

COMMENTARY

Drug safety: The roles of big data and clinical experience

DAVID HEALY

Abstract

Following the pharmacovigilance crises of 2004 involving the use of Vioxx and antidepressants in minors, medicine regulators turned to big data, also called real-world evidence, to support their efforts to establish the safety of treatment protocols. In many areas of drug development, big data can clearly play a part; but to date, it has not helped resolve safety issues. Developments in artificial intelligence may help clarify the respective roles of big data and clinical expertise in pharmacovigilance in surprising ways.

Keywords: *big data, real-world evidence, specific causation, general causation, regulation*

Introduction

In 1999, a new analgesic named Vioxx (rofecoxib) was launched. In 2004, evidence linking Vioxx to heart attacks led to its removal from the market [1]. Reports indicating that Vioxx might have given rise to thousands of cardiac events made it imperative to identify methods for establishing not just whether a drug might be associated with adverse events but how frequently they might occur.

In 2004, concerns also emerged regarding the suicide risks associated with the use of selective serotonin reuptake inhibiting (SSRI) antidepressants by minors. In addition, it became clear that more than half of the adult and paediatric antidepressant studies undertaken for the licensing of drugs showed negative outcomes but were published as positive [1]. Primary data from company studies were also inaccessible. There were, therefore, calls for the Food and Drug Administration (FDA) to do more to ensure the safety of treatments [2, 3].

In 2005, senators Chuck Grassley and Charles Dodd proposed the FDA Safety Act, which would split the FDA into licensing and post-approval divisions [4]. The licensing division would approve drugs, and the post-approval division would monitor their safety. This split never happened. Grassley and Dodd also proposed registering all clinical trials, which was incorporated in a 2007 FDA Amendments Act as clinicaltrials.gov [3]. A proposal to track links between clinicians and the industry was incorporated under the "Sunshine Act," within the 2010 Affordable Care Act [1].

Meanwhile, an internal FDA review concluded that the agency could not fulfil its mission because:

1. "its scientific base has eroded, and its scientific organizational structure is weak.

2. its scientific workforce does not have sufficient capacity and capability.
3. its information technology (IT) infrastructure is inadequate" [5].

Just then, a new route to potentially enhancing drug safety was opening up. New abilities to analyse Britain's General Practice Research Database both Medicare, Medicaid, and insurance databases offered the possibility of more easily identifying adverse events and estimating their frequency. This gave rise to what is now called "real-world evidence". Real-world evidence is often termed "big data" but neither term has been explicitly defined. Big data, it was claimed, would speed up therapeutic discovery, aid research on outcomes, and enhance safety surveillance.

Big data was incorporated in the 2007 FDA Amendments Act, which mandated that the FDA should create the Sentinel Initiative to collate data from medical encounters via electronic health records, hospital data, prescription redemptions, medical claims data, and adverse event reports. These have since been supplemented by data collected from social media reports and data from apps associated with wearable medical and fitness devices [2, 3].

Sentinel expanded further in 2017 into the Innovation in Medical Evidence Development and Surveillance (IMEDS) programme [6]. The European Medicines Agency (EMA) and Health Canada followed suit and established similar initiatives to aggregate medical data. In 2022, the EMA established the Data Analysis and Real World Interrogation Network (DARWIN EU). The FDA, the EMA, and Health Canada also met to align their collective efforts.

As the scale of such big data operations grew, the EMA and the FDA outsourced data collection to private companies, some of which are life science companies that also run studies for pharmaceutical companies and write up the study results — but (at least according to the regulators) they play no part in the safety decisions that may result [7].

Big data: the promise and the practice

A 2015 review noted that Sentinel "has yet to be the primary data source in identifying a single new drug risk that led to a significant regulatory action such as a drug withdrawal, boxed warning, restriction or contraindication" [8]. A study by Madigan et al [9] inserted some established adverse events into databases similar to those used by Sentinel and demonstrated that they could

fail to identify the event as an outcome. Many others have tested similar databases and failed to find known adverse events as well.

Shortly before pharmacovigilance turned to big data, polypharmacy had also emerged as a problem. Polypharmacy called for second-generation pharmacovigilance [10]. Establishing cause and effect between a drug and an event such as a fall or confusion is straightforward when a person starts taking a single drug; but it becomes more difficult when someone is on four drugs that might cause falls and six additional drugs.

