COMMENTARY

Cause, Effect, and Adverse Events: Evident-Based Medicine or Evidence-Based Medicine?

DAVID HEALY

_____

Abstract

From the late 1940s to 1991, the adverse effects of prescription drugs were primarily established through the publication of detailed case studies by doctors in medical journals. Subsequently, pharmaceutical companies would change the labels of medicines accordingly. This will be called "evident-based medicine" in this paper. After 1991, what is now called "evidence-based medicine" offers a markedly different view on establishing the adverse effects of a treatment, with randomised controlled trials (RCTs) held up as the gold standard. The differences between evidence- and evidence-based medicine are often framed in terms of the differences between specific and general causation. This article outlines the origins of these distinctions and the confusions they generate among both clinicians and the general public.

Keywords: *evident-based medicine, evidence-based medicine, specific causation, general causation, regulation*

Introduction

It is critical to good clinical practice to not only establish the links between a medication and its effects, both good and bad, but also to determine the best way to establish these links — both are central to medico-legal practice and public health policies. Our views on how best to establish such links have shifted dramatically since the 1930s, when the first antibiotics were introduced.

In this paper, we outline what was the standard medical position, "evident-based medicine", for the first five decades of the modern era. We then illustrate how the link between suicidality and antidepressants meshed with the emergence of evidence-based medicine to create a new narrative. While evidence-based medicine was initially viewed as a means to restrict pharmaceutical companies and establish boundaries for their claims, in medico-legal settings, evidence-based medicine became a means to undermine the judgement of both individual clinicians and patients [1].

The Standard Medical Position

In 1947, Austin Bradford Hill undertook the first randomised controlled trial (RCT), comparing tuberculosis patients who had been treated with streptomycin to those left untreated [1]. Hill's trial was not inspired by Ronald Fisher's famous thought experiment [2], which posited that randomisation may serve as a means of controlling unknown confounders, as the histories of RCTs might suggest. Rather, Fisher was attempting to mathematise expert knowledge — not conduct a trial. If an expert knew what he was doing, and if randomisation controlled for all trivial confounders, then the only thing that could interfere with the expert being right was chance, to which a statistically significant value could be applied [2]. This position might hold true as a mathematical abstraction, but does not apply to actual medical practice.

Hill's RCT proved that streptomycin works, but it failed to observe that its effects are weak, that patients developed a tolerance to it over time, and that some of them went deaf from the treatment. An earlier Mayo Clinic trial, which had a control group but was not randomised, had in contrast, demonstrated that streptomycin's effects were weak and short-lived and that it had significant adverse effects [3].

Hill used randomisation simply as a means of fair allocation [4, 5]. He did not assume that doctors were experts who knew what they were doing and did not assume that randomisation would control for unknown unknowns, one of which may be clinical ignorance. If a doctor is not aware that a new drug can cause a particular problem, they may not notice or record it, and the effect will not be reported in the academic literature.

Clinicians did not rush to adopt RCTs following Hill. In the 1950s, the leading American advocate of RCTs was Louis Lasagna, who considered them efficient demonstrations of efficacy [6]. He proposed that companies should be required to use them to demonstrate efficacy in addition to safety as part of the 1938 Food, Drugs, and Cosmetics Act [7].

In 1960, Merrell, hoping to market thalidomide in the US, asked Lasagna to test the efficacy of the drug. His RCT demonstrated that thalidomide was an effective and safe hypnotic with no side effects — the study failed to highlight the sexual dysfunction, agitation, suicidality, and peripheral neuropathy it is now known to cause [8].

In 1962, the thalidomide birth defect crisis forced American politicians to act. They modified the 1938 Food, Drugs, and Cosmetics Act to require the use of RCTs to demonstrate the efficacy of a treatment. This forced pharmaceutical companies to become the main sponsors of (what we now call) RCTs — in fact, given that these studies aim to meet standards set by regulators, rather than being designed to inform clinical practice, they are arguably better termed randomised controlled assays (RCAs), rather than RCTs [9].

