Randomized Controlled Assays and Randomized Controlled Trials: A Category Error With Consequences

David Healy, MD FRCPsych

Chief Scientific Officer, Data Based Medicine, Bangor, UK

In 1962, in the wake of the thalidomide crisis, a new Amendment to the Food and Drugs Act introduced Randomized Controlled Trials (RCTs) into the regulations governing the licensing of medicines. It was believed that requiring companies to demonstrate their products were effective through RCTs would contribute to safety. In 1962, RCTs were a little-understood technique. It was thought trials would produce generalizable knowledge with similar outcomes for successive trials. As a result, regulators adopted a criterion of two positive placebo-controlled trials for licensing medicine. For physicians keen to stall therapeutic bandwagons and eliminate ineffective treatments, a negative RCT result was a good outcome. When it made a gateway to the market, companies, in contrast, had an interest to transform RCTs from assessments that might throw up unexpected or negative results into Randomized Controlled Assays (RCAs) that efficiently generated approvable results. This article outlines the differences between RCTs and RCAs, the steps companies took to transform RCTs into RCAs, and the consequences of this transformation.

Keywords: pharmaceutical industry; psychiatric drug research; biological psychiatry; informed choice; ethics

There have been recent controversies about the approval of a range of drugs such as Aduhelm for Alzheimer's dementia (Woloshin & Kesselheim, 2022), Rivaroxaban for anticoagulation (Demasi, 2022a), and the antibiotic Ricarbrio (Doshi, 2023; Ross, 2023). Regulators like the Food and Drugs Administration (FDA) are criticized for approving minimally effective and potentially hazardous agents, for not adequately overseeing trials (Demasi, 2022b), and for abandoning science (Doshi, 2023; Ross 2023).

Some of these criticisms arguably hinge on a misunderstanding of the nature of the studies pharmaceutical companies undertake to get their drugs approved and the role of regulators in these approvals. Rather than make scientific evaluations, when approving products from food to automobiles, regulators review whether the product meets an assay standard. Viewing pharmaceutical company submissions for the approval of a new drug as assays rather than scientific studies may resolve some misunderstandings. The regulation of medicines, however, is further complicated by the fact that the assays companies submit for drug approval involve randomization, making them easily confused with Randomized Controlled Trials (RCTs) and pharmaceutical companies exploit this confusion for marketing purposes.

http://dx.doi.org/10.1891/EHPP-2023-0006

The origin of this confusion lies in the 1962 Amendments to the Food and Drugs Act, which incorporated RCTs into the regulatory apparatus for new drugs, aimed at demonstrating treatment effectiveness. The logic of regulation, however, is at odds with the logic of science, and ultimately, this has led company studies submitted as part of a licensing application for a new drug to transition from RCTs to Randomized Controlled Assays (RCAs).

The transition from RCTs to RCAs took place in the 1980s and is most readily visible in the efforts of regulators to establish the boundaries of their role in the licensing of antidepressant drugs.

As of 1962, company studies done for licensing purposes were called RCTs. The transition of company studies to RCAs in the 1980s make this designation inappropriate. It now constitutes a category error, in that RCAs, while bearing some superficial resemblances to RCTs, embody quite different processes. Designating them as RCTs leads to an inappropriate inclusion of RCAs in guidelines based on RCT evidence. This suits company marketing purposes but is at odds with the science of therapeutics.

Distinguishing between assays and trials that have many superficial similarities is not easily done. The role of regulation complicates the picture further. Across regulatory categories, assays have a defined testbed and standard to be met. Demonstrating that a standard has been met can range from ascertaining pH levels in a fluid, the number of particulates in automobile exhaust fumes, or demonstrating that an antidepressant has an effect on something like a Hamilton Rating Scale for Depression. In the case of medicines, meeting a standard suffices for licensing purposes, but it is a mistake to assume that the performance of a drug in an assay will inform therapeutics.

Further confusion stems from the widespread understanding that the regulation of medicines is designed to keep us safe. Efforts are also made to ensure the practice of medicine keeps us safe. These combined efforts lead to confusion, namely that the regulation of medicines should shape medical practice—so that doctors for example should not prescribe off-label. The label, however, is intended to constrain the claims companies can make rather than medical practice.

