International Journal of Risk & Safety in Medicine 25 (2013) 111–121
DOI 10.3233/JRS-130588
IOS Press

Opinion

Data based medicine and clinical judgement

David Healy[a,*], Derelie Mangin[b] and David Antonuccio[c]
[a]*Department of Psychiatry, Cardiff University, Cardiff, UK*
[b]*Department of General Practice, University of Otago, Christchurch, New Zealand*
[c]*Department of Psychiatry, University of Nevada, Reno, NV, USA*

Received 17 December 2012
Accepted 7 April 2013

Abstract. Randomized controlled trials (RCTs) are a useful tool to check the effectiveness of drugs but have come to shape the culture of medicine in a manner that increasingly compromises medical care. Dependence on RCT evidence is compromised by the well-known problems stemming non-publication of trials, lack of access to trial data, ghostwriting of those trials that are published and a variety of coding and other strategies to hide harms. But what is less appreciated is that whenever a drug and an illness can produce the same benefit or harm that the outcomes of RCTs can be profoundly misleading. This article gives examples of how RCTs can produce the wrong answer.

Keywords: Randomized controlled trials, evidence based medicine, Simpson's paradox

1. Introduction

Randomized controlled trials (RCT) have made critical contributions to healthcare. They have given rise to an approach to medicine now called Evidence Based Medicine (EBM) [1]. But it is becoming increasingly clear that there are a number of ambiguities in the term Evidence and that for instance practising Evidence Based Medicine may be quite different to practising Data Based Medicine.

Under the influence of EBM, RCTs are widely cited as offering gold-standard evidence in a manner that implies they will semi-automatically provide the best possible answers for almost any problem, in particular any issue to do with drug treatment. In contrast, while good clinical judgement once played a key role especially in delineating treatment related adverse events, when compared with RCT data clinical judgements are now likely to be dismissed as anecdotal. On the basis of an apparent absence of evidence from trials about specific adverse effects, doctors, patients and anyone involved in healthcare faced with these side effects are increasingly called on to doubt the evidence of their own eyes.

There are serious and unrecognized problems with taking an uncritical approach to the data about treatment induced adverse events that stem from RCTs. Some of these problems stem from a series of inappropriate data-management strategies that are relatively widely known, but there are more significant

*Address for correspondence: David Healy, Department of Psychiatry, Cardiff University, Cardiff, UK. E-mail: David.Healy54@googlemail.com.

problems. By their very nature RCTs conducted in illnesses offer so many ways to hide treatment related adverse events that claims about adverse events that arise from RCT data must be assumed to be worthless until proven otherwise. The problems are not confined to the adverse effects of treatment.

2. Miscoding and mislocation

Some of the problems surrounding RCT data and adverse events are relatively widely known. These problems hinge on a series of inappropriate data-management strategies.

For instance, it is known that in pharmacotherapy trials the side effects of a drug may be coded under disparate headings. Suicidal acts have been coded under headings ranging from anxiety, agitation, akathisia, emotional lability, thinking abnormally, abnormal dreams, psychosis and others through to aggravated depression or treatment ineffectiveness and sometime less obvious terms such as nausea. Allocation to various different codes may happen unintentionally, and can to some extent be overcome by standardised queries of coding systems like MedDRA. Intentional or unintentional, such miscoding divides and conquers what may be a problem for a pharmaceutical company, by artificially lowering the apparent rate of a particular problem.

Instances of problems being hidden in this way have been described for many adverse events. In general all such miscodings devalue clinical judgement; when a doctor faced with a problem she terms suicidal ideation queries a company to find if they have any other reports of such effects, she is likely to be told "no, we haven't". The message is that experts looking at these things have not seen it – you the coalface doctor are not going to pit your judgement against the experts are you?

The coding of data by pharmaceutical companies under diverse headings means that any tabulation of treatment related events that might appear on sites like clinicaltrials.gov must necessarily be suspect unless there is access to the raw data. But data in this case means information at the level of the individual patient rather than a listing of coded events. It may mean interrogating the patient.

