Ethical Human Psychology and Psychiatry, Volume 9, Number 3, 2007

The New Anecdotes

David Healy, MD, FRCPsych
North Wales Department of Psychological Medicine

This article deals with the interpretation and clinical implications of findings from randomized controlled trials (RCTs) of antidepressant drug efficacy. It will also deal with a related issue, namely the differential response of journals to articles reporting individual clinical case outcomes compared with articles reporting controlled trial outcomes.

Keywords: antidepressant effects; randomized controlled trials; case studies; suicidality

Multiple meta-analyses suggest that antidepressant randomized controlled trials (RCTs) show roughly 50% of patients responding positively on a rating scale measure as compared with 40% on placebo (Kirsch & Sapirstein, 1998; Kirsch, Moore, Scoboria, & Nicholls, 2002; Stone & Jones, 2006). These outcomes are represented in Figure 1. When unpublished studies are included, the data presented by the FDA in 2006 suggest a placebo response rate as high as 40%.

A statistically significant improvement compared with placebo is taken to indicate that the drug works. Regulators are obliged to approve such drugs, drug companies market them as effective, and clinicians prescribe them. As a result of such trials, the money, research, and culture in psychiatry has been based heavily on use of antidepressants. That is, clinical practice follows the dark column, and this generates the problems outlined below, which are likely to hold equally true for other areas of medicine in which the differences between the response rates to active drug and placebo are of the magnitude found in antidepressant trials.

While the above data can be read as evidence that the drugs "work," another option is to read the data as offering the possibility of quantifying the portion of the therapeutic response that can be attributed to a specific treatment (Paykel, 1988).

It is known that certain factors, such as the natural history of depression, mean that many people will improve within a few weeks whether they are treated or not. It is also widely thought that sensible advice from a clinician on matters of diet, lifestyle, alcohol intake, as well as work and relationship stress may make a difference. It is suspected that patient perceptions that they are being seen and cared for by a medical expert may make a difference, and this effect may be enhanced by being given a substance that they think will restore some chemical balance to normal—even if that imbalance is mythical and the substance is a placebo. The fact that the patients present themselves for treatment may make a difference. All of these factors are reflected in the placebo response, but it is very difficult to quantify the distinct contribution of these various components.

All of the same factors also contribute to the therapeutic response for those patients on an active drug. In contrast to the difficulties in quantifying the components of the placebo

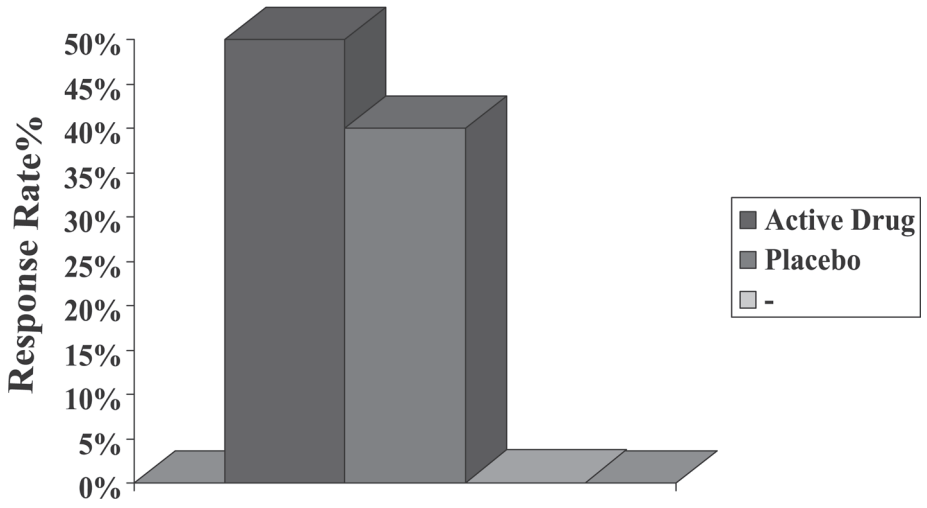

Figure 1. Drug versus placebo.

response, however, RCTs allow us to quantify the contribution made by the drug. In the case outlined above, the drug effect is 50% – 40% = 10%. That is, four out of five, or 80%, of the treated patients who improve would have improved had they received the placebo. Expressed as the number of patients needed to treat (NNT) in order for one patient to realize a specific benefit from drug treatment, the reciprocal of the risk difference, NNT = 1/10% = 10.

