



FEATURE

ESSAY

Clinical judgments, not algorithms, are key to patient safety—an essay by David Healy and Dee Mangin

When it comes to detecting harms related to drugs, clinicians' and patients' judgment trumps trials, say **David Healy** and **Dee Mangin**. Failure to realise this is the greatest threat to the safety of medicines

David Healy *professor*¹, Dee Mangin *professor*²

¹Department of Psychiatry, Bangor University, Wales; ²Department of Family Medicine, McMaster University, Ontario

Immediately on taking a selective serotonin reuptake inhibitor (SSRI), most people have some genital anaesthesia.¹ This may be aggravated on withdrawal of the drug and can remain for years after treatment has stopped, constituting post-SSRI sexual dysfunction (PSSD).² The first case of PSSD was reported to regulators in 1987, even before fluoxetine was approved. While sexual dysfunction features in the labels of SSRIs, neither genital anaesthesia nor PSSD does. The fluoxetine label states that “there are no adequate and well-controlled studies examining sexual dysfunction with fluoxetine treatment.” The citalopram label acknowledges “some evidence suggests that SSRIs can cause such untoward sexual experiences.”

A standard refrain is that randomised clinical trials of short duration and small size have limited ability to detect rare effects of drugs, implying that longer trials are all that's needed. But as indicated by the sexual effects of SSRIs, which are more common than their mood effects,¹ a possibly greater problem lies not in whether we can detect rare adverse events but in our limited ability to detect common ones. Fetishising RCTs as medicine's only true tool for establishing drug-effect relations may be one reason for this problem.

The gold standard way to miss adverse events

In 1962, in the wake of the thalidomide disaster, RCTs—a then poorly understood technique brought into the mainstream by the English epidemiologist Austin Bradford Hill—were adopted in amendments to the US Federal Food, Drug, and Cosmetic Act to buttress the safety of medicines by keeping ineffective drugs off the market, even though Bradford Hill's landmark RCT of streptomycin offered less information on the drug's benefits and side effects than prior clinical evaluation.

But Bradford Hill was not an uncritical advocate for RCTs. In 1965, while elaborating on Koch's postulates, he emphasised

the role of dose responsiveness and of challenge-dechallenge-rechallenge (CDR) in determining clinical causality, and he made clear his view that if RCTs were ever seen as the only way to evaluate a drug, the pendulum swinging from standard to controlled clinical observations “would not only have swung too far it would have come off its hook.”³

Louis Lasagna, another early advocate of RCTs, was responsible for their inclusion in the 1962 act as a gateway to the market,⁴ even though he had run a prior placebo controlled trial of thalidomide that did not raise any red flags regarding safety.⁵

But in 1983, Lasagna, like Hill, increasingly aware of the drawbacks of RCTs and faced with claims that spontaneous reporting was “the least sophisticated and scientifically rigorous . . . method of detecting new adverse drug reactions,” replied, “This may be true in the Webster's dictionary sense of sophisticated meaning ‘adulterated’ . . . but I submit spontaneous reporting is more ‘worldly-wise, knowing, subtle and intellectually appealing’ than grandiose, expensive Phase IV schemes [RCTs].”⁶

Nevertheless, the rhetoric of RCTs portrays them as a technique that offers the best control of bias and confounders. As the example of the sexual side effects of SSRIs shows, however, the necessary focus on a primary outcome opens RCTs to a most profound bias: assurance that the trial will deliver only information on what trialists wish to learn about and little to no information on those things that are not proactively assessed, such as most adverse events, irrespective of how common they may be.