This distinction was captured in an FDA hearing on Respiratory Syncytial Virus (RSV) vaccines in pregnancy (May 18, 2023) when FDA was challenged about a difficult problem, namely that giving RSV vaccines in pregnancy looks likely to reduce the efficacy of influenza and Tdap vaccines also given in pregnancy (El-Sahly & Toerner, 2023). The chair of the meeting asked if FDA could offer any recommendations on what to prioritize. FDA responded:

> It's a great question. We look forward to your discussion about this topic…All the data that is part of our review ends up in product labeling…It would be up to the provider to understand what's in labeling, to understand the data on the coadministration, and come to a determination about vaccine administration….Our deliverable to you is important information in the labeling so that a provider [a doctor] can decide [as regards the] use of the product.

RANDOMIZED CONTROLLED TRIALS

Before Thalidomide

The first RCT in 1948, designed and run by Austin Bradford Hill of Britain's Medical Research Council (MRC), compared streptomycin with treatment as usual for tuberculosis (Medical Research Council, 1948). It told us less about streptomycin than prior Mayo Clinic trials controlled without randomization, which made it clear that there was rapid tolerance to streptomycin and some people went deaf (Hinsaw & Feldman, 1944).

The MRC trial did not take inspiration from Ronald Fisher's thought experiment, in which he posited randomization as a means of controlling unknown confounders. Fisher was attempting to mathematize expert knowledge—not conduct a trial. He posited that if an expert knew what he was doing, and randomization 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 (Fisher, 1935).

In the later MRC trial, randomization was simply used as a means of fair allocation (Chalmers et al., 2012; Hill, 1984). There was no assumption that doctors knew what they were doing, and no reason to think randomization could control ignorance. The motivation for controlled trials was to eliminate ineffective treatments, as before streptomycin, many ineffective cures for tuberculosis had been touted.

Clinicians did not rush to adopt RCTs after Hill's trial. The leading advocate of RCTs in the United States was Louis Lasagna (Healy, 2021; Lasagna, 1998). Before 1962, Lasagna had run a placebo-controlled RCT of thalidomide, showing it to be effective and safe (Lasagna, 1960).

Clinical Trials and Regulation

In 1962, the thalidomide crisis in the United States led to amendments to the 1938 Food, Drugs, and Cosmetics Act. Prior drug regulation sought to ensure medicines were safe. Arguing ineffective drugs cannot be safe, Lasagna persuaded politicians to introduce a requirement for companies to demonstrate the effectiveness of their medicines in the amendments.

The amendment-specified companies should provide substantial evidence of effectiveness, which was codified to mean evidence from two RCTs. Companies turned to academic physicians to run these trials, which typically compared a new and an established drug. It was only in the 1980s when it was accepted that noninferiority to an existing drug did not provide robust evidence of effectiveness that placebo controls were introduced (Leber, 1996).

While a placebo requirement removed one avenue for getting minimally effective drugs on the market, it introduced another. With a later more explicit focus on assays, as outlined below, the requirement to beat a placebo opened the door to a licensing of drugs based on studies showing a minimal change on a surrogate marker with debatable clinical relevance and in some cases with more lives lost on active treatment than on placebo.

In the 1960s, the place of RCTs in clinical practice remained uncertain. Austin Bradford Hill offered this view in 1965:

> Frequently with a 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 .

During the 1960s, the FDA, in contrast, appears to have held the belief that RCTs had ushered in an era of regulation by science. This belief was challenged by controversies surrounding Panalba, a combination antibiotic, and by FDA caution in licensing beta-blockers for blood pressure (Healy, 2021). Lasagna and Wardell coined the term drug lag pointing to evidence that FDA was slower than European regulators to license beta-blockers and other drugs (Lasagna & Wardell, 1975; Wardell & Lasagna, 1975).

The FDA argued that evidence of lowered blood pressure was not evidence of effectiveness, but it capitulated. Lasagna, in turn, later recognized that studies of beta-blockers on blood pressure in clinical practice did little to benefit patients and often did harm (Jachuk et al., 1982; Lasagna, 1998). By the early 1990s, several well-powered clinical trials had made it clear that the drop in blood pressure beta-blockers produced did not necessarily improve life expectancy—at least not as much as other antihypertensives (Allhat, 2002, Dahlöf et al., 2002; Epstein, 2017).

It was also becoming clear that a drug class such as antihypertensives might contain quite different therapeutic principles, from thiazides, through beta-blockers, to Angiotensin Converting Enzyme (ACE) inhibitors. Clinical practice was not a matter of knowing, which was generically the best antihypertensive but which of these drugs was likely to be most effective for a specific patient. The same is true for most drug groups from laxatives to antidepressants. These different modes of action offer different therapeutic principles.