Solving polypharmacy might seem to call for the use of big data, but consider all the possibilities for, say, an event such as a suicide attempt while on an SSRI:

- The event may be prominent on starting treatment.
- It can remain for the entire time the person is on treatment.
- It may disappear if a person takes a recognised antidote, such as mirtazapine.
- It may disappear if the person takes an unrecognised antidote.
- It may not appear on a lower dose of treatment.
- It may only appear on a lower dose of treatment.
- It may only appear in a person genetically sensitive to the effect.
- It may not appear on starting treatment but may do so only on stopping the drug.
- If it appears on stopping the drug, it may be relieved by restarting.
- If it appears on stopping the drug, it may be made worse by restarting.
- If it appears on stopping the drug, it may be self-limiting over a course of time.
- If it appears on stopping the drug, it may endure for decades after.
- If it appears on stopping the treatment and lasts decades, it may be pitched as evidence that the drug does not cause the problem.

A potential antidote such as mirtazapine may block the suicidality (or other adverse effects) of an SSRI; but:

- may also itself cause suicidality on starting mirtazapine
- or may cause suicidality on stopping mirtazapine.

Mirtazapine is just one of the several other medications adolescents taking an SSRI may be on simultaneously —

increasingly, they are likely to be on up to 10 other drugs, some of which may combine with an SSRI to produce effects never seen before, and thus not yet present in coding dictionaries, and therefore invisible to big data.

Big data at present cannot resolve these complexities — in part because the effects of medicines on patients are never binary and therefore not readily amenable to algorithmic approaches. Without human judgement to differentiate between superficially similar events, big data risks making these events “disappear” from scrutiny. Garbage in, garbage out, as the saying goes. We need good clinical interviews to cut through the complexities, distinguish between similar-looking events, and form tentative judgements.

In addition, while the rhetoric of real-world evidence suggests real-time evaluation — and big data may offer quicker answers than clinical trials — amassing big data takes years, whereas clinicians may need to decide whether a new drug is causing problems within weeks of its launch.

Finally, despite good intentions, it may be misguided to attempt to improve drug safety by mandating that regulators also serve as assessors. Regulators are tasked with licensing drugs and labels (that the companies have manufactured and written, respectively). The companies are legally liable for safety, not regulators.

There will inevitably be a difference in outcomes if regulators are tasked with detecting signals and agreeing with companies on a change of wording in drug labels and if they are expected to be medical generalists who determine clinical causality with patient safety as their goal. The difference between the words “detecting” and “determining” brings this home.

Regulators and adverse events

Among their modes of action, the first tricyclic antidepressants (TCAs) are serotonin reuptake inhibitors. TCAs are more potent antidepressants than SSRIs and can cure melancholia, a severe mood disorder that SSRIs cannot help. Melancholics lose interest in everything, including sex. Nevertheless, doctors in 1960 could distinguish between a melancholia-induced and drug-induced sexual problem [11].

By 1970, we knew that one tenth of the antidepressant dose of clomipramine could numb genitals within 30 minutes. This effect could be used to treat premature ejaculation [11]. Genital numbness occurs in most people on SSRIs, but it does not feature on the labels of these drugs. By the 1980s, some people were known to have enduring sexual dysfunction after stopping SSRIs [11].

In some phase 1 company studies on SSRIs, half the healthy volunteers reported experiencing sexual dysfunction, which in some cases, continued beyond the brief 3-week exposure. Investigators (including myself) in later clinical studies were

told not to ask about sex, enabling companies to state that sexual dysfunction affected less than 5% of study participants [11].

By 1990, there was a steady stream of reports of SSRI-induced sexual dysfunction on the commencement of, during, and following treatment. Meanwhile, SSRIs were used widely for premature ejaculation, and one, dapoxetine, was specifically licensed for this purpose.

It took 15 years for post-SSRI sexual dysfunction (PSSD) publications to appear, in part because no one affected wanted their name in the public domain [12]. Two years after that first article, Audrey Bahrack, an academic psychologist at the University of Iowa, approached Senator Grassley, the senator for Iowa at the time, who flagged the issue with the FDA. Stephen Mason, an FDA acting assistant commissioner, responded to Senator Grassley, stating:

It is not possible for FDA in any individual case to determine if the discontinued SSRI, the underlying disorder, or some other unknown factor is responsible for causing sexual dysfunction [13].