In addition to miscoding, treatment related adverse events may be mislocated. Many trials have a washout period (sometimes called a placebo run-in phase) lasting a week or two before randomization proper. This is a period where patients may be asked to stop their prior antidepressants or other treatments. It is now clear that this is an extremely hazardous period, owing possibly to the withdrawal effects of prior treatments.

In the initial trials of selective serotonin reuptake inhibiting (SSRI) antidepressants Fluoxetine, Paroxetine and Sertraline, suicidal events that occurred during this washout period were later filed inappropriately as though such cases had been randomized to placebo [2]. This was spotted but ignored by all regulators and explained by companies in terms of placebo being equivalent to being on nothing.

Events from the post-trial follow-up period that have occurred in patients previously on placebo but after randomization put on active agents have also been filed under the heading of placebo [3]. Mislocations of this sort have been detected not just for suicide related adverse events on antidepressants but also for heart attacks on treatments like rofecoxib (Vioxx) and rosiglitazone (Avandia) [4].

Beyond miscoding and mislocation there is of course non-publication or ghost writing of essentially negative trials to the point where they are portrayed as positive [5]. These are all serious problems. They repeatedly devalue clinical judgements in that doctors making judgement calls are told there is nothing in the scientific literature to support their perceptions, whereas in fact there is.

Nevertheless many people would regard these as relatively superficial problems in that much of the problem it seems could be put right if there was the access to the data from clinical trials. There is a

growing consensus that there should be access to the data if the field of therapeutics is to be properly scientific [6]. In response pharmaceutical companies are offering "managed access" in lieu of open access.

3. Eclipsed by statistics

Statistical approaches pose a much greater threat to clinical judgement than mislocation, miscoding and non-publication. An inappropriate reading of statistical significance tests also hides a much greater number of adverse outcomes than happens by mislocation.

There is an impression that we undertake significance testing and controlled trials almost explicitly because the judgement of doctors is not reliable. For centuries doctors saw and still see possible beneficial effects of a treatment and attribute these benefits to a medicine when in fact placebo would perform as well.

While it is entirely appropriate that we hold judgements about effects that may be claimed as benefits to a particularly high standard, and exclude the possibility that these effects could have arisen in any way other than by virtue of the action of the drug, this is not the case for the harms that stem from treatments. In the case of a benefit, the evidence of a benefit can be believed; the point at issue is the origin of this benefit. If it is very clear the drug is producing the benefit no trial is needed. It is only where it is less certain the drug is producing a benefit there is a need for deliberation and scrutiny.

In the case of a harm, immediate action on the basis of a clinical observation is mandated, regardless of whether the harm arises from the drug. It is a serious matter to do anything to interfere with the process of risk minimization and especially to do so by persuading doctors that because a problem might have arisen by chance it can be ignored.

Where a substantially increased rate of adverse events on drugs does not reach statistical significance for one reason or another, so that a company can say there is still a greater than 5% probability that these observations could have arisen by chance, company rhetoric now routinely dissuades doctors or patients from taking action claiming that as these events might have arisen by chance they in fact did not happen - no increase in risk has been demonstrated. The fact that we cannot rule out a possible role of chance in generating the findings – although the likelihood of chance being involved is vanishingly small – should not lead anyone to think there is in fact no increase in risk [2, 3].

This use of statistics by companies is the most sophisticated version of the "doubt is our product" dynamic first pioneered by tobacco companies. In recent years it has been used by pharmaceutical companies to hide a much greater number of suicidal acts in the case of the antidepressants, and of heart attacks or other causes of mortality in the case of Vioxx [7] or Avandia [4], than was ever hidden by mislocating suicidal acts from washout periods to placebo, or otherwise allocating heart attacks to analytic subgroups as was done for Vioxx.

To claim no increase in risk exists when in fact there is an increase in risk but the role of chance in producing this clear increase has not been eliminated is to call on people to doubt the evidence of their own eyes. It is also to place the burden of proof on those adversely affected by a treatment rather than have that burden placed where it should be – which is on the sponsor of a medication that has produced a clear increase in risk.