This is true of hypoglycemic agents also. A National Institute of Health (NIH) trial aimed at establishing the respective places of tolbutamide, metformin, diet, and an appetite suppressant had a measure of effectiveness - mortality - as its outcome measure (Blackburn & Jacobs, 2017). This trial took 10 years from conception to early results, and despite the largest-ever patient cohort, it led to controversy rather than consensus (Healy, 2012).

A 10-year timescale is incompatible with a pharmaceutical company's need to get products on the market expeditiously. An endpoint like mortality almost inevitably entails a longer timeframe than that of a study that looks at an effect on a surrogate outcome, whether rating scales like the Hamilton Depression Rating Scale, blood lipids, or peak flow rates, measures when a treatment can be expected to show an effect in a matter of days or weeks.

In the 1970s, in the interests of the expedition, pharmaceutical companies stopped sponsoring academic clinicians to run their license application studies. Running these studies was contracted out to Contract Research Organizations. The writing of articles representing the results of trials was contracted out to Medical Writing Companies (Healy, 2021).

From the early 1970s, bodies like the NIH and MRC allocated less funding for trials given that pharmaceutical companies had to run trials for licensing purposes, and there was a belief that the process was sufficiently robust that broadly speaking there would be similar outcomes regardless of who funded the study.

With companies free to adopt protocols prepared a decade before by academic clinicians, and designing them for running assays rather than trials, the way was open to optimizing the studies companies undertook as RCAs.

RANDOMIZED CONTROLLED ASSAYS

The U.S. FDA regulates food and drugs. In the case of chocolate or butter, the FDA checks if a submitted substance meets an assay standard. If the product has a certain amount of cocoa solids, regulators allow companies to use the word chocolate. A certain amount of specific animal fats distinguishes butter from lard colored to look like butter. The regulator does not decide if butter is good for us or whether this is a good chocolate - but rather is decided by others. If the products, including medicines, meet the criteria, regulators allow words like butter, chocolate, and antidepressant to be used on product labels.

In the 1980s, Paul Leber, then head of the Central Nervous System Division within the FDA, introduced the concept that companies undertake assays rather than trials (Leber, 1998). This approach was adopted and regulators now review RCAs to see if the results meet an assay standard. Assays and clinical trials overlap in their use of randomization and surrogates, but can be distinguished.

Purpose

The leading factor that drives a distinction between assays and trials is the purpose of the exercise. Meeting an assay standard is designed to get a drug on the market. Assay protocols are standard but a clinical trial is designed to inform clinical practice. It is more likely to have a novel protocol and rather than expand the use of a medicine, the results of an RCT may restrict the use of treatments to subgroups who will benefit.

Internal Versus External Validity

A demonstration designed to meet a standard requires internal validity. Company assays, therefore, largely recruit uncomplicated samples not typical of clinical cases.

Clinical trials aim at informing clinical practice, for which external generalizability is important. The trials recently run by Britain's National Health Service to find which drug treatments worked for patients admitted to hospitals with COVID-19 is a good instance of this (Ahmad et al., 2022). Such trials embrace pragmatism whereas assays don't.

Assay Standard

In the case of assays for new products, companies will often request regulators to set a standard that will enable them to run an assay. Meeting the standard comes with an implicit agreement that the regulator will then let the product on the market.

This was the case with tacrine, the first cholinomimetic agent to be licensed for dementia (Leber, 1998). A standard was set based on a minimal change in the activities of daily living scale without establishing whether this was meaningful. Tacrine had no

evidence of enhanced life expectancy or slowing of disease progression. The tacrine standard set a precedent for subsequent drugs for dementia with many approved but none to date offering a significant benefit in clinical practice.

More recently, inclisiran, aimed at reducing levels of low-density lipoproteins, was licensed based on showing a fall in these lipoproteins without evidence of enhanced life expectancy (Ray et al., 2020). For a novel agent like inclisiran, companies will confirm with the regulator beforehand that if they run a study that meets a standard, such as the fall in low-density lipoproteins, the regulator will license the product on this basis.

Failed Assays

Regulators routinely designate negative studies in a license application as failed assays and pay no heed to such studies. In contrast, a negative RCT has important therapeutic implications.

The history of failed assays begins with Selective Serotonin Reuptake Inhibiting (SSRI) antidepressants (Healy, 2004). Put into licensing application assays with moderate to severely depressed patients, SSRIs have not beaten a placebo, while tricyclic antidepressants (TCAs) did. In assays of mild to moderately depressed people, SSRIs could demonstrate a benefit but often failed to beat the placebo, but if the SSRI failed to beat the placebo, a comparator TCA often failed also. When a drug that regulators view as having established efficacy (a TCA) failed to work in an assay, as in this case, the problem becomes one of assay sensitivity. The is viewed as being at fault—a high placebo response rate inhibits the demonstration of a supposedly known effect.