Effects can be missed in RCTs by design, such as when a vaccine trial omits adequate methods to detect possible autoimmune effects.⁷ Effects can also be missed because of the heterogeneity of the condition being investigated or where both condition and treatment give rise to similar states, as when

antidepressants trigger suicidality or rosiglitazone triggers heart failure in patients with diabetes.^{8,9}

Nor do RCTs automatically eliminate confounders. In trials on healthy volunteers, for instance, the sexual, suicidal, and dependence producing effects of SSRIs stood out, whereas in clinical RCTs these effects merged with superficially similar features of the participants' condition.¹⁰

Claims that randomisation, blinding, and pre-specified protocols protect against bias have aggravated these problems by allowing trial sponsors to turn to inexperienced investigators and to contract research organisations that conduct trials in settings of socioeconomic deprivation, instead of using senior clinicians who know the patients and their conditions.¹¹

As a result, whereas in 1962 it took two to three years for unanticipated effects of a drug—such as changes in hair texture and deep vein thrombosis on contraceptives,¹² or tardive dyskinesia on antipsychotics¹³—to be accepted in clinical practice, it now takes two to three decades. Suicidality and withdrawal syndromes resulting from leucotriene antagonists are among the commonest adverse effects among children reported to regulators,^{14,15} but these are still denied by many in clinical practice.

Linked to this neglect of harms, and an emphasis on efficacy, the numbers of patients taking drugs began rising, with 40% of the US population aged between 45 and 64 now on three or more medicines, and 40% of over 65s on five or more.¹⁶ Recent reports of stalled improvement in life expectancy may indicate that we need to give greater consideration to safety to achieve optimal outcomes.¹⁷

Pharmacovigilance

Even before 1962 the US Food and Drug Administration made efforts to set up adverse event reporting systems. This seemed obviously important, but early systems had a poor pick-up. Physicians' knowledge of drug effects continued to derive from clinical experience and from case reports in journals.

However, an increasing industrial use of RCTs, allied to a sequestration of clinical data from company trials, slowly relegated clinical evaluations that drug X caused effect Y, even when buttressed by CDR evidence, to the status of "anecdotes." This came to a head in 1991, when Eli Lilly appeared at a regulatory hearing over fluoxetine. The company successfully pitted a meta-analysis of its trials ("science") against convincing case reports of fluoxetine induced suicidality that incorporated CDR and dose responsiveness ("anecdotes").⁴ After 1991, influential journals stopped taking case reports, and regulators became reluctant to warn about hazards for fear of deterring people from seeking treatment. As a result, significant and common harms of treatments are now being contested for decades.

Regulators have since extended postmarketing surveillance systems to allow reporting by non-medical clinicians and the public. They have also enhanced the quality of reports by incorporating cause and effect algorithms. While still subject to vast under-reporting, postmarketing reports can offer further information through use of proportional reporting ratios and related metrics. We also have increasing capability to track the effects of drugs and vaccines in registries, and it may be possible at some point to extract information from routinely collected clinical data. But despite this, there are increasingly long delays from first reporting of common and rare hazards to their acceptance by clinicians. Regular calls are made for new pharmacovigilance methods aimed at detecting rare events not

discovered by RCTs, but—while welcome—these will not solve the basic problem.

Many more effects are likely to come into view as a result of judicious efforts to reduce polypharmacy.¹⁸ While a patient's first exposure to a drug offers an opportunity to detect and explore its effects, so too does pausing treatment. Such dechallenge opportunities, however, return us to the original issue: how to restore confidence in doctors' and patients' ability to detect treatment effects objectively. There will be no RCTs to help or hinder us in this new domain, and establishing effects shown by deprescribing is not something for which we need novel pharmacovigilance techniques.

Objectivity

The rhetoric of RCTs claims that they offer a more objective way than case reports to establish treatment effects. This claim carries some weight in regulatory and legal settings because case reports are anonymous, which reduces their status to hearsay, whereas RCTs are viewed as being an exception to the hearsay rule, even though it is increasingly unlikely that anyone linked to a trial would ever be able to attest to the clinical reality behind it.

Furthermore, while the original use of statistical significance tests and confidence intervals had an understandable basis in the real worlds of fertilisers and astronomy, their use in RCTs does not have an established clinical reference point. In clinical RCTs these statistical approaches offer ways to describe data; they do not constitute objective knowledge. No RCT tells a clinician how to treat the patient in front of her—it's largely a matter of chance whether the drug works for that patient.

