Misinterpretation of randomized trial evidence: do antidepressants work?

On February 29 this year I was due to give a talk about the pharmacological management of depression to a large meeting of psychiatrists. As my radio alarm clock came on I heard the top headline on the BBC Radio 4 Today programme 'New study suggests that antidepressants don't work'. The headline arose from a meta-analysis published by Irvine Kirsch, Professor of Psychology at the University of Hull, and colleagues (Kirsch et al, 2008).

This is not the first time Professor Kirsch has induced a storm about antidepressants. A previous paper in the *BMJ* (Moncrieff and Kirsch, 2005) arguing that antidepressants had little or no clinically significant effects led to a deluge of correspondence from doctors challenging the conclusions and saying they would continue to prescribe antidepressants (www.bmj.com/cgi/eletters/331/7509/155#112381). I must make a declaration of interest at this stage. I prescribe antidepressants. I believe that they do work and this is what I said in my lecture in February even after having read Professor Kirsch's recent paper.

Clinical *vs* statistical significance

At the heart of the argument is how clinical, as opposed to statistical, significance is judged. Clinical trials most commonly assess severity of depression using scales such as the Hamilton Depression Rating Scale (HDRS; Hamilton, 1967). In 2005, Moncrieff and Kirsch argued that categorical outcomes (e.g. response *vs* non-response) can be misleading. If the distribution of severity ratings is moved by just a small extent (e.g. reduced by one HDRS point by the active treatment), and the cut-off for the categorical outcome is chosen carefully, then it can appear that the treatment led to a large increase in the percentage of responders (Moncrieff and Kirsch, 2005). However, categorical outcomes in depression are generally standardized with regulatory bodies accepting a definition of response as a 50% decrease in HDRS score. Nevertheless Montcrieff and Kirsch's argument is statistically correct, and for the same reasons the National Institute for Clinical Excellence (NICE), when reviewing the evidence base in preparing their depression guidelines (National Institute for Clinical Excellence, 2004), decided that the primary measure they would assess was the difference in mean end point rating scale scores, rather than response or remission rates.

NICE *a priori* set a minimum antidepressant-placebo mean end point difference of three points on the HDRS as the cut-off for clinical significance. This figure is entirely arbitrary and not based on evidence other than the guideline development group's view that anything less might be hard to see in an individual patient in the clinic.

Assessing clinical effectiveness

However, there are many problems in using randomized controlled trial data to determine clinical effects sizes. Randomized controlled trials 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 (i.e. change in rating scale score at the end of a trial compared to the baseline) in a randomized controlled trial results from many processes.

Extreme values at baseline can 'regress to the mean', there can be measurement bias and the disorder itself might spontaneously improve. A randomized controlled trial controls for these through the processes of randomization. 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 an active drug. These processes, in essence, are the 'placebo effect' and in a randomized controlled trial are controlled for through blinding of patients and raters.

Finally, for the active drug alone, there is the specific therapeutic effect. This is calculated by subtracting the response to placebo from the response to the drug. The null hypothesis is that this difference is zero. To be statistically robust randomized controlled trial data are analysed on an 'intention to treat' basis including all randomized patients. This leads to small effect sizes compared to a 'per protocol' analysis that looks at the outcomes of patients who complete the study. However, in clinical practice we are concerned about how much benefit a patient will get from a drug if he/she takes it for long enough (safety is a separate issue). Further, it matters little whether the patient responds because of a placebo effect or the specific pharmacological actions of the drug, as long as he/she gets better. This is because it is ethically not possible to prescribe placebo (which also don't work well unless the doctor also believes the prescription is an active drug – Gracely et al, 1985).

Note that most psychological therapies have been assessed by comparing outcomes with 'waiting list controls' or 'treatment as usual'. This combines the therapeutic effect of the therapy with any placebo response. If antidepressants were similarly assessed the difference in HDRS end point scores would be way in excess of three points in pretty much all randomized controlled trials.