In the license application for sertraline in 1990, a majority of assays were negative (Turner et al., 2008). Senior FDA officials made it clear that regulations require two positive assays, stating that they could potentially license a drug if 2 out of 100 assays were positive (Leber, 1990). This illustrates the role of assays in regulation.

Regulation and the Medical Literature

In 2002, faced with a license application for paroxetine for pediatric depression containing three assays of paroxetine, all negative (329, 377, and 701) FDA agreed to approve paroxetine and wrote to GlaxoSmithKline (GSK) as follows (Katz, 2002):

> Given the fact that negative trials are frequently seen, even for antidepressant drugs that we know are effective, we agree that it would not be useful to describe these negative trials in labeling.

Study 329 had been published a year previously claiming paroxetine worked well and was safe (Keller et al., 2001). This published claim of a benefit was not born out by internal GSK documents, by an FDA review, or by a subsequent independent analysis of the data (Le Noury et al., 2015).

Erick Turner and colleagues noted that of adult studies done as part of a licensing application for SSRIs and related antidepressants, 31% of trials viewed by the FDA as negative were published as positive (Turner et al., 2008). Thomas Laughren, a senior FDA figure, also made it clear that 46% of assays submitted as part of a licensing application for antidepressants were negative (Laughren, 2001).

After licensing, regulators monitor the wording of adverts, but they do not view it as their job to police the medical literature any more than they would police the medical or academic literature on the benefits or drawbacks of butter or chocolate. They are not scientists who contest claims in the published literature.

Assay Preparation

Company studies of drugs in clinical populations are commonly called drug trials (assays), but giving a drug to healthy volunteers is more appropriately called a drug trial (assay), with clinical trials (assays) better termed as treatment trials (assays). Treatment assays are always preceded by phase 1 assays, whereas RCTs are not. In a healthy volunteer phase 1 assay, the confounders that stem from treating a medical condition, which is not reliably controlled by randomization, are eliminated and the effects of the drug are more salient.

These phase 1 studies have a standard assay format. They seek to establish what effect an agent might have, for instance, on required standards such as clotting times or the levels of other commonly used drugs. While conducting these, companies can also monitor an agent's other effects. In SSRI phase 1 studies, participants had sexual effects, agitation, suicidality, and dependence, some of which did not remit when treatment stopped (Healy, 2020). Many of these assays remain unpublished and of those that are published, details like this are not reported.

Informed by phase 1 assays, the designs of treatment assays can minimize the chance of detecting unwanted effects, eliminating, for instance, any instruments that might detect suicidal or sexual effects. There are, to this day, no company studies designed to explore either suicidality or sexual dysfunction induced by antidepressants. Avoiding such effects can make an assay more efficient for regulatory purposes, but it poses problems for clinical practice.

Assay Endpoints

Whether done to obtain a license or to inform clinical practice, assays and trials have a primary endpoint. A primary endpoint centers on one effect, among more than a hundred, small molecule drugs have. The word primary, however, suggests this is the most common effect. In addition, for two decades, regulators have claimed drugs have been licensed based on a positive benefit–risk ratio, a statement that implies the primary outcome is the most common of a drug's effects.

However, SSRI antidepressants have a minimal effect on depression rating scales, often barely distinguishable from placebo in a 6-week assay. The focus needed to detect the desired commercial effect meant that these assays missed a genital numbing that happens (close to universally) among participants within 30 minutes of the first pill. This effect had been known for over a decade before the licensing of SSRIs (Beaumont, 1996), and it underpins the later widespread use of SSRIs and licensing of dapoxetine for premature ejaculation.

An intense focus on one effect can make common effects of greater importance to the patient disappear. This is a hazard shared by both assays and trials, but it would be expected that clinicians knowing of such effects, undertaking an RCT to establish a beneficial effect, would include measures to take these other effects into account. In an RCT, a possible but marginal benefit would not be pitched as a favorable benefit–risk

ratio when a problem of considerable significance to patients occurs more commonly than the primary endpoint.

In addition, the primary endpoint of an assay is often a matter of bureaucratic convenience. A 2–3 point lowering of an aggregate depression rating scale score is the antidepressant assay standard today, but this is of uncertain clinical significance. In a meta-analysis of all antidepressant assays combined, this standard was met, while the data from these assays otherwise pointed to a doubling of the suicidal event rate, with more people dying on active treatment than on the placebo (Stone et al., 2009).