These statistical sleights of hand *will not* be put right by access to data. Using these approaches pharmaceutical companies have been able to hide dead bodies in the broad light of day. Medicine is unique in this regard; in any other area of life from physics to sport 10 is greater than 3 but in medicine 3 may be greater than 10 if the 3 is statistically significant but the 10 is not [2].

4. Language games

Controlled trials are commonly portrayed as value neutral. Trials are designed in almost all cases to test the effects of treatment. But in fact our language slips and the effects that might lead to a market authorization are called benefits.

In the case of the antidepressants for instance, the trials show effects on rating scales, while trials for statins, bisphosphonates or respiratory drugs show effects on lipid levels, bone densities and peak flow rates. In the case of all these drugs there may be more lives lost on the active treatment than placebo, or fewer people returned to work on active treatment, so to portray the demonstrated effect as a benefit is to make an unwarranted and unsupported claim.

There is a benefit to a pharmaceutical company to being able to market an effect as a benefit, but even in the case of sleep induction by an hypnotic, many potential takers may not regard this type of sleep inducing effect as beneficial. Except in the case of lives saved, it typically remains to be demonstrated that an effect is in fact beneficial and it is ordinarily only an individual doctor and patient who can make that determination.

Trials could be set up to demonstrate a benefit, but few are. The proponents of EBM meanwhile typically play into pharmaceutical company hands by routinely referring to the role of trials in establishing benefits.

5. RCTs in clinical settings

Despite the non-publication of trials, ghost writing of publications, miscoding of data, mislocating of events, misuse of statistical significance testing, and co-opting of language, the faith of most of those involved in healthcare in controlled trials remains unshaken. Most scientists believe that if only the data were made publically available, that many medical problems could be solved and some believe that controlled trials demonstrate cause and effect.

RCTs were introduced to healthcare to contribute to safety by eliminating ineffective treatments. They are most useful and contribute to safety when they show that certain claims are unwarranted.

Trials arose in agricultural settings and to this day the additional confounders that can arise in clinical settings from the interplay between a disease and its treatment remain poorly understood.

There are two ways in which a disease being treated can come into play to conceal treatment induced effects, adverse or other. One lies in disease heterogeneity and the other arises when there is a variable effectiveness of treatment and both the disorder and its treatments can give rise to a superficially similar problem as when both depression and antidepressants give rise to suicidal acts.

5.1. Disease heterogeneity

In the late 1980 s, Lilly undertook a trial of Fluoxetine (Prozac) in a group of patients with what was first termed recurrent brief depressive disorder (RBDD). In this trial placebo was sweepingly statistically superior to Prozac. The study was published four years later shorn of its data except for the broad claim that the numbers of suicide attempts in the Fluoxetine and placebo groups were the same [8].

In the early 1990 s, SmithKline Beecham undertook a study (protocol 106) of a closely related SSRI antidepressant Paroxetine (Seroxat, Paxil) in the same hospital centre, in the same diagnostic group of patients, possibly with some of the same patients who had been in the Prozac trial. This second study was terminated early, and the results were never published. The rate of suicidal acts on Paroxetine was three-fold higher than on placebo (data available from the authors).

Table 1
Suicidal acts in major depressive disorder trials

Major Depressive Disorder Trials (MDD)	Paroxetine	Placebo	Relative risk
Number of suicidal acts/Number of patients	11/2943	0/1671	Inf (1.3, inf)

Table 2
Suicidal acts in major depressive and recurrent brief depressive disorder trials

	Paxil	Placebo	Relative risk
MDD trials acts/Patients	11/2943	0/1671	Inf (1.3, inf)
RBDD trials acts/Patients	32/147	35/151	0.9
Combined acts/Patients	43/3090	35/1822	0.7

Several years later SmithKline Beecham undertook yet another trial (protocol 057) in a similar group of patients [9]. There are several quite different extant sets of figures from this study; the figures used in this article are from a SmithKline Beecham publication.

In April 2006, GlaxoSmithKline issued a press release with the following figures for suicidal acts in the Paroxetine trials in their most important therapeutic area, Major Depressive Disorder [10]. Conceding that there was a risk of suicide for this patient group was extremely problematic for the company.

