

---

# The dilemmas posed by new and fashionable treatments

David Healy

---

A new drug is introduced to the market. It has been approved (after stringent scrutiny) by regulators, who require ever more convincing evidence for safety and efficacy. Aside from the increased costs of the new treatment compared with the old, what could the problems be?

Even on the cost front, many would argue that there is little cause for concern. We have entered an era where placebo-controlled clinical trials demonstrate that new treatments work, in contrast to the demonstrations of efficacy of the sort available for earlier treatments. There is general agreement that, if we were to operate only in accordance with the demonstrations of efficacy from clinical trials of the type now done, the health services would be more effective and efficient and, ultimately, costs would fall. Furthermore, the use of many new drugs in recent years has appeared to be justified by economic models based on figures from clinical trials and a range of assumptions, such that a new antidepressant or antipsychotic costing several thousand pounds a year can be transformed (by costs offsets) into a treatment that is less expensive than that using an older agent costing £50 per year or less.

---

## Treatment effects and treatment effectiveness

---

There are many problems with this scenario, however. When they were introduced, randomised controlled trials (RCTs) were a significant step forward in terms of evaluative technologies for new treatments. The assumption of a null hypothesis

means that their primary purpose was to show that treatments did not work – to stop therapeutic bandwagons in their tracks. Within psychiatry, for instance, the first RCTs demonstrated that cortisone did not work for schizophrenia (Rees, 1997). A recent illustration of this function of RCTs lies in demonstrations that debriefing, which had all but become a social movement (Raphael *et al*, 1995), does not work – at least when given indiscriminately (Bisson *et al*, 1997).

What RCTs did historically was to demonstrate to the opponents of treatments such as chlorpromazine that the first antipsychotics did have some treatment effect, whatever these critics might still think about the overall benefits. Now, in complete contrast to the original intentions behind their use, RCT evidence is used to fuel therapeutic bandwagons. It is sold as evidence that the treatment works (actually does good) rather than evidence that treatments have an effect (which may be put to good use in judicious hands). There is no philosophical or methodological basis for this development.

Randomised controlled trials originated within epidemiology. Some epidemiologists had and continue to have considerable misgivings about the capacity of randomisation to overcome the problems of external validity that result from the sampling methods adopted by this approach. The alternative is to use large simple trials with 'hard' end-points such as mortality (Healy, 1997). The problems inherent in RCTs are compounded in company-sponsored RCTs, which explicitly recruit samples of convenience. This approach offers internal validity, in the sense of providing an assay system that detects a treatment effect, but the external validity of these samples remains unclear. As a

result, a majority of current trials in any area of medicine have the power to disconfirm the null hypothesis that the treatment does not differentiate from placebo, but this evidence does not support extrapolations to the likely effectiveness of treatment. Such extrapolations at present can be based only on clinical judgement.

Distinctions between treatment effects and effectiveness are a particular problem in the case of clinical trials in psychiatry, where the end-points of treatment are surrogate ones based on changes in rating scale scores rather than demonstrations of return to work, reduced mortality or absent bacterial loads. There are four potential domains of measurement: observer-based disease-specific rating scales, such as the Hamilton Rating Scale for Depression (HRSD); patient-based disease-specific rating scales, such as the Beck Depression Inventory (BDI); observer-based non-disease-specific scales of global functioning; and patient-based non-disease-specific scales of global functioning (Quality of Life scale, QoL). It might be possible to provide better estimates of therapeutic effectiveness if a clear treatment effect could be demonstrated on ratings scales from all four domains. As a matter of fact, however, there is not a single antipsychotic or antidepressant that has been demonstrated to have treatment effects across all these domains. In the case of the antidepressants, demonstrations of treatment effects have largely been on the basis of instruments from the first domain. The work of Weissman *et al* (1974) on social adaptation shows that while antidepressants may lead to symptomatic improvements, the broader functioning of the patient may not normalise for a long time afterwards. In the case of trials with the selective serotonin reuptake inhibitor (SSRI) antidepressants, QoL scales have been used in as many as 100 trials with data from fewer than 10 reported (Healy, 2000).