Some rarer and longer term effects of a drug, such as diabetes or birth defects, may require signal detection methods operating on large databases. But, for most adverse effects, seasoned clinicians allied to increasingly health literate patients are better placed than RCTs to determine causality. This also applies to many of the benefits of treatment, with patients and clinicians routinely making the call as to whether or not a treatment is working. This is particularly the case where common effects need input from patients able to distinguish treatment effects from superficially similar condition effects, such as the sexual or suicidal effects of a variety of drugs.

Science advances knowledge by generating data as we avail ourselves of new techniques, such as a drug, instrument, or method of evaluation, to throw up new observations. The mission of science, however, is not to replace judgment by technique. Patients constitute the core clinical dataset and are present in the flesh to be cross examined. It's when two parties with different perspectives agree that we can begin to have solid knowledge.

If there is a mismatch between what a doctor and patient judge is happening in response to a drug and what RCTs appear to show, this should not be framed as a choice between anecdote and science. The job of science in this instance is to account for what is happening to the patient and secondarily to explain any mismatch. Many factors may result in indubitable drug effects not showing up in RCTs, but a mismatch may also result from research misconduct, ghost writing, and data sequestration, as happened with antidepressants and suicidality. Industry RCTs generate numbers, but at present these numbers and the data behind them are sequestered so that no one can engage with them.

The way forward

Clinicians once had drug bulletins and case reports to provide rounded assessments of a treatment, but these sources are now difficult to access. Guidelines, in contrast, have become omnipresent but are based on RCTs so tend to extol the benefits of treatments and largely ignore their harms. Beyond guidelines, in the past 30 years drug companies have gained control of the distribution channels for information on drug effects—another problem that cannot be remedied with a new evaluation technique.

Drug labels may be a way forward. When adverse events are reported to drug companies, they, unlike regulators, are obliged to track down the patient or their doctor and their medical record. While companies make every effort to explain away the reported effects, when this is not possible—and standard assessments of causality point to a link—they include these effects in the label in carefully crafted postmarketing experience sections. However, clinicians commonly misread these as containing only anecdotes, in contrast to the apparent objectivity of the adverse events reported in RCTs, which are listed earlier in the label.

Neither companies nor regulators are as well placed as doctors and patients to make judgment calls on adverse effects. Reports of the same event from a series of named clinicians and patients could not be dismissed as hearsay in regulatory or legal settings, where a cross examination of an individual case remains the essence of objectivity, endorsed by authoritative texts.^{19 20}

We can foster earlier recognition of common adverse effects if doctors and patients send reports to companies and regulators and insist that their names and full contact details be left on the reports. Regulators should follow up such reports and make a causal determination, as companies do. This reporting is a role that falls on clinicians, who by virtue of prescription only arrangements are part of the regulatory apparatus. If clinicians insist that certain effects are happening on treatment, regulatory officials need to reflect this view. The fact that this is not happening is a clinical failure.

Journals such as *The BMJ* should publish case reports on adverse events that meet criteria for causality. The names of patient and clinician should be attached, and ideally both parties should indicate their willingness to be cross examined.

Clinical practice has been and should remain an exercise in judgment driven by the evidence that a doctor and patient have in front of them, rather than by thoughtless adherence to what a manual says. When it comes to wider debate about a drug's effects, we may need to designate RCTs as offering hearsay evidence, at least in respect of adverse events, given the limitations of RCTs and current sequestration of their data.

Biography

Dee Mangin is professor of family medicine at McMaster University, Hamilton, Ontario, and Otago University, Dunedin, New Zealand. Her work focuses on rational, person focused use of treatments and innovative models of primary care. David Healy is professor of psychiatry in Bangor University, Wales. His work focuses on the history of physical treatments in mental health and adverse events. In 2012 they cofounded RxISK.org, a patient powered website focusing on adverse event reporting and provision of information on drug side effects. This work includes development of TaperMD (<https://tapermd.com>), a tool that supports TAPER (<https://taper.fammedmcmaster.ca>), a team based clinical pathway for evaluation and reduction of inappropriate polypharmacy.