Now, contract research organisations (CROs) carry out company studies, with medical writing companies presenting the results and the study data sequestered behind commercial confidentiality clauses. Rather than meeting legal or scientific criteria for evidence, assay materials like this are, strictly speaking, hearsay. The results of company assays are often called "evidence" in medical settings; but legally, they are more appropriately designated as "hearsay", in that no one can be brought into court to testify to the context of any information drawn from these assays.

In contrast to the role of RCAs after 1962, the place of RCTs in clinical practice remained uncertain. Lecturing to doctors in 1965, Hill said good clinical interviewing — listening closely to and looking at patients — was the best way to establish the effects of a drug on a patient. RCTs, he said, give us the average effects of a drug. This can tell us whether a relatively weak drug has some effect, for instance, but it tells doctors little about how to treat the patient in front of them.

> Frequently with a new discovery… the pendulum at first swings too far… Given the right attitude of mind, there is more than one way we can study therapeutic efficacy. Any belief the controlled trial is the only way would mean not that the pendulum had swung too far but that it had come off its hook. [4]

By the early 1980s, however, many doctors were arguing that case reports, like the ones Hill advocated for, offered "the least sophisticated and scientifically rigorous… method of detecting new adverse drug reactions" [10]. We should depend instead on RCTs, they claimed, which in practice by then meant the company RCAs.

In response to this point, in 1983, Lasagna re-articulated what at that point was still the standard clinical position, stating that "This [claim] may be true in the dictionary sense of sophisticated meaning 'adulterated'… but I contend that spontaneous reporting is more 'worldly-wise, knowing, subtle and intellectually appealing' than grandiose, expensive RCTs" [11].

In this cited passage, Lasagna is advocating for an "evident-based medicine" approach, which was the gold standard for Hill and almost all clinicians through to 1990. There is a wonderful irony here in that Lasagna ran an RCT on thalidomide that missed all of its significant adverse effects, and so he knew from firsthand experience that RCTs, if not designed to specifically detect effects, may miss them completely. His response offers an implicit critique of what came to be called "evidence-based medicine" in the 1990s.

Evident-Based or Evidence-Based Medicine?

In 1990, three Boston clinicians reported on six people who became suicidal while on fluoxetine (Prozac), a novel selective serotonin reuptake inhibiting (SSRI) antidepressant [12]. Using standard methods to establish causality — namely, reviewing prior medical histories, assessing each case for all possible causes, reducing the dose or halting the drug, rechallenging if

indicated, and paying heed to patients able to distinguish illness-induced suicidality from drug-induced suicidality — it was possible for doctors to establish that fluoxetine could cause suicidal events in some people.

Other groups published similar findings, adding evidence of causality such as the fact that administering an antidote could mitigate the effect. A Yale group reported that 1 in 7 children became suicidal on fluoxetine [13], a figure later replicated in paroxetine trials [14].

In response, Eli Lilly, the makers of Prozac, claimed in a 1991 *BMJ* article that an analysis of their trials did not support this claim [15]. Lilly characterised the published cases as anecdotal, adding that the plural of anecdote is not data, that depression and not fluoxetine caused suicides, and that RCTs are the science of cause and effect. This article effectively created evidence-based medicine.

In fact, Lilly's Beasley et al article demonstrates an excess of suicidal events on Prozac compared to placebos. This excess was downplayed on the basis that it was not statistically significant. When an event filed as a "placebo suicide" is returned to the pre-randomisation wash-out phase of the trial from whence it came, the excess of events on fluoxetine is statistically significant [16]. The figures for suicidal events on paroxetine and sertraline, other SSRIs then being developed, showed a similar excess of suicidal events — and a regulation-breaching transfer of events from the wash-out phase to the placebo arm of the trials [16].

In addition, regulators, companies, and medical journals, then and now, have applied statistical significance tests and confidence intervals to the trial data in a manner generally condemned by medical statisticians [9, 17].

Company trials are not demonstrations that we know what we are doing, as outlined by Fisher [2], which might make significance tests appropriate. Additionally, the variation among subjects in company trials is not the same as the random errors caused by faulty instruments, that Gauss addressed with confidence intervals in astronomy [9, 17]. Technically, we should view an excess of suicidal events as an excess of events until there is consensus on how these might have arisen.