Typically, outcomes in one domain, such as a 50% drop in an HRSD score, are presented as evidence of treatment effectiveness. This is clear evidence of a treatment effect but it does not necessarily support claims for efficacy, for example if the drop is from 38 to 19. Were convincing scores on rating scales across the range of domains of measurement available, there would still remain the problem of factoring recent evidence of discontinuation syndromes (Viguera *et al*, 1997; Tranter & Healy, 1998) into any extrapolation from demonstrations of treatment effects to claims for treatment effectiveness. If the stabilised patient relapses on discontinuation, the final outcome may be worse than non-treatment. Treating and stopping treatment is in general not the same as not treating in the first instance, and we rarely know sufficient about the natural history of either the treated or untreated states that we manage to make any confident claims for efficacy.

There are further problems with the current evidence base. Derived from epidemiological studies, RCTs essentially provide evidence of associations. But as in studies of smoking and lung cancer or diet and cardiac disorders, for instance, such evidence points to a link between events rather than an explanation of how or why these events may be linked. Indeed, arguably, epidemiological studies of this type, which link drugs to a therapeutic outcome, have obscured the mechanisms whereby these events are linked by deflecting our attention away from what the drug actually does to bring about the association. For example, Jick *et al* (1995) compared suicide rates after 172 000 prescriptions for antidepressants in primary care in the UK and found a higher rate on fluoxetine than on other antidepressants – but this study necessarily left uncertain the nature of the mechanism producing this association.

In the case of the antidepressants, clinical trials may suggest to the unwary that a group of pharmacologically diverse agents, which almost certainly bring about their benefits by producing distinctive functional effects, produce common treatment effects. The SSRIs were in fact synthesised in the first instance to do something functionally (not biochemically) different from the older tricyclic agents. Interpreting the trial evidence as evidence that these agents all 'work' diverts attention from the question of how they are working. Through what functional effects does a noradrenergic selective agent bring about its benefits compared with an SSRI? Preclinical work indicates that one set of drugs is energy enhancing, while the other is more serenic (anxiolytic). But our recent mesmerised focus on RCTs has obscured these distinctions in clinical practice. Prescribing without knowing what potentially beneficial effects an agent produces is not likely to lead to either rational or good practice. If we do not know what these diverse agents do to get patients with depression better, how can we know which of them to give the patient in front of us?

The discussion so far has focused on the relatively simple case of depression. The apparently clear-cut effects on HRSD scores in short-term trials of these agents have contributed to the impression that it is possible to assess the efficacy of our treatments in complex conditions such as manic-depressive disease or schizophrenia. But consider the problems in bipolar disorders. No single rating scale can be used in a condition that cycles from one pole to its polar opposite. If we use frequency of episodes as an end-point, thousands of patients would have to be recruited across multiple centres and sustained within an experimental protocol for years in order to produce a convincing demonstration of prophylaxis. This cannot be simply done. Even the

resources of the largest pharmaceutical companies have not been able to support trials like this. As a result, the use of anticonvulsants, sometimes called mood stabilisers, in mood disorders is underpinned by evidence of a treatment effect in depression or in mania but not evidence of effects in manic-depressive disease. In the same way, there is little evidence on the extent to which antipsychotics work for schizophrenia over and above their treatment effect in acute psychotic states and in some maintenance studies.

There are further problems for anyone who wishes to go beyond the statement that treatment effects can be demonstrated to a claim that treatments have been shown to work. In placebo-controlled clinical trials, the placebo also has effects. In short-term trials, based on changes in rating scale scores, it becomes a practical impossibility to abstract the placebo component of efficacy from any specific component of efficacy and determine how much, if at all, the 'active' treatment is 'working'. While there may be efficacy in some patients, in general, superiority over placebo in a clinical trial is a demonstration of an effect rather than a demonstration of efficacy.