Analysis of trial data

Professor Kirsch's recent meta-analysis used data obtained from that submitted to the Food and Drug Administration (FDA) in the USA for fluoxetine (five studies), paroxetine (16 studies), venlafaxine (six studies) and nefazadone (eight studies). The primary assessment of the combined data confirmed that actually these drugs are effective antidepressants in that there was a highly significant difference in mean HDRS end point score for active drug *vs* placebo (Kirsch et al, 2008).

Despite the limitations of using randomized controlled trials to assess clinical effect sizes, Kirsch and his colleagues argued that these studies demonstrate that antidepressants do not have a clinically sig-

nificant effect in depressed patients because the difference between drug and placebo was 1.8 HDRS points, not meeting NICE's arbitrary criteria of three. Further analysis suggested that the difference only exceeded three in studies where the baseline severity of patients' depression was 'very severe' (probably more akin to moderate to severe depression as defined by NICE and where antidepressants are recommended to be used), the authors concluding that this was the result of a decrease in the placebo response rather than an increase in the therapeutic effect of the active drugs. This conclusion is entirely fallacious since the magnitude of the therapeutic effect is the difference between active drug and placebo, not the absolute response to active drug.

What was all the fuss about?

In fact Kirsch's meta-analysis simply confirms that antidepressants do work, as well as previous findings that placebo–drug differences increase with increasing severity of baseline depression (Khan et al, 2002). So why was there such a media storm over this publication? Antidepressants, particularly the selective serotonin-reuptake inhibitors, have become the new whipping boys in psychopharmacology, superseding the benzodiazepines for reasons that are not clear (Nutt and Malizia, 2008).

Another issue that the media focused on when reporting the Kirsch paper was the non-disclosure of negative trials by pharmaceutical companies, with an implication that obtaining such data was behind the 'new' finding that antidepressants 'don't work'. It is certainly true that through industry's desire to hide bad news and editors' lack of interest in publishing negative studies problems have arisen. However, regulatory authorities are requiring registration of all studies at their outset so that negative ones can't be 'lost'. In addition many companies are now making all of their data available via the internet and regulators and independent academics need to keep up pressure for them to do this.

It should also be noted that contrary to the media's implications, none of the data used by Kirsch and colleagues was 'hidden' – they obtained it from the FDA to whom it had been submitted by the relevant companies. Further, the FDA had analysed all of this data and came to the conclusion that not only did the drugs work but that the effect was significant enough for them to grant marketing licences and the drugs to be prescribed to patients. I for one will continue to do so. **BJHM**

KEY POINTS

- Randomized controlled trials are tools to assess whether or not drugs work.
- Analysis of trials of antidepressant drugs confirms that they work.
- Randomized controlled trials are poor ways of assessing clinical effectiveness of drugs.
- Recent media scares should not dissuade clinicians from using antidepressants to treat depression in line with National Institute for Health and Clinical Excellence guidelines.

R Hamish McAllister-Williams
Reader in Clinical Psychopharmacology and Honorary Consultant Psychiatrist
Psychiatry
Royal Victoria Infirmary
Newcastle upon Tyne NE1 4LP

Gracely RH, Dubner R, Deeter WR, Wolskee PJ (1985) Clinicians' expectations influence placebo analgesia. *Lancet* **i**(8419): 43

Hamilton M (1967) Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* **6**(4): 278–96

Khan A, Leventhal RM, Khan SR, Brown WA (2002) Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* **22**(1): 40–5

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008) Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* **5**(2): e45

Moncrieff J, Kirsch I (2005) Efficacy of antidepressants in adults. *BMJ* **331**(7509): 155–7

National Institute for Clinical Excellence (2004) *Depression: Management of depression in primary and secondary care*. Clinical Guideline 23. National Institute for Health and Clinical Excellence, London

Nutt D, Malizia A (2008) Why does the world have such a 'down' on antidepressants? *J Psychopharmacol* (in press)