This study of patients with Major Depressive Disorder shows a statistically significant increase in the suicidal act risk on Paroxetine. The full press release, however, combined the "depressed" patients from protocols 106 and 057 with patients from the MDD trials (See Table 2).

When the two datasets, from Major Depressive and Recurrent Brief Depressive Disorder trials, are added together the increased risk of suicidality not only vanishes but Paroxetine becomes apparently protective against suicide risk. In fact it is possible to increase the number of suicidal acts on Paroxetine in the RBDD group by 20 and achieve the same outcome.

In these two protocols (106 and 057), 298 patients had 67 suicidal acts between them – this is 100 times more suicidal acts per patient than in the set of major depressive disorder trials (In fact a handful of patients in 106 and 057 had close to half the suicidal acts between them).

This paradoxical outcome is a variation on what has been termed Simpson's paradox [11]. Simpson's paradox arises when collapsing trials together based on the simple addition of all events leads to a reversal of the direction of effects seen in a majority of studies [11]. This paradoxical effect is most likely to happen if the event rates in studies differ markedly – as they do in the MDD and RBDD studies outlined here.

In the case of protocols 106 and 057, the timing of these studies makes it conceivable that some of those involved in the design of these studies had accepted that SSRIs like Paroxetine cause suicide and embarked on a series of studies that used a problem they accepted the drug causes to hide a problem that the drug causes.

But whether there was intent or not in this set of studies, something similar is possible in principle in single studies or combinations of studies if the clinical population recruited is heterogeneous in respect of a particular treatment effect. It would for example be quite possible to diagnose these RBDD patients has having MDD and recruit them to an MDD trial and for the contribution from these patients to raise the background placebo rate, thereby concealing the adverse event.

Unless the treatment effect, be it a respiratory, gastro-intestinal, rheumatological or other system event is fully understood and in particular its link to treatment, the heterogeneity of clinical populations makes

it possible that effects linked to treatment will not emerge as clearly linked to treatment and indeed it can even appear that treatment is having the opposite effect in general to an effect it is clearly having in subtypes of the disorder.

5.2. It's the disease not the drug

In 1990 concerns about a suicide risk of antidepressants arose with the publication of case studies in which suicidality emerged on Fluoxetine, cleared when treatment stopped and re-emerged on the reinstitution of treatment. These reports fulfilled all the standard canons for determining cause and effect [12]. In the clinical judgement of the senior investigators involved, the drug was causing some people to become suicidal.

However, the field was swayed instead by rhetorical arguments that the clinical trial data showed no risk, even though the data showed a clear but not statistically significant increase in risk on active treatment as noted above. But even a statistically significant increased risk would not have been the appropriate way to demonstrate cause and effect.

Over 15 years later when a sufficiently large number of trials were assembled, and rates of suicidal acts on selective serotonin reuptake inhibitors (SSRIs) did show a statistically significant increase in the relative risk of a suicidal act compared to placebo, FDA officials stated that this statistically significant doubling of risk demonstrated a causal effect. It was a mistake to link this doubling of risk to causality.

FDA conceded causality in this case because of a precedent set in the 1980 s in a series of tort cases involving claimed injuries following breast implants, and Bendectin. There has since been a de facto medico-legal convention that a demonstration of cause and effect requires a statistically significant doubling of the risk of the adverse event in question [13, 14].

This understanding has arisen by default, most likely because it suits the interests of the corporations produced drugs or implants or other goods that might cause injuries, and the convenience of bureaucrats who want a simple rule to which they can appeal.

The case of suicide on antidepressants shows why and how this understanding is wrong. There are 3 steps to the argument. First, at a meeting a year after the original tricyclic antidepressant, Imipramine, was introduced in 1958 several clinicians stated on the basis of clinical observations involving the emergence of the problem on exposure to the drug (challenge) and clearing up of the problem on discontinuation of treatment (dechallenge) that Imipramine could directly cause suicide by increasing agitation. There was no dissent [15].