The rate of improvement among people on placebo is 40% (Figure 2). The NNT for the placebo group, therefore, is 1/40% = 2.5. If the culture, money, and research in psychiatry and in disciplines faced with comparable data are to follow the evidence, there is a case for saying they should follow the pale rather than the dark column or, alternately, that more effort should be put into determining what factors make for a specific drug response rather than attempting to ensure that people are treated indiscriminately on the basis that the drugs work.

If medicine were set-up to deliver the "pale" response, we would achieve only 80% of the benefit realized by the treated groups in these trials. Treatment decisions, however, should be based not only on benefits, but also on weighing the benefits against treatment-induced harm. As the placebo response is associated with fewer adverse events, it offers a safety advantage to mitigate any loss of benefit from a specific treatment. In fact, the data above are not inconsistent with the possibility that in real life the drugs may do more harm than good.

In practice, psychiatry behaves as though the data from Figure 1 look like the data from Figure 3. This offers a notional set of outcomes for a treatment like penicillin for a condition like fulminating pneumonia. In this case, patients, physicians, and all others would want practice to follow the dark column and not the pale column. Physicians with an excellent bedside manner who failed to prescribe penicillin would be likely to be sued.

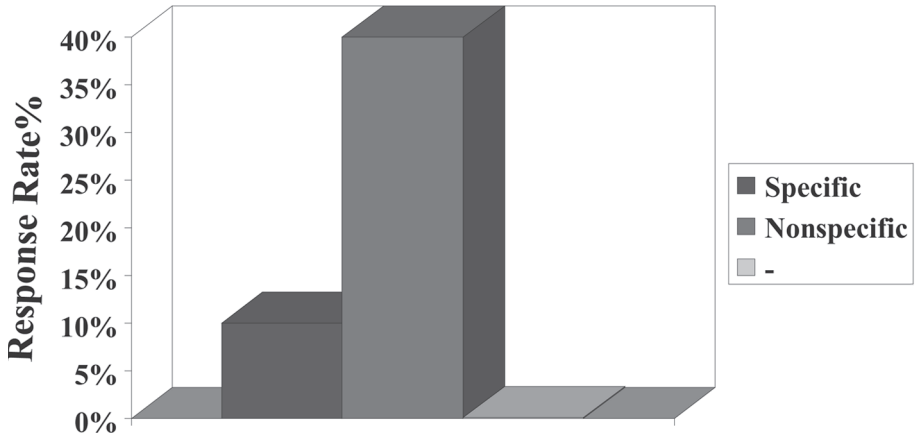

Figure 2. Components of therapeutic response: Specific drug versus nonspecific placebo.

MEDICAL JOURNALS

One of the first randomized and parallel group placebo controlled trials in medicine compared reserpine and placebo in a group of anxious depressives (Davies & Shepherd, 1955). The results showed reserpine to be effective in this population, with a treatment effect size as large as that later shown by fluoxetine or other antidepressants in this patient group. But although published in the *Lancet*, this RCT had almost no impact.

Reserpine instead developed a reputation for causing depression on the basis that it triggered suicides. The reported suicides were in people taking reserpine for hypertension, and these may have been mediated through treatment-induced akathisia (Healy & Savage, 1998). Had Davies and Shepherd's trial registered more widely, and reserpine been recognized as an antidepressant, we might never had had the selective serotonin reuptake inhibitors (SSRIs), or at least would not have been likely to have views that depression involved a lowering of serotonin that these drugs normalized, as reserpine lowers serotonin and other monoamines.