Phase 1 SSRI assays also pointed to a physical dependence on these agents, which later treatment assays side-stepped. We now have a public health problem with 15% of the population of many Western countries on these medicines for years, primarily because they are unable to get off them, even though when introduced the clinical understanding was these drugs would only be used for 3–6 months (Healy, 2004).

These are very real problems, but it is not clear that regulators should be held responsible for these problems.

Assays and Adverse Effects

RCTs of this type begun by Bradford Hill explicitly embraced clinical judgment, in part because that was the norm in prior controlled clinical trials, and in part because exploring the possible clinical niche for a new treatment with unknown features, almost by definition, cannot be done by an algorithm or with a standardized protocol.

In practice, investigators in company assays are required to adhere to the protocol with all questions delivered according to designated checklists—such as the Hamilton Depression Rating Scale. The Hamilton depression Rating Scale has items for sleep, sex, anxiety, agitation, and suicide, all of which may be affected by the condition or treatment. The rater is asked to rate suicidality from 0–4, depending on its absence or a recently attempted suicide but is not asked to judge whether the treatment or the condition has caused the problem. If the suicidality rating increases, there is a default to the condition.

Companies have developed a rationale for this, namely that it is only after an assay is over, and we have statistically significant results for X that it can be claimed that a drug does X. As no assay is designed to investigate effects other than on the primary endpoint, adverse effects are rarely likely to be statistically significant, and if written up, it will at best be noted "as reported" rather than likely linked to treatment (Hudson, 2000; Puliyel & Naik, 2018).

An investigator's hunch that an event is linked to treatment is viewed as anecdotal in the absence of statistically significant evidence that the drug can cause events like this. There is no scope in an assay to adjust the dose, to dechallenge and rechallenge the patient or introduce an antidote, or to investigate a link between treatment and an effect.

This is not inconsistent with Bradford Hill's view that trials can help in establishing one important effect of a drug, such as whether there is a therapeutic effect, but it leaves us without a method to establish the adverse effects of the treatment.

In addition to a primary focus on therapeutic effect, in assays, adverse effects are coded and grouped. Companies appear to have availed of opportunities these processes offer to a greater extent than has been noted in RCTs. In Study 329, children attempting suicide were coded as emotionally labile (Le Noury et al., 2015). Behavioral events

such as suicidality were grouped in a neurological cluster along with common adverse effects like headaches and dizziness. This drowned out a strong signal for suicidality on paroxetine that became visible when these behavioral effects were placed in a psychiatric higher-order grouping (Le Noury et al., 2015).

The coding and grouping of adverse effects are the first act of authorship in a study. Peer reviews of company studies never explore the replicability of coding or grouping. This problem could be overcome if the data from the studies were available, but assay data is not ordinarily available in contrast to the norms of science which put a premium on access to the data.

Surrogate Outcomes

Company assays commonly adopt surrogate endpoints, such as blood pressure, blood sugar levels, bone densities, peak flow rates, or rating scale scores rather than improved function or lives saved. Surrogates facilitate speedy assays and efficient regulation.

The relationship between effects on surrogates and benefits such as lives saved, or function restored, is uncertain. In antidepressant assays, for instance, there are more suicide attempts and completed suicides in those on active treatment compared with placebo, even while changes in surrogate depression rating scale scores suggest these drugs "work."

Adjusting risk surrogates such as lipid levels, blood pressure, or rating scale scores, offer better templates for assays than saving lives from heart attacks, strokes, cancers, or psychoses. If these assays are mistakenly viewed as trials, the management of risks becomes a more certain way for companies to make money rather than treating illnesses. Reviewing this scenario, Goldman Sacks recently declared saving lives is not a good business model (Kim, 2018).

Significance Testing

In the 1980s, the leading figures in medical statistics argued that RCTs and epidemiological studies in medicine did not test hypotheses but rather offered estimations of treatment effects undertaken against a background of considerable uncertainty. Therefore, statistical significance testing should be replaced with Confidence Intervals (CIs) (Carver, 1978; Cohen, 1994; Greenland, 2006; Gardner & Altman, 1986; Guttman, 1985; Johnstone et al., 1986; Oakes, 1986; Poole, 1998; Rothman & Greenland, 1998; Rozeboom, 1997; Salsburg, 1986; Sterne & Davey Smith, 2001).

The desirability of using CIs rather than significance testing rapidly became the orthodox point of view. But, when it comes to reporting assay results to regulators or in journals, CIs default into significance testing. Healthcare academics substitute a 95% CI for study data instead of significance testing (Healy, 2006).