Second, there is considerable clinical trial evidence that Imipramine, Clomipramine and other tricyclic antidepressants are more effective than SSRIs. Specifically, they are effective in melancholic depressions where SSRIs are not [12]. This means that they are therefore effective in a patient group at a substantially higher risk of suicide than those outpatient or primary care depressed patients that were entered into SSRI trials [12].

Third, it follows from this that in a placebo controlled trial of Imipramine in melancholia the rate of suicidal acts in the placebo arm would likely be higher than in the placebo arm of SSRI trials while Imipramine would likely lower the rate of suicidal acts by successful treatment of melancholia in the treatment arm in a way that SSRIs could not have done in mildly depressed patients at little risk of suicide.

From this it follows that while the relative risk of a suicidal act on SSRIs compared to placebo was roughly 2.0, the relative risk of a suicidal act for Imipramine or other tricyclic agents in melancholia trials might well be less than 1.0, perhaps as low as 0.5.

On the basis of a relative risk less than 1.0, many academics and regulators like FDA would not be prepared to concede that the drug being tested could cause suicide, although using the criteria still embodied in standard textbooks of adverse event causality – namely challenge-de-challenge and re-challenge along with dose-responsiveness - these older agents that also inhibit serotonin reuptake unquestionably do cause suicide.

The notion that a treatment that on the one hand causes a problem might in good faith trials, without any manipulation of the data or statistical artefact, give rise to a relative risk <1.0 for that problem should pose difficulties for any simple explanations of what RCTs do. It makes it clear that whatever else RCTs do, they do not simply demonstrate cause and effect. This possibility has been noted in abstract terms by others [16–18].

6. Recovering uncertainty

This thought experiment raises several epistemological issues.

First, giving a primacy to RCT data over other data throws us into the unusual position of having to concede that the suicidality SSRIs induced only came to light as a result of the artefact of their being tested in populations at minimal risk of suicide.

Second, given what we now know about the behaviour of an effect like suicide, we can construct studies to make it appear or disappear. But knowing about an effect to this extent raises the question as to what exactly RCTs establish. It also points to an adverse effect of RCTs – they can be deliberately manipulated to conceal problems. Where an adverse event is as well understood as suicide on antidepressants it is quite possible to design trials to produce any predetermined relative risk between 0.1 and 10.0.

Third, in the case of a relative risk of a suicidal act in depression trials, if the drug has also been effective and helped some of those with illness linked suicidality, then the resulting overall relative risk is in fact a compound of risks. The treatment induced component of that compound is therefore almost certainly greater than the overall relative risk suggests, but we have no way of knowing by how much greater it is. It is unknowable.

RCTs are often held to establish the frequency with which certain effects happen. When both the illness and its treatments give rise to similar problems RCTs cannot in principle offer good data on the frequency of a problem. Claims that RCTs offer reliable estimates of the frequency or reproducibility of certain observations stemming from RCTs are however widely used to exclude clinical judgements in legal cases on the basis that clinical judgement is not reliable.

We can map out the dilemmas that arise in the case of the antidepressants and suicide because these drugs and this problem are relatively well characterized. Comparable scenarios can be mapped out for some arrhythmias on anti-arrhythmics, for beta agonists given for asthma, as well as for certain vaccines. In principle such difficulties will potentially arise in every case in which both an illness and its treatment give rise to at least superficially similar problems. If the adverse effect and the treatments are not well understood, the results that emerge from these trials become impossible to interpret other than to say these are the data that emerged from this particular assay.

A "blame the illness" dynamic feeds into a clinical bias to see problems as stemming from diseases rather than their treatments. Perhaps recognizing the merits of this defence, companies marketing antipsychotics in recent years faced with elevated rates of diabetes argued that schizophrenia gives rise to diabetes without any evidence to support their position, and found this defence worked [19]. Almost anything it seems can be portrayed as a risk of the illness.

7. Balancing RCTs and clinical judgement

RCTs are immensely useful clinical tools. But for the effect RCTs produce to be of benefit, and to outweigh the adverse effects they can produce (of which more below), there are a number steps that need to be taken, some of which involve RCTs and others that do not.