These statistical approaches, however, offer a stop-go mechanism for approvals from regulators and journals and for warnings they might feel are company business rather than journal or regulatory business. Statistics used in this way are not anchored in the real world. They are models. Their use to describe the distribution of data in company assays has immense rhetorical value, but it risks compromising clinical care.

Specific and general causation

In the late 1980s, important medico-legal events unfolded that impacted how the arguments developed from 1990 onwards. There was a growing appreciation that classical

clinical methods alone could not settle claims linking medicines such as doxylamine to birth defects, linking other drugs to cancers that might appear years later, or linking devices like breast implants to connective tissue diseases. In legal cases such as these, clinical evaluations need to be supplemented with epidemiological and other methods [18].

In contrast, in legal cases through to the late 1990s, whenever an injury developed in clinical practice before a doctor's or patient's eyes — unlike the reports in drug-induced birth defect cases — the experts' medico-legal reports continued to offer views about the cause and effect using the methods adopted by Teicher et al for fluoxetine-induced suicidality and advocated by Lasagna.

From the late 1990s, however, company lawyers in SSRI and suicidality cases embraced Lilly's argument that RCTs are the science of cause and effect, in contrast to anecdotal case reports. This argument entered the medico-legal domain in the form of distinctions between general and specific causation.

"General causation", then, refers to evidence from epidemiological and randomised studies. Absent such evidence that a drug could in principle cause a particular event, courts were invited to dismiss expert reports which to a reasonable degree of medical certainty demonstrated drug X had caused problem Y in a particular case — even in cases offering a clear link. Judges, it was argued, had a duty to gatekeep science and not admit junk science. This led to pre-trial Daubert hearings in cases where adverse events followed medication or treatment [19].

"General causation" implies an objectivity not found in "specific causation". However, general causation also refers to average effects that happen in no one individual, while specific causation refers to a considered judgement of what has happened in individual cases.

This has led to two decades of confusion, which a 2021 Southern District of California opinion may clarify [20]. It stated that:

> Courts **define general causation** to mean 'whether the substance at issue had the capacity to cause the harm alleged'.

Applying this definition to the accepted capacity of husbands to murder wives, let us consider what kind of evidence would help in a legal case. Using company views on general causation, husbands will always be acquitted as controlled studies will always show that husbands *on average* do not murder wives.

Traditional legal approaches to causation — which depend on an examination and cross-examination of individual case evidence — would, in contrast, enable courts to decide, to a reasonable degree of certainty, that the husband had murdered his wife. It would seem entirely inappropriate to describe a legal verdict in a case like this as "anecdotal".

Similarly, while controlled studies now show that on average, a drug can cause suicidal events, standard medico-legal approaches to case analysis, as advocated by Lasagna and Teicher, can enable a court to decide the drug has *not* caused this suicidal or homicidal event if the clinical features of the case do not map onto a strong specific causation case.

There are clear cases where examination and cross-examination of individual cases must be supplemented with input from controlled studies, but in cases of drug-induced injury that depend primarily on specific causation factors, a "specific causation" approach trumps what has been termed "general causation", and, indeed, offers the best basis for establishing general causation as recently legally defined.

This is not just a medico-legal matter. Clinical practice in general is like the practice of law. It calls on doctors to come to a consensus with their patients and colleagues as to whether, in this specific case, a medicine is causing a problem or not. If the consensus is that the illness rather than the medicine is causing the problem, the appropriate clinical input might lead to double the treatment dose, whereas if the view taken is that the medicine is causing a problem, then the appropriate course of action is to reduce the dosage or stop it entirely.

Regulation and therapeutics

The distinction between specific and general causation has widened the divide between therapeutics and regulation. Clinicians necessarily make decisions about specific causation in daily practice, along with decisions following on from that, such as reducing or increasing the dose of a medicine. Up to 1991, by publishing case reports, they played a significant role in establishing adverse events in the minds of their colleagues.