Statistical significance testing is more appropriate for RCAs than for RCTs, which call for a demonstration of a known effect. The default into treating CIs as significance tests seems likely driven in part by the need for regulators and companies to have a Stop–Go signal, which statistical significance provides but CIs do not, unless adapted for that purpose. This default is particularly problematic in the case of the adverse effects of a medicine, when the assays bringing it to market, they have not been designed to measure with precision effects other than the primary endpoint, leaving companies to claim that their drug has no adverse effects, as none are statistically significant.

The continuing de facto use of a surrogate form of statistical significance testing sitting beside universal condemnation of this use by medical statisticians points to a profound incoherence that has a possible explanation in a failure to distinguish between RCAs and RCTs.

CONSEQUENCES OF A CATEGORY ERROR

There are several consequences of a failure to distinguish between assays and trials.

Effectiveness and Safety

The 1962 FDA statute required companies to demonstrate their treatments are effective, which to many likely means saving lives or restoring function. The statute does not say to demonstrate an effect in an assay. Lives have rarely been saved in the assays used to bring medicine on the market for the last four decades. The life-saving effects of Triple Therapy for AIDS, for instance, were discovered in later clinical practice rather than through company assays.

In 1962, it was a reasonable expectation that ensuring effectiveness would contribute to safety. In practice, assumptions of established effectiveness lessen the prior premium the 1938 Food, Drugs, and Cosmetics Act put on assessing safety, particularly if a therapeutic effect is *ipso facto* viewed as establishing a favorable benefit–risk ratio. The licensing of a drug is not held up now, as it once was under the 1938 Act, on the basis that even though effective, a sleeping pill, thalidomide, should not cause peripheral neuropathy.

RCTs emerged in the early 1950s, fostered by a need to establish if any of the many treatments on the market, which companies claimed worked, actually did work. Many did not. By the 1960s, we had an expanding arsenal of treatments that worked very well, all brought to market without the use of RCTs. Many of these were better than the remedies that succeeded them.

In 1965, Bradford Hill suggested RCTs were now needed to evaluate which of the available drugs were best in class (Hill, 1966). This is not something an RCT can establish definitively, but efforts to compare treatments can inform clinical practice albeit in studies that take a long time.

One of the primary medical needs today is to reduce medication burdens, stimulated by having too many drugs that are individually effective, but which compromise effectiveness when combined. It is unlikely that RCTs can make much of a contribution to evaluating this medical need.

We need companies to have an efficient way to bring a drug to the market. Assay systems that demonstrate an anxiolytic or sedative effect, or an effect on a blood parameter or the generation of antibodies, could be even more efficient than running treatment assays notionally aimed at establishing effectiveness in a condition. Some assays could be run on healthy volunteers.

Making it clear that RCAs are used to license rather than evaluate drugs would make salient the need for a later evaluation of a drug's place in therapy. This could be done by replacing phase 4 company assays with RCTs run by independent healthcare bodies designed to inform therapeutics and establish the hazards of treatment rather than to secure a marketing niche.

As things stand, participants in company assays generate a state of legal jeopardy for themselves and others in that companies will claim that their trials did not show their drugs can cause injuries like this as the results did not reach statistical significance (Healy, 1999, 2023).

In 1983, Lasagna, who had introduced RCTs into the 1962 regulations, countered an emerging view that RCTs were needed to establish if a drug had an adverse effect, arguing that case reports by clinicians were more rigorous and appropriate (Lasagna, 1983).

The company studies Lasagna mandated in 1962 now form the core of Evidence-Based Medicine. In 1998, he stated these studies "invariably fail to tell the physician what he or she wants to know which is which drug is best for Mr. Jones or Ms. Smith—not what happens to a nonexistent average person" (Lasagna, 1998).

Therapeutic Principles or Magic Bullets

Getting a drug licensed as an antihypertensive, hypoglycemic, laxative, or antidepressant implies that these drugs are magic bullets rather than therapeutic principles.

In the case of laxatives, there are four therapeutic principles—bulk forming, stool softening, and osmotic or stimulant agents. The clinical skill is to pick out the therapeutic principle most likely to help a specific patient. Simply prescribing any laxative because a company has been licensed to use this word risks creating treatment-resistant constipation.

Among antidepressants, SSRIs have a serene effect. Noradrenaline reuptake inhibitors enhance drive and vigilance, mirtazapine and trimipramine stimulate appetite and help sleep, while TCAs combine some of these principles and add a degree of euphoria through their anticholinergic effects. Unlike SSRIs, TCAs effectively treat melancholia (Bech, 1998).