Or consider the case of the hypnotics. RCT evidence may show that a hypnotic has a clear effect without any need to employ a rating scale. Patients, however, may not wish to take such treatments. In this sense, despite evidence that the treatment can be said to work in one dimension of value, this hypnotic does not work for a subgroup of patients in other dimensions. Further trials are called for, to establish how much such a treatment is valued, but these are never undertaken for hypnotics. In the case of sleep and hypnotics, however, people are probably confident enough in their own judgement to ignore their clinician or any expert if need be. In the case of anxiety, depression, manic-depression or schizophrenia, the situation is more ambiguous and it behoves the clinician, on behalf of the patient, to know how much treatments actually are valued. But there is no evidence of this sort.

---

## Marketing the evidence

---

The problems outlined above are in a very real sense academic. In the real world, the problems with the evidence facing clinicians are even graver. First, clinical trials that do not favour a company's interest are frequently not reported. This leads to a situation where the greatest single determinant of outcome of a published study appears to be its sponsorship (Freemantle *et al.* 2000; Gilbody & Song, 2000). Second, as mentioned above, there is no

obligation on companies to report all the data from within trials that are published. In the case of the SSRIs, for example, there has been almost universal non-reporting of QoL data (Healy, 2000). Finally, there is an overreporting of favourable studies. At international meetings and in peer-reviewed journals, senior experts in the field, who have had no participation in a study, present data from company trials in a manner that leaves those attempting to meta-analyse the results confused as to how many trials there actually have been. A recent estimate has been that this process leads to a 25% overestimate of the efficacy of new antipsychotics, for instance (Huston & Locher, 1996; Rennie, 1999).

Aside from the underreporting, selective reporting and overreporting, an ever-increasing proportion of the literature on treatments is ghost written. This applies particularly to material appearing in journal supplements as the proceedings of satellite symposia or consensus conferences. These papers commonly have the names of senior figures in the field on them but it is by no means clear that these experts have even seen the paper to which their names are attached. On the basis of a survey of review articles on the use of antidepressants in depression complicated by physical disorders, my estimate is that up to 50% of the review articles appearing in respectable Medline-listed journals on new drugs or aspects of their use either appear in supplement form, are ghost written or are written by company personnel.

It is common for philosophers and sociologists of science to investigate the emergence and dominance of paradigmatic views. None have hitherto considered the possibility that the convergence of views among experts constituting a paradigm might stem from the fact that a common set of articles is produced in communication agencies with the names of various experts almost randomly attached as appropriate for the occasion. This has clear implications for the sociology of science, but does any of this have any significance for clinical practice? Surely clinicians are trained to review papers and assess the literature critically. Indeed, their duty under prescription-only arrangements is to determine the true hazards of new agents and distinguish 'hype' from genuine advances.

Unfortunately, prescription-only arrangements also mean that the full weight of the pharmaceutical industry can be brought to bear on a very small number of purchasers as opposed to being spread across an entire market-place. It would be a mistake to believe that this weight will be without influence. While dependence on benzodiazepines was clearly a therapeutic problem, the wholesale switch from the use of tranquillisers in the 1980s to antidepressants in the 1990s, with the same patients

being diagnosed as having anxiety disorders in one decade and depressive disorders in another, stems to a considerable extent from the marketing power of pharmaceutical companies channelled through prescription-only arrangements. (And in all likelihood, as the SSRIs come off patent in the near future these same patients will once more be diagnosed as having anxiety disorders, to be treated with anxiolytics rather than tranquilisers.) In the case of the antipsychotics, an earlier generation of weakly neuroleptic antipsychotics was replaced by a generation of neuroleptics. The past 5 years, however, have seen a wholesale switch from neuroleptics back to a group of compounds that, in terms of receptor profile and efficacy, are indistinguishable from first-generation antipsychotics such as chlorpromazine, chlorprothixene and levomepromazine (Pedersen & Bogeso, 1998; Healy, 2001). Neither of these switches can be justified on the basis of clinical trial evidence.