When asked later why this trial failed to have an impact, Shepherd suggested that clinicians were unused to seeing medical data presented in this way. They were used to being presented with case reports (Shepherd, 1998). The discoveries of new benefits or new hazards of drugs up till then had all been presented in terms of case reports. The first paper on lithium outlined 10 cases given the drug (Cade, 1949). The first paper on the antidepressant imipramine outlined the effects on 40 patients given the drug (Kuhn, 1958). The first papers on the antipsychotic chlorpromazine outlined its effects on a series of patients (Delay, Deniker, & Harl, 1952).

And the two papers preceding Davies and Shepherd's reserpine trial were case series of patients becoming agitated and suicidal on reserpine (Smirk & McQueen, 1955; Wallace, 1955). These articles trumped Davies and Shepherd's RCT. There is no reason to think that either the findings of the case reports or those of the RCT, even though superficially contradictory, are wrong. Indeed, there is a close parallel with early trials and case reports of suicidality for fluoxetine.

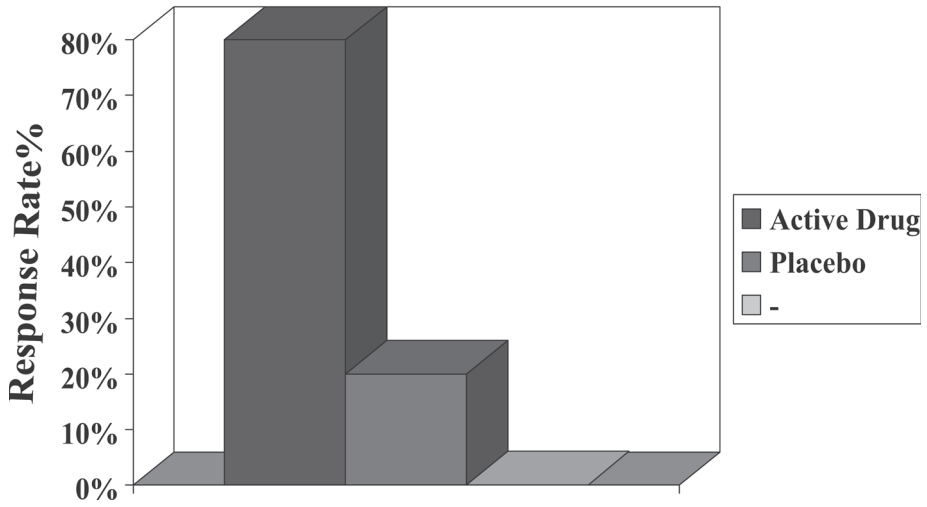

Figure 3. Penicillin versus placebo.

Since 1955, the pendulum has swung in favor of RCTs. The leading journals in medicine devote their space to RCTs and all but refuse to take case reports. The Cade, Kuhn, and Delay papers might not now be published in a major journal. While economic considerations linked to the potential for purchases of reprint requests may play some part in the changing character of clinical journals, the dominant factor is a perception that the evidence from RCTs trumps all other sorts of evidence, especially that from case reports. This is the case, even when case reports come from the most senior clinicians in the field and outline effects that follow challenge-dechallenge and rechallenge paradigms and when these effects are explicable in terms of known pathological mechanisms (Teicher, Glod, & Cole, 1990).

Far from being pleased with the growing "scientific" character of their journals, clinicians can be heard to moan that their journals have become sterile. While the impact factor of the *British Journal of Psychiatry* is greater than that of its stable mate *Psychiatric Bulletin*, many clinicians confess they are more likely to read *Psychiatric Bulletin*. There is perhaps something of an assumption among leading journals that these clinicians are failing to come to grips with evidence-based medicine. But perhaps there is a greater philosophical justification for this recalcitrant position than is often conceded.

If we return to the data in Figure 1, we see that of 5 people responding to an antidepressant 1 responds specifically to the drug while the other 4 would have responded to placebo. Thus in any sample of 10 patients, with drugs like the antidepressants, 1 responds to the drug while 9 do not. In preferentially accepting RCTs—as they are currently framed and interpreted—over case reports, journals risk privileging the experiences of the 1 specific drug responder over the 9-fold larger pool of other responders or nonresponders. This is laid out schematically in Figure 4.