This quote brings out the FDA's and big data's problems in determining adverse events. Contrast this with FDA Guidance for reviewers assessing adverse events [14]:

Discuss the adequacy of the applicant's efforts to detect specific adverse reactions that are... predicted on the basis of the drug class (e.g., sexual dysfunction with SSRI antidepressants)... [and] also discuss whether the applicant should have made efforts to assess certain events that it did not assess... [and] also discuss pertinent absence of findings for a drug (Example ... sexual dysfunction (any antidepressant).

If doctors stated that it is not possible in an individual case to determine whether the discontinued SSRI, the underlying disorder, or some other unknown factor has caused sexual dysfunction, suicidality, or protracted withdrawal, medical practice would have to cease overnight. Faced with a suicidal patient, a doctor must *decide* whether the drug or the condition is the cause. The appropriate responses for these two causes are diametrically opposite — in one, the dose must be lowered; in the other, increased. Get it wrong and the patient may die. The FDA, however, cannot interview patients or doctors to supplement the details they have.

A decade later, with my colleagues, I submitted petitions to the FDA and the EMA to have PSSD added to antidepressant labels [15]. I persuaded over 80 individuals with PSSD to append their names to reports of their condition and convinced over 30 doctors to write reports on their patients' conditions, indicating that there was no way other than the drug to explain them. The FDA and EMA were offered the list of names along with the reports to facilitate interviews that might establish causality. The FDA declined the offer. The EMA accepted but, as per its standard operating procedure, removed all names and contacted no one.

Any doctor who learns that a patient could take a hard-bristled brush to her genitals and feel nothing would realise that this must be a drug-induced problem. No psychiatric disorder gives rise to states like this. This is why interviewing patients can eliminate confounders.

In 2004, the FDA opened MedWatch to reporting by the public. Prior to this, in the US, the public could report adverse events to companies alone, who then determined causality in the way clinicians did. In a significant indicator that regulators cannot determine treatment-related causality, US companies began encouraging doctors and the public to report adverse events to the FDA rather than to them. When events are reported to companies, they now file the report and forward them to the FDA, rather than making a judgement of causality themselves [7].

From big data to artificial intelligence

Before clinical medicine turned to epidemiology and controlled trials, pharmaceutical companies had already turned to big data for drug development. Prior to 1980, drugs were discovered by noting unexpected effects in people or by judging the effects of likely candidates on animal models of diseases or physiological systems [7].

A growing knowledge of receptors, and the growing ability of computational chemistry to tailor molecules to bind to receptors, made it possible to screen thousands of molecules per day. Pharmacologists expected that this rational drug development would produce an increased flow of new drugs. Computational chemistry did lead to the discovery of a number of ligands that bind to receptors, but it did not lead to clinically significant new medicines [1].

The Human Genome Project, similarly, fanned expectations that our increased ability to generate genomic data would lead to safe and effectively tailored medicines. But there was no increase in clinically significant medicines [7].

As noted above, developments in our database processing capabilities led to improved detection of signals but have not led to the determination of significant adverse events. In contrast, our ability to process large amounts of epigenetic data for "signals" of possible harm and to pick out teratogens and carcinogens based on their epigenetic profiles may contribute to drug safety [16, 17].

Epigenetic studies offer profiles, rather than the "average effects" that epidemiological and controlled studies generate. Average effects run the risk of suggesting that there are no hazards — on average. Epigenetic data is more readily viewed as containing heterogeneous subsets of predictive factors that support diverse rather than unitary outcomes.

Artificial intelligence (AI) has also enhanced our ability to produce new ligands without leading to the creation of new medicines [18, 19]. One reason for this is that availing of our increasingly precise chemical capabilities is like availing of

the pixelated radiographic imagery that enables AI to outdo radiologists in detecting abnormalities. Clinical encounters, in contrast, are “pixelated” these conversations are not reducible to numbers.

What works clinically is often at odds with the myth of the magic bullet. The pharmaceutical industry seeks magic bullets that hit a defined target without collateral damage, but many medicines offer therapeutic principles that compensate for a loss of function by inhibiting another function. This level of complexity is at present beyond the capacity of algorithms to resolve.

Our emerging AI capabilities are likely seen by companies and regulators as allowing them to access even bigger datasets, thereby enhancing their abilities to interrogate real-world evidence. This approach seems to make sense, but our track record while using such approaches suggests that AI will fail in this domain. At least, there is no evidence to date that AI used in this way is likely to contribute to a significant degree.

In contrast to AI-enhanced big data crunching, if we develop what is often called “full AI” — systems that can learn — these systems might settle the dispute between clinical experience and big data. If an AI system has a brief to keep us alive and the scope to learn from mistakes, will it trust medical literature that only reports the average effects of drugs, or will it trust doctors’ abilities to distinguish between specific treatment effects?