Companies are legally responsible for drug labels. Until 2000, companies assessed the adverse effects reported to them by doctors or patients in the same way as clinicians did in clinical practice. This approach was embodied in the guidance that the Food and Drug Administration (FDA) put out for companies, recommending that company doctors contact people with possible adverse events to determine causation [21, 22].

After 2000, the argument that controlled studies were the only way to determine adverse effects meant that it was no longer acceptable for companies to determine cause and effect using specific causation methods, as per FDA guidance. This problem was solved in several steps.

First, incoming reports were designated as having been "reported" to a company and then passed on to regulators, rather than being examined for causality.

Second, as was standard in Europe, American doctors and patients were encouraged to report to the FDA rather than to manufacturing companies. Standard regulatory practice removes personal identifiers. Regulators, therefore, can never specifically link a drug to an individual event.

Third, company assays have a treatment benefit as the primary endpoint. This implies that the benefit is the most common effect and other effects are rare or occur outside the timeframe of the study. As company assays are not designed to explore other effects, these rarely feature to a statistically significant extent. As a result, these assays support positive risk–benefit claims for drugs.

The benefit, however, may not be the most common effect of a drug. The sexual effects of SSRIs are immediate (evident within an hour of taking the first pill) and several times more common than any mood benefit [23].

In 1991, the FDA also opted not to warn users about the excess of suicidal events on SSRIs on the basis of a risk–benefit analysis, claiming that such warnings might deter people from seeking the benefit [19]. A system geared toward not warning patients about hazards in order to avoid deterring them from seeking a benefit is not one that is incentivised to assign causality to a treatment hazard.

Discussion

In the 1990s, influenced by the emergence of evidence-based medicine, which prioritised RCTs over "evident-based medicine" reports, clinical journals stopped accepting clinical reports for publication. For journals, there was less of a legal risk and more money to be made from reprints of company assays demonstrating treatment benefits than from case reports of adverse events.

Free pharmacovigilance bulletins, which flagged up the possible hazards of new drugs, were sent to clinicians in many countries; but this stopped in the late 1990s. Instead, clinicians got regular and free access to the latest treatment guidelines, which only outlined the benefits of treatment and were based primarily on published company assays.

In most countries, health services now follow treatment guidelines, as the best possible evidence is supposed to produce the best patient outcomes as well as the most efficient and legally defensible services. This has increasingly put doctors at risk of moral hazard. There is no longer any incentive for them to recognise the harms that could arise from a treatment.

These developments have widened the pharmacovigilance divide between the clinical need to have rapid access to reliable information on a full range of drug effects, especially for recently released medicines, and the mandate of regulators, which is to monitor drug labels and seek company agreements to include representative, or "average", treatment effects on drug labels.

It can now take decades for serious and common problems, such as post-SSRI sexual dysfunction (PSSD), or the behavioural effects of isotretinoin, fluoroquinolone antibiotics, and leukotriene antagonists, to turn up on drug labels.

Longstanding concerns about the neuropsychiatric effects of montelukast, a leukotriene receptor antagonist used for asthma, for instance, led to successive modifications to the label, culminating in a black box warning in 2020, over two decades after the release of the drug. The FDA characterised this warning as a response to patients' and medical professionals' convictions regarding the associated hazards, rather than being based on controlled studies [24].

A 2023 British Commission on Human Medicines Isotretinoin Expert Working Group reviewed the data regarding the psychiatric and sexual side effects linked to isotretinoin against a background of convincing case reports of harms dating back decades. It concluded that an association could not be ruled out and that the individual experiences of patients and families raised concerns that warranted warnings [25]. Dermatologists branded the Medicines and Healthcare products Regulatory Agency (MHRA) requirement to raise a warning a retreat from "evidence" to "misinformation" [26].

Viewing convincing "evident-based medicine" reports as misinformation suggests that warnings by regulators are now read like the "May contain nuts" labels on foods — which are viewed, even by those with severe nut allergies, as companies and regulators protecting themselves from liability, rather than as information deserving serious consideration.

It appears that unless doctors play a part in generating warnings, as had been standard practice up to 1990, they will no longer take drug labels seriously — even though regulators traditionally have, and to this day still do designate prescription drugs as unavoidably hazardous. This is not a recipe for safe clinical practice.

