although the mean differences for the four antidepressants did vary somewhat (venlafaxine \( d = 0.42 \), paroxetine \( d = 0.47 \), fluoxetine \( d = 0.22 \) and nefazodone \( d = 0.21 \)). However there was a significant effect of baseline severity of depression on response to antidepressant versus placebo. Data are only shown in relation to the standardised mean difference scores. All the studies except one had approximate baseline Hamilton scores ranging from 23 to 30 and they showed a similar level of improvement with antidepressant (\( d \) around 1.24) irrespective of baseline severity. One outlying study with a lower baseline severity of depression (HDRS approximately 17) had a smaller \( d \) with active drug. Improvement with placebo decreased with increasing baseline severity across all studies. The difference between antidepressant and placebo was projected to exceed the NICE criteria of \( d \geq 0.5 \) when baseline severity exceeded a HDRS score of 28.

Kirsch and colleagues explored the publications relating to the data they obtained from the FDA. For those studies that had been published (at least once) there were frequent inconsistencies in the reported data with that submitted to the FDA, mainly minor differences in the number of patients reported in the studies.

CONCLUSIONS DRAWN BY THE AUTHORS
The authors state that they find that "efficacy reaches clinical significance only in trials involving the most extremely depressed patients, and that this pattern is due to a decrease in the response to placebo rather than an increase in the response to medication." Their description of the very severe nature of the depression of the patients included in the analysis is based on American Psychiatric Association (APA) definitions of cut-off scores for the HDRS with 25 or more being "very severe" depression.

COMMENTS ON THE AUTHORS' CONCLUSIONS
Undoubtedly the findings in this analysis are robust, as far as the studies included in the analysis are concerned. The choice of the data set was based on logical reasoning in trying to avoid publication bias. However it does not include all possible data from studies completed subsequent to FDA submissions. Nevertheless, in line with many previous analyses (including NICE's own), the meta-analysis demonstrates that antidepressants are significantly better than placebo. Further, in line with previous evidence, the drug-placebo difference increases with increasing severity of baseline illness. The conclusion that this is due to a decrease in response to placebo rather than an increase in effectiveness of the drug is entirely fallacious because the magnitude of the therapeutic effect is the difference between active drug and placebo, not the absolute response to active drug.

The subsequent arguments around the conclusions drawn then come down to semantics: how is severity of depression and clinical significance defined. Although Kirsch and colleagues use the APA definition of "very severe" depression (HDRS >23), this is not a universally accepted cut-off. Others have argued that an HDRS of at least 30 indicates severe depression. Whatever the rights and wrongs of this debate, it is hard to accept the language used in the paper's conclusions that only "the most extremely depressed patients" showed a clinically significant response to antidepressants. At most only two out of 35 trials were conducted in inpatients, and suicidality is a virtually universal exclusion criterion from a clinical trial of antidepressants. Many of the studies included in the analysis were conducted in the USA. Such studies rely heavily on patients recruited by advertisement and who may join the study in order to receive free medication and care. It has been argued that there is a tendency in such trials for inflation of baseline HDRS scores so as to allow patients to meet the study entry criteria.

With regard to clinical significance, NICE a priori set a minimum antidepressant-placebo mean end-point difference of three points on the HDRS. As already stated, this figure is entirely arbitrary and not based on any evidence other than the guideline development group's view that anything less might be hard to see in an individual patient in the clinic. However there are many problems with applying this criterion to data obtained from randomised controlled trials (RCTs). RCTs are simply experimental tools used to test hypotheses—they are not well designed to assess clinical effectiveness. The reason for this is that a response to a drug (that is, change in rating scale score at the end of a trial compared to the baseline) in an RCT results from many processes. Extreme values at baseline can regress to the mean, there can be measurement bias and the illness itself might spontaneously improve. An RCT controls for these through the process of randomisation. Improvement can also result from patients (and doctors) being conditioned to expect a response and attaching meaning to any "side effect" as implying the person is on active drug. These processes,
Correspondence to: Professor J C Markowitz, Research Psychiatrist, New York State Psychiatric Institute; Clinical Professor of Psychiatry, Weill Medical College of Cornell University, New York, USA; jcmv42@columbia.edu

A letter from America: rescuing inpatient psychiatry

Professor J C Markowitz

In recent decades, American insurance companies have severely restricted medical reimbursement and access to treatment. Within this general trend, they have discriminated against psychiatry more than most other specialties, and have slashed in particular the most expensive facet of psychiatric care: inpatient treatment.¹ The mean length of psychiatric stay in American hospitals has fallen from months to a handful of days. Realistically, how effectively can clinicians treat severely ill psychiatric patients in a three or four day hospital stay? Meanwhile, reimbursement for inpatient care has fallen so low (in some cases to 39% less than actual costs) that American hospitals have begun reducing psychiatric beds as well as lengths of stay.² Patients in managed care plans are less likely to receive inpatient care than patients in fee-for-service plans.³

This drastic change in inpatient psychiatric care has had little to do with its clinical effectiveness and much to do with the managed care mantra of cost cutting. The promise that increased reimbursement for less costly day hospitals and outpatient services would compensate for this inpatient decline has proven false. This unholy precedent is being echoed in other countries as governments try to conserve medical funds.⁴

Managed care, alas, is a euphemism for managed cost. Caring for patients is beside the point. Its goal is to save money and to earn money for insurance “stakeholders” rather than to provide optimal treatment for “consumers” (the erstwhile patients). The underlying philosophy is hard-nosed...
Do antidepressants work? A commentary on "Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration" by Kirsch et al

R H McAllister-Williams

Evid Based Mental Health 2008 11: 66-68
doi: 10.1136/ebmh.11.3.66

Updated information and services can be found at:
http://ebmh.bmj.com/content/11/3/66

These include:

References
This article cites 5 articles, 1 of which you can access for free at:
http://ebmh.bmj.com/content/11/3/66#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
EBMH Notebook (73)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/