If a doctor prescribes an SSRI to a patient simply because it is an antidepressant, instead of attempting to decide whether a serene effect is likely to be more appropriate for this patient because they are anxious, a drive-enhancing effect because they are fatigued, or a TCA because they have melancholia, she puts the patient on track to a treatment-resistant depression.

As noted above, the treatment of hypertension and type II diabetes are similarly beset by failures to select the appropriate therapeutic principle. This likely holds true in most areas of medicine.

Paradoxically if companies had distinguished between these therapeutic principles and applied them to license the serenic effects of SSRIs, for instance, rather than supposed effectiveness in depression, they would have been able to run more efficient assays with much more convincing results. Opening the black box of drug labels would likely contribute to therapeutic effectiveness and be safer. It would, however, hand over more control of the drug a patient gets to their clinician than happens currently, which companies might resist.

The Evidence in Evidence-Based Medicine

For journal editors, the appearances of randomization in controlled studies reported in articles with distinguished authors, and FDA approval of the drug offering no objections to claimed benefits and freedom from harms, opens the door to the publication of

company assays in distinguished medical journals where they are read as clinical trials rather than assays. This is a significant source of confusion.

The nature of data is central to any distinction between trials and assays. Demonstrations of a drug meeting a standard effect in assays do not need to be accompanied by the names of those on whom the drug was tested. Companies claim they withhold names to preserve confidentiality, but they withhold these details even in the case of healthy volunteer assays.

From an evidential point of view in legal settings, figures for assays aimed at generating standard ranges for blood pressure, hemoglobin, or other parameters can be admitted as technical materials without access to the names of the people used to establish that range but technical materials are not clinical evidence.

Data is often incorrectly taken to mean figures. In a scientific study of a drug rather than an essay, people are the data (Blumsohn, 2006; Dyer, 2010). We do not know what happened on a drug in a clinical trial or clinical practice (Lasagna, 1983) without people's names and the ability to contact them or their relatives and access their medical records. The lack of access to the people in company assays means that on this basis also, company assays do not meet the legal standard for medical as opposed to technical evidence.

In admitting claims based on RCAs into legal proceedings, courts are admitting hearsay, which is normally excluded legally. At this point, courts become party to the creation of a state of legal jeopardy in that someone injured by treatment cannot bring people previously similarly injured in an assay into court to be examined and cross-examined. Some courts will have to address this. In contrast, doctors and patients who are the authors or subjects of published reports outlining an injury can be brought into court (Healy, 2023).

This point about evidence equally applies to treatment guidelines based primarily on company assays.

CONCLUDING REMARKS

In the 1950s, many drugs did not work, and our need was to rigorously test claims for effectiveness. Testing safety claims is the more important need at this time.

There is a degree of urgency to address the arrangements under which medicines are brought to market in that we are moving into an era of virtual assays with prepopulated lists of adverse effects on digital devices and assay reports written by artificial intelligence (AI) within weeks of completing data collection. AI will make assay processes even more efficient and are certain to be adopted.

It is not in anyone's interest to make it harder for companies to get new medicines that are likely to work on the market. Simply distinguishing between RCAs and RCTs may neutralize claims that companies' conflicting interests are corrupting science and do a good deal to adjust perceptions of our current problems.

Life expectancies have been falling before COVID-19, and increasing medication burdens are a plausible factor in this. A move to reduce medication burdens (Garfinkel & Mangin, 2010; Healy & Mangin, 2019) will need judgment calls about which medicines to stop. This may restore salience to the question

of whether it is right to equate evidence from RCAs and RCTs and may also restore a place for clinical judgment in clinical practice and research.

REFERENCES

Ahmad, R., Gordon, A. C., Aylin, P., Redhead, J., Holmes, A., & Evans, D. P. (2022). Effective knowledge mobilisation: Creating environments for quick generation, dissemination, and use of evidence. *BMJ (Clinical Research Ed.)*, *379*, e070195. https://doi.org/10.1136/bmj-2022-070195

ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. (2002). Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic. *Journal of the American Medical Association*, *288*, 2981–2997. https://doi.org/10.1001/jama.288.23.2981

Beaumont, G. (1996). The place of clomipramine in psychopharmacology. In D. Healy (Ed.), The Psychopharmacologists, volume 1 (pp. 309–328). London: Chapman & Hall.

Bech, P. (1998). Measurement and organization in Psychopharmacology. In D. Healy (Ed.), The Psychopharmacologists vol 2 (Vol. 2, pp. 499–520). Arnold London.