This appears to be a swing to a new form of anecdotalism. This swing is not without consequences for both journals and clinicians. For journals, there remains the fact that the first discoveries of a new drug benefit or hazard are more likely to come in the form of a case report than in an RCT. In focussing on RCTs only, journals risk missing out on

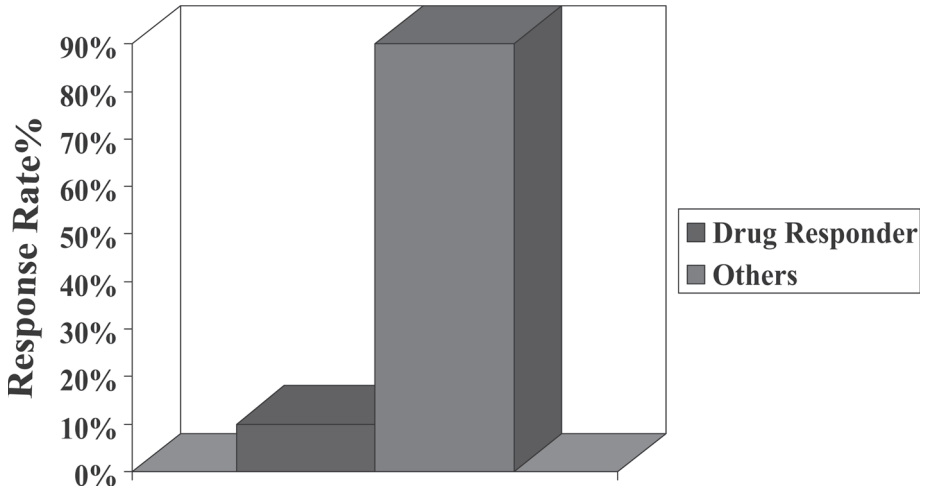

Figure 4. Specific drug responders versus others.

breakthrough papers. They also risk losing their readership, as the material fails to stimulate clinicians.

For clinicians, there is a further problem. We have moved away from a world in which clinicians were slow to use new drugs and when they did so, if their patients responded paradoxically, they stopped the treatment and wrote up the outcomes. But now driven by evidence that is less generalizable than commonly thought, clinicians rapidly take up the newest treatments. Faced with patients who turn suicidal, for instance, they consult the RCT evidence base that will commonly not list such effects, or may list them under codes such as emotional lability, which few clinicians will realize means suicidality. Failing to see evidence of a hazard, the clinician in this case may even double the dose of the new agent.

FROM ANECDOTES TO DATA

One reason that antidepressants have been so commercially successful is that their lack of generalizable efficacy and their hazards are not apparent in journal articles. Efficacy results are reported as *significantly better than placebo* and safety results are reported as *not statistically different from placebo*, statements that are both true and misleading (Healy, 2006a, 2006b).

An alternative is to report benefits in terms of odds ratios and confidence intervals to quantify the magnitude of an effect instead of reporting benefits in terms of the dichotomous classification of yes/no created by significance tests. If antidepressant results had been presented in this fashion, the most likely point estimate in individual trials would have been circa 1.5 with confidence intervals for a majority of trials that included 1.0. Recent FDA reviews of all antidepressant studies show that the confidence interval for all depression trials in children and adolescents straddles 1.0. The odds ratio for a benefit over placebo in 18–25-year-olds is 1.54 (95% C.I., 1.34, 1.76), for 25–64-year-olds is 1.84 (95% C.I., 1.77, 1.93) and for 65 and over is 1.39 (95% C.I., 1.24, 1.57) (Stone & Jones, 2006).

This style of reporting might be enhanced by two additional considerations. First, given that companies and others now recognize that selective reporting of trials is in its own right a new anecdotalism, and there is agreement in principle to register all clinical trials, it should be possible to have the confidence interval for an individual trial presented alongside a revised odds ratio and confidence interval for all trials carried out for that agent in that condition. Where trials are registered but the results are unavailable, the results from these trials could be incorporated by assigning an odds ratio of 1.0, along with an appropriate confidence interval, and a number of subjects equal to the mean number drawn from available trials.