Randomised controlled trials produce main effects and side-effects. By convention, the main effect of antidepressants is taken to be on mood, and effects, for example, on sexual functioning are designated side-effects. In fact, sexual functioning may be more reliably affected by an SSRI than mood. Where up to 200 patients may be needed to demonstrate a treatment effect for an SSRI in depression, as few as 12 may be needed to demonstrate efficacy for premature ejaculation (Waldinger *et al.*, 1994). Evidence of the potentially beneficial effects of SSRIs on aspects of sexual functioning such as premature ejaculation was kept almost entirely out of the public domain by companies for two decades (Healy, 1997). This should make it clear that the designation of a main effect of a compound is essentially an arbitrary decision, related to company economics and far from value-free (Healy & Nutt, 1998).

The licensing system was put in place to constrain the claims that companies can make, not to regulate clinical practice. Increasingly, however, there has been confusion on this point, and many clinicians feel that they can only prescribe compounds for their licensed indication. This confusion has come about since the 1962 amendments to the US Food and Drug Act, where the requirements for drug licensing moved from demonstrations of treatment effects to demonstrations of effects for particular disease conditions. With the restriction of drug treatments to disease states, companies have more aggressively marketed medical disease models such as panic disorder and social phobia as a means of selling compounds (Healy, 1997). This helps further the link between the claims that a company can make regarding their compound and perceptions that clinicians may have of the appropriate use of those compounds and it leads to an indiscriminate usage

of many drugs for 'depression' on the basis that they have been demonstrated to be antidepressants. In fact, a license is an acknowledgement that a treatment effect can be demonstrated, not that treatments work. It can be issued even if the majority of patients the drug is given to in clinical trials fail to show this effect – as was the case with a number of the SSRI antidepressants.

In 1860, faced with the medical arsenal, Oliver Wendell Holmes stated:

"I firmly believe that if the whole materia medica as now used were to be sunk to the bottom of the sea, it would be all the better for mankind and all the worst for the fishes." (cited in Young, 1992, p. 19)

The perception now is that new evaluative methods have pushed bad medicines out of the arsenal. In fact, there is every reason to suspect that RCTs are pushing good therapies out of health care. Psychiatric units that once had active occupational therapy sections and social programmes are now reduced to boring sterile places where only things that have been 'shown to work' happen. Patients are not exercised, nor taken out on social activities, nor involved in art, music or other therapies. If they leave hospital for psychosocial reasons, it is likely to be because of boredom. One reason for this is that RCTs, as currently interpreted and allied to the patenting system, provide evidence that can be used for lobbying purposes. In contrast, other non-specific approaches will remain, like placebo, undeniably but unprovably effective and consequently unsponsored.

Much of the above could be countenanced if RCTs had done something to restrain therapeutic zeal (the *furor therapeuticus*). There is little evidence for this. In recent years there has been a mass medicalisation of a range of nervous conditions in primary care. Only time will tell how appropriate such medicalisation is. But what is clearly inappropriate is the current lack of monitoring of the therapeutic impact of intervening in these conditions. In practice, on the basis of weak evidence of treatment effects, we have done a great deal to detect such conditions and advocate that subjects are given treatment, but little to monitor whether treatment has in fact delivered the desired result. Because these agents have been shown by RCTs to 'work', we have promoted a situation, virtually free of warnings, in which primary care prescribers and others, besieged by the mass of community nervous problems and all but impotent to do much for them, have been trapped by the weight of supposed scientific evidence into indiscriminately handing out psychotropic agents on a huge scale.