7.1. RCT solutions

First it must be noted that a number of these problems arise in RCTs where the effect in question is not the primary outcome. They arise in trials therefore not designed to look specifically at an issue. One option is to simply say that if a study has not been specifically designed to look at the issue in question, we have in fact no good clinical trial evidence on the issue. A failure to build a statement like this into all statements about RCTs risks compromising the credibility of RCTs.

Two further sets of RCTs have the capacity to reveal the effects of treatment without the confounding or effect modification that may stem from an associated illness. These are the phase one or healthy volunteer trials that companies and universities conduct. At present clinical trials in patient groups (phase 2 and 3 trials) are registered so we know what trials are happening even though we cannot get the data from these trials. But there is no register for phase 1 studies. And where access to the data from clinical trials remains problematic because of patient confidentiality, there should in principle be no problem with access to the data from healthy volunteer studies.

The second set are dose-response studies. These are rarely done at present.

Phase 1 studies can be surprisingly powerful. Consider a never-published study of 12 volunteers conducted in Leeds in 1983, 9 years before the antidepressant Sertraline was marketed, and 21 years before FDA required it to carry suicide warnings. This study of Sertraline was terminated early because all of the women on Sertraline had become anxious or apprehensive, noting thoughts of aggression and related difficulties. Pfizer concluded that Sertraline had caused these changes. One of the senior investigators further noted that comparable results had been seen in healthy volunteers for other SSRIs then in development [12].

There are in fact several known healthy volunteer suicides and episodes of violence in the phase one studies of SSRIs. In order to detect an adverse event like suicide, the event must exceed the base rate in the untreated clinical population. The salience of the event is therefore much more marked in a healthy population. This is the reverse of the situation with protocol 057 and 106.

Dose response studies were once common in therapeutics and generations of doctors learned to start low and go slow. The drug most associated with this was digitalis but it applied in principle to all drugs. Dose responsiveness of this type will be accepted by regulators including FDA as evidence of efficacy to this day.

In the case of the SSRIs dose response studies would have forced companies to specify the intermediate variable on which their drug has effects and where they could demonstrate dose-response effects that cannot be demonstrated in clinical syndromes.

Essentially SSRIs produce a form of emotional numbing that may be of therapeutic benefit. Increasing the dose of treatment increases this effect – except in those who are agitated instead, who also show a dose response effect. Demonstrating a dose-responsive emotional numbing would have made clear what the drug does that gets patients better and would have created a space for clinical judgement. Rather than giving these drugs because they have been supposedly shown to work in depression, doctors and patients

would have been in the business of attempting to establish whether a desired effect was in fact happening or not. They would have been called on to make case specific judgements.

7.2. Non-RCT solutions

With adverse events that stem from both an illness and its treatment, the question is what weight to put on observations from controlled trials that have not been designed to investigate the issue but give rise to a non-significant increase or decrease in the relative risk for the event versus observations that incorporate challenge, de-challenge and re-challenge (CDR) relationships, along with evidence of dose-responsiveness, and reversal by antidote.

One doctor who reports that a patient develops an adverse event on treatment and who, because of CDR, dose response and other relationships, links this to treatment, might not be believed. If, a thousand doctors outline similarly good quality reports (and even more so, if each knows there are 999 other reports) the field is likely to believe the outcome. Somewhere in between we cross a credibility threshold for believing reports like this – is it 5 good reports, 10 or 100 reports? What weight should be put on such reports as compared with data from RCTs where the event in question has not been the primary outcome measure?

In the case of antidepressants and suicide, 6 good clinical observations from one clinical centre turned out to be correct, but made no difference to the general perception of the issues. There is likely a cognitive bias in that events like suicidal acts and violence seem much less likely to be attributed to prescription drugs than dependence and withdrawal for instance. In the case of suicide on antidepressants this bias has not been overcome even though regulators have placed Black Box warnings on the drugs.