Second, all articles reporting clinical trials could contain a paragraph or footnote, which gives the odds ratio for the benefit of a drug like penicillin for a condition like pneumonia, along with its confidence interval. This may make it clear that a benefit in a clinical trial does not mean the drug in question is comparable to penicillin for pneumonia.

If the results from antidepressant trials had been presented in this fashion, it is likely that most confidence intervals for individual trials would have been broad and might have included 1.0. This would have indicated not that the findings are not significant and should have been disregarded, but rather that the treatment has benefits but that further scientific input was needed to specify the characteristics of responders and nonresponders. Findings presented in this way would also have offered scientific support for a presentation of case reports that with appropriate controls, such as challenge and dechallenge, might have made it clear that new drugs, as reserpine and fluoxetine once were, even when effective for some might trigger clinical deterioration and even an outcome like suicide in others.

REFERENCES

Cade, J. F. J. (1949). Lithium salts in the treatment of psychotic excitement. *Medical Journal of Australia, 2,* 349–353.

Davies, D. L., & Shepherd, M. (1955). Reserpine in the treatment of anxious and depressed patients. *Lancet, 1,* 117–121.

Delay, J., Deniker, P., & Harl, J. M. (1952). Utilisation en thérapeutique psychiatrique d'une phénothiazine d'action centrale elective. *Annales Medico-Psychologique, 110,* 112–131.

Healy, D. (2006a). Did regulators fail over selective serotonin reuptake inhibitors. *British Medical Journal, 333,* 92–95.

Healy, D. (2006b). The antidepressant tale. Figures signifying nothing? *Advances in Psychiatric Treatment, 12,* 320–328.

Healy, D., & Savage, M. (1998). Reserpine exhumed. *British Journal of Psychiatry, 172,* 376–378.

Kirsch, I., & Sapirstein, G. (1998). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention and Treatment, 1.* Retrieved July 15, 2002, from http://www.journals.apa.org/prevention/volume1/pre0010002a.html

Kirsch, I., Moore, T. J., Scoboria, A., & Nicholls, S. S. (2002). The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention and Treatment, 5.* Retrieved July 15, 2002, from http://www.journals.apa.org/prevention/volume5/pre0050023a.html.

Kuhn, R. (1958). The treatment of depressive states with G22355 (imipramine hydrochloride). *American Journal of Psychiatry, 115,* 459–464.

Paykel, E. S. (1988). Antidepressants: Their efficacy and place in therapy. *Journal of Psychopharmacology, 2,* 105–118.

Shepherd, M. (1998). Psychopharmacology: Specific and non-specific. In D. Healy (Ed.), *The Psychopharmacologists* (Vol. 2, pp. 237–257). London: Arnold.

Smirk, F. H., & McQueen, E. G. (1955). Comparison of rescinamine and reserpine as hypotensive agents. *Lancet, 1,* 115–116.

Stone, M., & Jones, L. (2006). *Clinical review: Relationship between antidepressant drugs and adult suicidality*. Retrieved December 20, 2006, from http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006–4272b1-index.htm

Teicher, M. H., Glod, C., & Cole, J. O. (1990). Emergence of intense suicidal preoccupation during fluoxetine treatment. *American Journal of Psychiatry, 147,* 207–210.

Wallace, D. C. (1955). Treatment of hypertension. Hypotensive drugs and mental changes. *Lancet, 1,* 116–117.

Acknowledgments. In recent years the author has had consultancies with, been a principal investigator or clinical trialist for, been a chairman or speaker at international symposia for, or been in receipt of support to attend meetings from all major pharmaceutical companies. In the past 5 years he has been an expert witness for the plaintiff in legal actions involving SSRIs and has been consulted on a number of attempted suicide, suicide, and suicide-homicide cases following antidepressant medication, in most of which he has offered the view that the treatment was not involved.

Correspondence regarding this article should be directed to David Healy, MD, FRCPsych, North Wales Department of Psychological Medicine, Cardiff University, Wales LL57 2PW. E-mail: Healy_Hergest@compuserve.com