Mathematical loss functions can constrain machines to learn. That learning depends on binary options — but, as noted above, medicines do not typically provide binary options. In clinical practice, the binary option lies in deciding between two judgement calls. Does a good clinician’s ability to distinguish good and bad treatment effects produce better results than medical practice based on company studies that yield average effects that rarely (if ever) implicate a medicine as the cause of a problem? While an initial view of AI suggests that it may contribute to drug safety through “better” big data analysis, this latter consideration suggests the almost diametrically opposite conclusion. We will only be able to grasp the future of AI-supported drug safety protocols if the thought experiment just outlined is actually ever carried out.

Conclusion: From Schrödinger to artificial intelligence

The contest between clinical determination and signal detection maps on to arguments about general and specific causation. This offers a medico-regulatory version of Schrödinger’s thought experiment, which posited that if there were a cat in a closed box, we can never know if it is dead or alive unless we open the box.

Quantum physics uses probabilities and statistics to manage random subatomic data to make predictions. This big data strategy contrasts with events such as the apple falling on Newton’s head, which led to Einsteinian physics. Schrödinger

said that quantum physics offers no way for us to know whether the cat is alive or dead, leaving us no option until we open the box to view it as being simultaneously alive and dead.

Similarly, a probabilistic approach attempts to manage, for example, many possible confounders while establishing whether montelukast — a drug for asthma — can cause neuropsychiatric effects such as anxiety, suicide, and memory problems. Signal detection alone points to many potential confounders: asthma itself or its treatment with steroids, beta-1 agonists, or antihistamines can each trigger anxiety and depression. The asthmatic population itself has an almost universal background of anxiety and depression [20]. Does the asthma patient in the “box” who takes montelukast and becomes anxious have a montelukast-induced event or not? Of course, once we open the box, we can see if the cat is dead or alive — and, in this case, we can interview the patient to determine the causality.

Medical practice largely deals with a patient’s history — an open box. Signal detection attempts to decide if montelukast has caused a problem without opening the box. There is a difference between an assessment of what has happened and efforts to map possible happenings and decide from that what has happened.

That said, big data — in the form of well-designed epidemiological studies — certainly has a role to play in linking, for instance, birth defects to treatments, a circumstance where the events do happen within a closed box.

Big data also has a role in the evaluation of problems such as substance-induced cancers or cardiovascular problems that only manifest decades after an exposure. Epigenetic studies, as noted above, may help make these studies more precise [1].

Concerns about the possible adverse effects of Covid vaccines shed an ironic light on this issue. In the face of voluminous adverse events reporting — on a scale that has certainly led to the withdrawal of previous vaccines — regulators (and politicians advised by them) have claimed that this time, there is no evidence of causality. They have done so primarily on the basis that these adverse event reports have not been accompanied by statements from *clinicians* that the vaccine seems the most likely cause of the cardiac or neurological event in question.

Google’s AI system, Bard, has recently run into related difficulties. The name “Bard” connotes an element of creativity in storytelling. The algorithms shaping Bard’s storytelling are programmed to avoid mistakes arising from racial bias. As a result, Bard’s “true account” of history can end up with images of non-Aryan Nazi troops. Managing confounders (bias) is important and might help us better understand certain events, but we still need human

judgement calls to distinguish between the confounders that do and do not apply in specific situations.

In this author's opinion, if an AI system is to determine what is happening to a patient present in person, and whether they have a treatment-related problem, it must inevitably place clinical judgement (specific causation) above controlled studies (general causation).

In addition to buttressing the role of clinicians in such situations, an AI system that can learn might also be deployed to enhance clinical practice by, for instance, tracking the paths clinicians take through the minefield of alternative medication options when attempting to reduce the medication burden. No randomised controlled trials (RCTs) or Guidelines will ever support clinicians in this area.

AI compiling and evaluating data from the efforts of clinicians and patients working together to manage highly individual situations, however, may help reveal some of the biases that get in the way of successfully managing or treating specific cases.

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

To cite: Healy D. Drug safety: The roles of big data and clinical experience. *Indian J Med Ethics*. Published online first on February 28, 2025. DOI: 10.20529/IJME.2025.017

Manuscript Editor: Veena Johari

Peer Reviewer: Malini Aisola

Copy editing: This manuscript was copy edited by The Clean Copy.