There have been moves in recent years by the Cochrane Centre and leading medical journals to encourage companies to publish all their data. The

implication appears to be that if all the data are published the field will become scientific. In fact, publication of all the data will just produce acceptable business practice in contrast to the currently unacceptable practice; the systematic concealment of data about a new car, for instance, would constitute bad business practice rather than bad science. It will take considerably more than more transparent publication practices to produce good science. Good science will result only from studies that are designed to answer scientific questions rather than from ones designed to support regulatory applications or market penetration.

---

## Coda

---

Colleagues and I recently reported the first results of a study in North Wales undertaken within a population that has been stable for 100 years in terms of population numbers, age cohorts, ethnic mix and rurality (Healy *et al.* 2001). This demonstrated that there has been a three-fold increase in the rate of detentions into psychiatric services and a 15-fold increase in the rate of admissions since the introduction of the psychotropic drugs. The inter-illness intervals for bipolar disorders appear to have got shorter rather than longer, despite the availability of supposedly prophylactic treatments. Overall, for all psychiatric conditions patients now appear to spend longer in a service bed than they would have done 50 or 100 years ago. Such findings are compatible with our treatments having effects that may be used judiciously, but in many instances are probably not being used to their best advantage; they are incompatible with our treatments being effective in practice for a majority of the patients to whom they are given.

When chlorpromazine was introduced, Evarts from the National Institute of Mental Health cautioned that the new treatment assessment and drug development methods being proposed were problematic (Evarts, 1959). Had fever therapy and, later, penicillin not been discovered as treatment for general paralysis of the insane (GPI), he noted, chlorpromazine would also have been used for dementia paralytica. The research methods we have subsequently relied on, against his advice, exclusively for dementia praecox and manic-depressive illness would have demonstrated chlorpromazine's utility for GPI. The failure of cases of GPI to clear up in response to chlorpromazine would have justified the production of an ever-increasing number of essentially similar agents. A research and therapy establishment would have arisen on the back of these

efforts, which, Evarts predicted, would have actively inhibited the discovery of an effective treatment for GPI.

The example of GPI and penicillin demonstrates that everybody knows when a treatment really works without the need for RCTs – the problem vanishes. Notwithstanding this, we work in an era that, for a range of reasons, sets great store on evidence-based medicine. RCTs and the embodiment of evidence derived from them in guidelines have become a solution for complexity and a substitute for wisdom and in some cases for common sense. This suggests a blind spot on our part when it comes to evidence about evidence.

There is, however, one advantage in the new arrangements. The first antipsychotics and antidepressants led to the emergence of antipsychiatry and a questioning of the legitimacy of psychiatry. Such a scenario is unlikely to be repeated. The market development plans of drug companies for recent and future generations of psychotropic agents include the establishment of, or penetration of, patient support groups. Psychiatrists who might once have been vilified when they advocated new physical treatments to patient groups are more likely to find themselves vilified now if they fail to endorse enthusiastically the latest treatments.

Perhaps it is now time for psychiatrists, like focus-group-oriented politicians, to follow rather than to lead. A growing string of cases, from the sacking of Nancy Olivieri from the University of Toronto for the publication of clinical trial results inconvenient to the sponsoring company to the suing of Ian Oswald in the UK for his concerns about the concealment of study data, demonstrate that fashionable treatments increasingly pose dilemmas that go beyond any problems in the evidence base or in the way that evidence is marketed.

---

## References

---

- Bisson, J. I., Jenkins, P. L., Alexander, J., *et al* (1997) Randomised controlled trial of psychological debriefing for victims of acute burn trauma. *British Journal of Psychiatry*, **171**, 78–81.
- Evarts, E. (1959) A discussion of the relevance of effects of drugs on animal behavior to the possible effects of drugs on psychopathological processes in man. In *Psychopharmacology: Problems in Evaluation* (eds J. Cole & R. Gerard), pp. 284–306. Washington, DC: National Academy of Sciences/National Research Council.
- Freemantle, N., Anderson, I. M. & Young, P. (2000) Predictive value of pharmacological activity for the relative efficacy of antidepressants drugs. Meta-regression analysis. *British Journal of Psychiatry*, **177**, 292–302.
- Gilbody, S. M. & Song, F. (2000) Publication bias and the integrity of psychiatry research. *Psychological Medicine*, **30**, 253–258.
- Healy, D. (1997) *The Antidepressant Era*. Cambridge, MA: Harvard University Press.