Competing interests: DH is an expert in a legal case involving antidepressant withdrawal. DM is a member of the Integrated Patient Safety Action Plan Medication Safety Leads Group, Patients for Patient Safety, Canada.

Patient and public involvement: DH and DM are involved with a patient adverse event reporting website, RxISK.org, whose operation has influenced the views offered here.

Provenance and peer review: Commissioned; externally peer reviewed.

- 1 Healy D. Citizen petition: Sexual side effects of SSRIs and SNRIs. *Int J Risk Saf Med* 2018;29:135-47. 10.3233/JRS-180745 29733031
- 2 Healy D, Le Noury J, Mangin D. Enduring sexual dysfunction after treatment with antidepressants, 5 α -reductase inhibitors and isotretinoin: 300 cases. *Int J Risk Saf Med* 2018;29:125-34. 10.3233/JRS-180744 29733030
- 3 Hill AB. Reflections on controlled trial. *Ann Rheum Dis* 1966;25:107-13. 10.1136/ard.25.2.107 5326437
- 4 Healy D. *Let them eat Prozac*. New York University Press, 2004.
- 5 Lasagna L. Thalidomide—a new nonbarbiturate sleep-inducing drug. *J Chronic Dis* 1960;11:627-31. 10.1016/0021-9681(60)90061-8 14414320
- 6 Lasagna L. Discovering adverse drug reactions. *JAMA* 1983;249:2224-5. 10.1001/jama.1983.03330400070028 6339764
- 7 Joelsing F. What the Gardasil testing might have missed. *Slate*. 2017. <https://slate.com/health-and-science/2017/12/flaws-in-the-clinical-trials-for-gardasil-made-it-harder-to-properly-assess-safety.html>.
- 8 Healy D. Did regulators fail over selective serotonin reuptake inhibitors? *BMJ* 2006;333:92-5. 10.1136/bmj.333.7558.92 16825234
- 9 Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007;356:2457-71. 10.1056/NEJMoa072761 17517853
- 10 Healy D, Mangin D, Antonuccio D. Data based medicine and clinical judgement. *Int J Risk Saf Med* 2013;25:111-21. 23796470
- 11 Healy D. *Pharmageddon*. University of California Press, 2012.
- 12 Venning GR. Validity of anecdotal reports of suspected adverse drug reactions: the problem of false alarms. *Br Med J (Clin Res Ed)* 1982;284:249-52. 10.1136/bmj.284.6311.249 6799125
- 13 Healy D. *The creation of psychopharmacology*. Harvard University Press, 2002. 10.4159/9780674038455.
- 14 Moore TJ, Furberg C, Cohen MR. Adverse drug effects in children under age 18. *Quarterwatch*. Jan 2014. Institute of Safe Medication Practices. www.ismp.org/Quarterwatch.
- 15 Gov.UK. Montelukast: reminder of the risk of neuropsychiatric reactions. Sep 2019. <https://www.gov.uk/drug-safety-update/montelukast-singulair-reminder-of-the-risk-of-neuropsychiatric-reactions>.
- 16 National Center for Health Statistics. Health, United States, 2016: with chartbook on long-term trends in health. 2017. Table 79, p 293. <https://www.cdc.gov/nchs/hus/contents2016.htm#079>.
- 17 Ho JY, Hendi AS. Recent trends in life expectancy across high income countries: retrospective observational study. *BMJ* 2018;362:k2562. 10.1136/bmj.k2562 30111634
- 18 Garfinkel D, Mangin D. Feasibility study of a systematic approach for discontinuation of multiple medications in older adults: addressing polypharmacy. *Arch Intern Med* 2010;170:1648-54. 10.1001/archinternmed.2010.355 20937924
- 19 Federal Judicial Center. Reference manual on scientific evidence. National Academies Press, 2011. <https://www.fjc.gov/sites/default/files/2015/SciMan3D01.pdf>.
- 20 Dukas MGN, Schwartz B, Mildred M. *Responsibility for drug induced injury*. IOS Press, 1998.

Published by the BMJ Publishing Group Limited. For permission to use (where not already granted under a licence) please go to <http://group.bmj.com/group/rights-licensing/permissions>