In addition to compromising clinical care, this practice risks thwarting justice. Courts expect experts on either side of a legal action to be able to agree on the specifics of a case. If one clinical expert is convinced by the specific details but another is not convinced, primarily because they view controlled studies as trumping individual cases, no matter how convincing the individual case, the legal system is paralysed.

Lay jurors may now be expected to form a perspective on the hearsay nature of company assays, the appropriateness of ghostwritten studies that transform negative results into positive reports, and the sequestration of company data. They may be able to make an appropriate diagnosis in very clear-cut clinical cases; but in more complex cases, can it be left to jurors to decide the medical diagnosis?

This also applies in clinical practice. As things stand, patients may be correct to view their treatment as causing their problem; but if the "expert", their doctor, denies that a link is possible, then their predicament becomes a very difficult one.

"Advertisements" now tell both doctors and patients that regulators such as the European Medicines Agency (EMA) employ thousands of scientists to keep us safe [27]. But medical regulatory authorities are bureaucratic agencies, with no training or track record in establishing the adverse effects of drugs. Faced with other faulty products, we contact manufacturers and experts with domain expertise, not regulators. So why not do the same here?

As the adverse effects of certain treatments have increasingly failed to register in the minds of both the public and medical professionals, our life expectancies have begun falling again — just as our fertility rates fell below reproductive replacement rates — leaving us to face a polypharmacy pandemic [7].

Reducing medication burdens has become a pressing priority, meanwhile [28]. The peculiarity/variability of the medical burden at an individual level means that no randomised or other controlled studies nor guidelines can ever steer us through the medical minefields we now have. Safety considerations call for close attention to what is evident in specific patients and what likely differs from patient to patient. This polypharmacy pandemic calls for a restoration of the premium clinical expertise enjoyed all the way to 1990.

Authors: David Healy (david.healy@rxisk.org, https://orcid.org/0000-0002-6340-9247), Chief Scientific Officer, Data Based Medicine, UNITED KINGDOM.