- (2000) The assessment of outcome in depression. Measures of social functioning. *Reviews in Contemporary Pharmacotherapy*, **11**, 295-301.
- (2001) *The Creation of Psychopharmacology*. Cambridge, MA: Harvard University Press.
- & Nutt, D. (1998) Prescriptions, licenses and evidence. *Psychiatric Bulletin*, **22**, 680-684.
- , Savage, M., Michael, P., *et al* (2001) Psychiatric service utilisation: 1896 and 1996 compared. *Psychological Medicine*, **31**, 779-790.
- Huston, D. & Locher, M. (1996) Redundancy, disaggregation and the integrity of medical research. *Lancet*, **347**, 1024-1026.
- Jick, S., Dean, A. D. & Jick, H. (1995) Antidepressants and suicide. *BMJ*, **310**, 215-218.
- Pedersen, V. & Bogeso, K. (1998) Drug hunting. In *The Psychopharmacologists* (ed. D. Healy), pp. 561-580. London: Arnold.
- Raphael, B., Meldrum, L. & McFarlane, A. C. (1995) Does debriefing after psychological trauma work? Time for randomised controlled trials. *BMJ*, **310**, 1479-1480.
- Rees, W. L. (1997) The place of controlled trials in the development of psychopharmacology. *History of Psychiatry*, **8**, 1-20.
- Rennie, D. (1999) Fair conduct and fair reporting of clinical trials. *JAMA*, **282**, 1766-1768.
- Tranter, R. & Healy, D. (1998) Neuroleptic discontinuation syndromes. *Journal of Psychopharmacology*, **12**, 306-311.
- Viguera, A. C., Baldessarini, R. J., Hegarty, J. D., *et al* (1997) Clinical risk following abrupt and gradual withdrawal of maintenance neuroleptic treatment. *Archives of General Psychiatry*, **54**, 49-55.
- Waldinger, M., Hengeveld, M. H. & Zwinderman, A. H. (1994) Paroxetine treatment of premature ejaculation: a double-blind randomized placebo-controlled study. *American Journal of Psychiatry*, **151**, 1377-1379.
- Weissman, M. M., Klerman, G. L., Paykel, E. S., *et al* (1974) Treatment effects on the social adjustment of depressed patients. *Archives of General Psychiatry*, **30**, 771-778.
- Young, J. H. (1992) *Pure Food*, p. 19. Princeton, NJ: Princeton University Press.

A conference by FOCUS, at The Royal College of Psychiatrists Research Unit

## ***Bridging the gap between policy and practice in CAMHS***

Westminster, London

4-5 December 2001

A major 2 day multi-disciplinary event for child and adolescent mental health professionals, organised jointly by FOCUS and the Faculty of Child and Adolescent Psychiatry at The Royal College of Psychiatrists.

The conference will focus on policy initiatives over the last 3 years that have affected CAMHS, debate and discuss key new initiatives and priority areas and explore innovative ways to present the current evidence base.

The programme will include national strategies for CAMHS, the National Service Framework for Children, the role of the Children's Taskforce, the fit between evidence and policy, culturally appropriate services and the impact of the NHS Plan among other key issues for CAMHS development.

For booking forms and further information, please contact  
Catherine Ayres on 020 7227 0822  
or [catherine.ayres@virgin.net](mailto:catherine.ayres@virgin.net)

**FOCUS**



Visit our website: [www.focusproject.org.uk](http://www.focusproject.org.uk)