Current antidepressant trials run for approximately six weeks and involve changes on rating scales such that a collection of "adverse" effects – sedation, emotional numbing, increased appetite – can lower rating scale scores, producing an apparent benefit. In addition, the FDA currently licenses drugs on the basis of two positive trials, even if such trials are nested among a larger number of negative trials. The FDA concedes that this process means that, in the case of the antidepressants, all we have is the signal of a treatment effect rather than a demonstration of effectiveness.

With the current system, it would seem entirely possible to put alcohol, nicotine, diazepam or dexamphetamine through the system and get approval as an antidepressant. Giving any of these for 6–8 weeks would probably not cause significant clinical problems and indeed might cause fewer problems than the SSRIs have caused in these trials.

Because of their familiarity with alcohol, and with the reputation of these other drugs, most people would know that taking these "antidepressants" beyond 6 weeks might not be a great idea. But most patients and many doctors are disarmed by testing processes, which transform prescription-only drugs (on prescription-precisely because they may turn out to be risky) into risk-free drugs. What is needed is a metric that assumes ab initio that novel agents will come with problems such as dependence and other consequences if used in the longer term.

We have stepped back from viewing new drugs as poisons to be treated warily on the basis that their risks have not yet been demonstrated and in practice we increasingly assume that the lack of evidence means these new drugs pose no risks. It would be more in keeping with traditional clinical practice during the early life of a drug to assume a range of hazards are likely to happen, with an appropriate adjustment made later, only if it transpires a drug is safer than average.

8. Adverse consequences of an exclusive reliance on RCTs

In order to run a trial of a drug in many clinical conditions, where mortality or return to work are not available as outcomes, it is necessary to construct rating scales or other surrogate outcomes. In the case of female sexual dysfunction, the rating scales used include items like clitoral numbing. In this case, the clinical trial process forces women to attend to aspects of functioning that they would not ordinarily attend to. Were any of these treatments ever to come on the market, the marketing of the effects of drugs on clitoral sensitivity risk seriously affecting the experience of love-making.

The marketing of these treatments would also come wrapped in packages that encourage women with hypoactive sexual desire disorder to have their testosterone levels checked rather than have them review the state of their relationships.

The clinical encounter is also a relationship, and good care involves a great deal of sensitivity to the dynamics of the relationship. In a rather similar way to the way RCTs risk changing the experience of love-making, clinical trials have affected the nature of clinical experience. The doctor has been numbed to the reality of the patient who has become increasingly invisible. Clinical encounters have become a factory exercise that aims at implementing impersonal protocols, algorithms and guidelines based on miscoded data in ghost written publications from trials not designed to or incapable of detecting many of the significant effects of treatment.

Freud introduced us to the notion that we cannot believe everything a patient or doctor says; that there can be layers of meaning behind a statement. In this sense he was unquestionably correct even if this understanding does not offer specific cures for particular conditions. The problem with the Freudian approach is when enthusiasts discount all claims of abuse on the basis that some such claims may well be wrong.

In a similar fashion, RCTs have made us aware of the biases that doctors and patients bring to treatment. But just as it was a mistake to follow Freud blindly, and we have had to learn the hard way that many statements of abuse should be believed, so also it is a mistake to destroy the ability of doctors and patients to draw lessons from their own experience.

When they hear the term Evidence Based Medicine, most people likely hear Data Based Medicine. In fact, because of ghost writing, non-publication of studies, and statistical modelling of data rather presentations of real data, EBM delivers “evidence” rather than data. Data Based Medicine exists at the level of the patient or the doctor-patient relationship. The patient is ultimately the source and guarantor of the data and any selections of the data that might be presented as evidence [20].

At present, the studies that come closest to being Data Based Medicine are the case reports typically dismissed by companies and regulators as anecdotal. In fact if they incorporate appropriate causality algorithms these studies are likely to be more scientific than much of what passes as the scientific literature in medicine these days. The objectivity in these cases ultimately comes from teamwork. The more that individual cases or groups of cases are exposed to the critical gaze of a number of individuals with diverse backgrounds, the more likely is it that any links between a treatment and its effects will be established.