Copyright and license

References

1. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *Br Med J (Clinical Research Ed.).* 1948 Oct 30; 2(4582): 769–82.
2. Fisher RA. *The design of experiments.* London: MacMillan; 1935.
3. Hinshaw H, Feldman WH. Evaluation of chemotherapeutic agents in clinical tuberculosis. *Am Rev Tuberc.* 1954; 50 (3): 202–13.
4. Hill AB. Reflections on controlled trial. *Ann Rheum Dis.* 1966 Mar; 25(2): 107–13. https://doi.org/10.1136/ard.25.2.107
5. Chalmers I, Dukan E, Podolsky S, Smith GD. The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries. *J R Soc Med.* 2012 May; 105(5): 221–27. https://doi.org/10.1258/jrsm.2012.12k029
6. Lasagna L. Back to the future: Evaluation and drug development 1948–1998. In: Healy D (ed), *The Psychopharmacologists,* Vol 2. London: Arnold; 1998; pp 135–66.
7. Healy D. *Shipwreck of the singular: Healthcare's castaways.* Toronto: Samizdat Press; 2021.
8. Lasagna L. Thalidomide – a new non-barbiturate sleep-inducing drug. *J Chronic Dis.* 1960 Jun; 11: 627–31. https://doi.org/10.1016/0021-9681(60)90061-8
9. Healy D. Randomized controlled assays and randomized controlled trials: A category error with consequences. *Ethical Hum Psychol Psychiatry.* 2023 Oct; 25(2): 119–34. http://dx.doi.org/10.1891/EHPP-2023-0006
10. Rossi AC, Knapp DE, Anello C, O Neill RT, Graham CF, Mendelis PS, Stanley GR. Discovery of adverse drug reactions: A comparison of selected phase IV studies with spontaneous reporting methods. *JAMA.* 1983 Apr; 249(16): 2226–8. https://doi.org/10.1001/jama.249.16.2226
11. Lasagna L. Discovering adverse drug reactions. *JAMA.* 1983; 249 (16): 2224–5. https://doi.org/10.1001/jama.1983.03330400070028
12. Teicher MH, Glod C, Cole JO. Emergence of intense suicidal preoccupation during fluoxetine treatment. *Am J Psychiatry.* 1990 Feb; 147(2): 207–10. https://doi.org/10.1176/ajp.147.2.207
13. King R, Riddle M, Chappell P, Hardin M, Anderson G, Lombroso P. Emergence of self-destructive phenomena in children and adolescents during fluoxetine treatment. *J Am Acad Child Adolesc Psychiatry.* 1991 Mar; 30(2):171–6. https://doi.org/10.1097/00004583-199103000-00003
14. Le Noury J, Nardo JM, Healy D, Jureidini J, Raven M, Tufanaru C, Elia A-J. Restoring study 329: A randomised, controlled trial of the efficacy and harms of paroxetine and imipramine in the treatment of adolescent major depression. *BMJ.* 2015 Sep 16; 351: h4320. https://doi.org/10.1136/bmj.h4320
15. Beasley CM, Dornseif BE, Bosomworth JC, Sayler ME, Rampey Jr AH, Heiligenstein JH, et al. Fluoxetine and suicide: A meta-analysis of controlled trials of treatment for depression. *BMJ.* 1991 Sep 21; 303(6805): 685–92. https://doi.org/10.1136/bmj.303.6804.685
16. Healy D. Did regulators fail over selective serotonin reuptake inhibitors? *BMJ.* 2006 Jul 8; 333(7558): 92-5. https://doi.org/10.1136/bmj.333.7558.92
17. Healy D. The antidepressant tale: Figures signifying nothing? *Adv Psychiatr Treat.* 2006 Jan 2; 12(5): 320–8. https://doi.org/10.1192/apt.12.5.320
18. Angell M. *Science on trial.* New York: WW Norton & Co.; 1996.
19. Healy D. *Let them eat Prozac.* New York: New York University Press; 2004.
20. United States District Court, S.D. California. *In re incretin-based therapies products liability litigation.* California: United States District Court, Southern District of California; 2021. Case No.: 13-md-2452-AJB-MDD.
21. Food and Drug Administration. Guidance for industry: Good pharmacovigilance practices. Rockville, MD: US Dept of Health and Human Services; 2005 Mar [Cited 2025 Jan 8] Report: Available from: https://www.fda.gov/media/71546/download
22. Food and Drug Administration. Conducting a clinical safety review of a new product application. and preparing a report on the review. Rockville, MD: US Dept of Health and Human Services; 2005 Feb [Cited 2025 Jan 8]. Available from: https://fda.report/media/71665/Conducting-a-Clinical-Safety-Review-of-a-New-Product-Application-and-Preparing-a-Report-on-the-Review.pdf
23. Healy D. Antidepressants and sexual dysfunction: A history. *J R Soc Med.* 2020 Apr; 113(4): 133–5. https://doi.org/10.1177/0141076819899299
24. Clarridge K, Chin S Eworuke E, Seymour S. A boxed warning for montelukast: The FDA perspective. *J Allergy Clin Immunol Pract.* 2021 Jul; 9(7): 2638–41. https://doi.org/10.1016/j.jaip.2021.02.057
25. Mahase E. Isotretinoin: Regulator adds prescribing safeguards after review of side effects. *BMJ.* 2023 Nov 2;383: p2545. https://doi.org/10.1136/bmj.p2545
26. Affleck A. New isotretinoin legislation is based on misinformation. *BMJ.* 2024 Feb 26: 384: q438. https://doi.org/10.1136/bmj.q438
27. European Medicines Agency. Regulatory Science Research Needs: 2024 update – Draft. 2024 Nov 13[Cited 2025 Jan 8]. Available from: https://www.ema.europa.eu/en/documents/other/draft-regulatory-science-research-needs-2024-update_en.pdf
28. Garfinkel D, Mangin D. Feasibility study of a systematic approach for discontinuation of multiple medications in older adults: addressing polypharmacy. *AMA Arch Intern Med.* 2010 Oct 11; 170(18): 1648–54. https://doi.org/10.1001/archinternmed.2010.355