Copyright and license

©*Indian Journal of Medical Ethics* 2025: Open Access and Distributed under the Creative Commons license (CC BY-NC-ND 4.0), which permits only non-commercial and non-modified sharing in any medium, provided the original author(s) and source are credited.

References

1. Healy D. *Pharmageddon*. Berkeley, California: California University Press; 2012.
2. Lenzer J. Big data's big bias: Bringing noise and conflicts to US drug regulation. *BMJ*. 2017; 358: j3275. <https://doi.org/10.1136/bmj.j3275>
3. Avorn J, Kesselheim A, Sarparwari A. The FDA Amendments Act of 2007: Assessing its effects a decade later. *N Engl J Med*. 2018; 379: 1097–9. <https://doi.org/10.1056/NEJMp1803910>
4. Lenzer J. Legislation introduced to create a new drug safety office at the FDA. *BMJ*. 2005; 330: 1044. <https://doi.org/10.1136/bmj.330.7499.1044-f>
5. FDA Subcommittee on Science and Technology. FDA Science and Mission at Risk. FDA; 2007 [cited 2025 January 11]. Available from: https://www.fda.gov/ohrms/dockets/ac/07/briefing/2007-4329b_02_01_FDA%20Report%20on%20Science%20and%20Technology.pdf
6. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. FDA; 2018 [cited 2025 January 11]. Available from: <https://www.fda.gov/media/120060/download>
7. Healy D. *Shipwreck of the singular: Healthcare's castaways*. Toronto, Canada: Samizdat Press; 2021.
8. Moore TJ, Furberg CD. Electronic health data for postmarket surveillance: a vision not realized. *Drug Saf*. 2015; 358: 601–10. <https://doi.org/10.1007/s40264-015-0305-9>
9. Madigan D, Ryan PB, Schuemie M, Stang PE, Overhage M, Hartzema A, Suchard M, DuMouchel W, Berlin JA. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol*. 2013; 178: 645–51. <https://doi.org/10.1093/aje/kwt010>
10. Laporte JR. Fifty years of pharmacovigilance: Medicines safety and public health. *Pharmacoevid Drug Saf*. 2016; 25: 725–32. <https://doi.org/10.1002/pds.3967>
11. Healy D. Antidepressants and sexual dysfunction: A history. *J Roy Soc Med*. 2020; 113: 133–5. <https://doi.org/10.1177/0141076819899299>
12. Bahrick AS. Post SSRI sexual dysfunction. *American Society for the Advancement of Pharmacotherapy. Tablet*. 2006; 7: 2–11.
13. Mason S. Letter to Senator C Grassley from S Mason, FDA assistant commissioner. Dec 2008 [Available with author].
14. Food and Drug Administration. Guidance for industry: Good pharmacovigilance practices. Rockville, Md. US Dept of Health and Human Services: March 2005 [cited 2025 January 11]. Available from: <https://www.regulations.gov/docket/FDA-2004-D-0041>
15. Healy D, Le Noury J, Mangin D et al. Citizen petition: Sexual side effects of SSRIs and SNRIs. *Int J Risk Saf Med*. 2018; 29(3-4): 135–147. <https://doi.org/10.3233/JRS-180745>
16. Carter C, Blizard R. Autism genes are selectively targeted by environmental pollutants including pesticides, heavy metals, bisphenol A, phthalates and many others in food, cosmetics or household products. *Neurochem Int*. 2016; 101: 83–109. <https://doi.org/10.1016/j.neuint.2016.10.011>
17. Smith MT, Guyton K, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, DeMarini DM, Caldwell JC, Kavlock RJ, Lambert PF, Hecht SS, Bucher JR, Stewart BW, Baan RA, Coglianò VJ, Straif K. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. *Environ Health Perspect*. 2016; 124: 713–21. <https://doi.org/10.1289/ehp.1509912>
18. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: What is realistic, What are illusions? Part 1: Ways to make an impact and why we are not there yet. *Drug Discov Today*. 2021; 26: 511–24. <https://doi.org/10.1016/j.drudis.2020.12.009>
19. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 2: A discussion of chemical and biological data. *Drug Discov Today*. 2021; 26: 1040–52. <https://doi.org/10.1016/j.drudis.2020.11.037>
20. Goodwin R, Jacobi F, Thefeld W. Mental Disorders and asthma in the community. *Arch Gen Psychiatry*. 2003; 60(11): 1125–1130. <https://doi.org/10.1001/archpsyc.60.11.1125>