Conflicts of interest

D.H. has had consultancies with, been a principal investigator or clinical trialist for, been a chairman or speaker at international symposia for, been in receipt of support to attend meetings or lecture fees

from Astra-Zeneca, Boots/Knoll Pharmaceuticals, Eli Lilly, Janssen-Cilag, Lorex-Synthelabo, Lundbeck, Organon, Pharmacia & Upjohn, Pierre-Fabre, Pfizer, Rhone-Poulenc, Roche, Sanofi, GlaxoSmithKline, Solvay; D.H. has been an expert witness for the plaintiff in 20 legal actions involving SSRIs and suicide, homicide, dependence and birth defects and has been consulted on several attempted suicide, suicide, and suicide-homicide cases concerning people taking antidepressants, in most of which he has offered the view that the treatment was not implicated. He has also been an expert witness in one patent case, and one securities case.

D.H. is involved in a patient adverse event reporting website – Rxisk.org

D.M. declares no support from any organisation for the submitted work; DM is involved in a patient adverse event reporting website – Rxisk.org

D.A. declares no support from any organisation for the submitted work.

References

[1] Sackett DL, Rosenberg WM. The need for evidence-based medicine. J Roy Soc Med 1995;88(11):620-4.
[2] Healy D. The Antidepressant Tale: Figures Signifying Nothing? Advances in Psychiatric Treatment 2006;12:320-8.
[3] Healy D. Pharmageddon. University of California Press, Berkeley, 2012.
[4] Cohen D. Rosiglitazone: What went wrong? BMJ 2010;341(7772):530-5.
[5] Turner EH, Mathews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358(3):252-60.
[6] Godlee F. Clinical trial data for all drugs in current use. BMJ 2012;345:e7304
[7] Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ. VIGOR Study Group. Comparison of upper gastrointestinal toxicity of Rofecoxib and Naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med 2000;343(21):1520-8.
[8] Montgomery DB, Roberts A, Green M, Bullock T, Baldwin D, Montgomery SA, Lack of efficacy of Fluoxetine in recurrent brief depression and suicidal attempts. Eur Arch of Psychiatry Clin Neurosci 1994;244:211-5.
[9] Verkes RJ, Van der Mast RC, Hengeveld MW, Tuyl JP, Zwindermann AH, Van Kempen GM. Reduction by Paroxetine of suicidal behavior in patients with repeated suicide attempts but not major depression. American J Psychiatry 1998;155(4):543-7.
[10] GlaxoSmithKline (GSK). Paroxetine Adult Suicidality Analysis, 2006. gsk.com/media/paroxetine/briefing_doc.pdf. [Accessed Aug 8th 2007].
[11] Cates CJ. Simpson's paradox and calculation of number needed to treat from meta-analysis. BMC Medical Research Methodology 2002;(2):1. www.biomedcentral.com. doi:10.1186/1471-2288/2-1
[12] Angell M. Science on Trial. Norton WW, New York; 1997.
[13] Healy D. Let them eat Prozac. New York: New York University Press; 2004.
[14] Persuad N, Healy D. Epidemiological evidence in forensic pharmacovigilance. Int J Risk Saf Med 2012;24(1):31-5. doi:10.3233/jrs-2012-0550
[15] Davies EB. Proceedings of the symposium held at Cambridge, 1959. Cambridge University Press, 1962.
[16] Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. International J Epidemiology 1988;15(3):412-8.
[17] Lanes SF, Biological interpretation of relative risk. Drug Safety 1999;21:75-9.
[18] Rothman KJ, Greenland S, Lash T. Modern epidemiology. 3rd ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2008.
[19] Le Noury J, Khan A, Harris M, Wong W, Williams D, Tranter R, Healy D. The incidence and prevalence of diabetes in patients with serious mental illness in North West Wales: Two cohorts 1875-1924 and 1994-2006 compared. BMC Psychiatry 2008;8:67. doi:10.1186/1471-244X-8-67
[20] Dyer C, Blumsohn A. Academic who took on industry. BMJ 2010;340:22-23. See General Medical Council. Fitness to Practise Panel 2009; available from www.thejabberwock.org/doc/doc00001.pdf: page148 (D3/1).