Blackburn, H., & Jacobs, D. R. (2017). The university group diabetes program 1961-1978, pioneering randomized controlled trial. *International Journal of Epidemiology*, *46*(5), 1354–1364. https://doi.org/10.1093/ije/dyw168

Blumsohn, A. (2006). Authorship, Ghost-science, access to data and control of the pharmaceutical scientific literature – who stands behind the word? *American Association for the Advancement of Science, Professional Ethics Reports*, 19. ww.aaas.org/ssp/sfr1/per/per46.pdf

Carver, RP. (1978). The case against statistical testing. *Harvard Educational Review*, *48*(3), 378–399. https://doi.org/10.17763/haer.48.3.t490261645281841

Chalmers, I., Dukan, E., Podolsky, S., & Davey Smith, G. (2012). The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries. *Journal of the Royal Society of Medicine*, *105*(5), 221–227. https://doi.org/10.1258/jrsm.2012.12k029

Cohen, J. (1994). The earth is round (P<.05). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Dahlöf, B., Devereux, R. B., Kjeldsen, S. E., Julius, S., Beevers, G., de Faire, U., Fyhrquist, F., Ibsen, H., Kristiansson, K., Lederballe-Pedersen, O., Lindholm, L. H., Nieminen, M. S., Omvik, P., Oparil, S., Wedel, H., & LIFE Study Group. (2002). Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study (LIFE): Lancet. *Lancet (London, England)*, *359*(9311), 995–1003. https://doi.org/10.1016/S0140-6736(02)08089-3

Demasi, M. (2022a). Rivaroxaban: Lancet opens investigation into anticlotting drug trial after BMJ report. *BMJ (Clinical Research Ed.)*, *379*, 3071. https://doi.org/10.1136/bmj.o3071

Demasi, M. (2022b). FDA oversight of clinical trials is "grossly inadequate," say experts. *BMJ (Clinical Research Ed.)*, *379*, 2628. https://doi.org/10.1136/bmj.o2628

Doshi, P. (2023). Did FDA break its own rules in approving the antibiotic recarbrio. *BMJ*, *381*, p1048. https://doi.org/10.1136/bmj.p1048

Dyer, C. (2010). Aubrey blumsohn: Academic who took on industry. *BMJ (Clinical Research Ed.)*, 340, 22–23. https://doi.org/10.1136/bmj.c1096

El-Sahly, H., & Toerner, J. (2023). *Pfizer's Maternal RSV Vaccine. Food and Drug Administration, Vaccine and Related Biological Products Meeting, May 18, 2023.*

Epstein, D. (2017). *When evidence says no but doctors say yes* (Vol. 22). The Atlantic.

Fisher, R. A.. (1935). *The design of experiments*. Macmillan.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than P-values: Estimation rather than hypothesis testing. *British Medical Journal (Clinical Research Ed.)*, *292*(6522), 746–750. https://doi.org/10.1136/bmj.292.6522.746

Garfinkel, D., & Mangin, D. (2010). Feasibility study of a systematic approach for discontinuation of multiple medications in older adults: addressing polypharmacy. *Archives of Internal Medicine*, *170*(18), 1648–1654. https://doi.org/10.1001/archinternmed.2010.355

Greenland, S. (2006). On bayesian perspectives for epidemiologic research (reply). *International Journal of Epidemiology*, *35*(3), 777–778. https://doi.org/10.1093/ije/dyl081

Guttman, L. (1985). The Illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis*, *1*(1), 3–9. https://doi.org/10.1002/asm.3150010103

Healy, D. (1999). Clinical trials and legal jeopardy. *Bulletin of Medical Ethics*, *153*, 13–18. https://doi.org/2016-01-08T21:44:45Z

Healy, D. (2004). *Let them eat prozac*. New York University Press.

Healy, D. (2006). The antidepressant tale: Figures signifying nothing. *Advances in Psychiatric Treatment*, *12*(5), 320–327. https://doi.org/10.1192/apt.12.5.320

Healy, D. (2012). *Pharmageddon*. University of California Press. https://doi.org/10.1525/9780520951815

Healy, D. (2020). Antidepressants and sexual dysfunction, a history. *Journal of the Royal Society of Medicine*, *113*(4), 133–135. https://doi.org/10.1177/0141076819899299

Healy, D. (2021). Shipwreck of the Singular. In *Healthcare's castaways*. Toronto, Samizdat Press.

Healy, D. (2023). Diagnosis, verdict, conclusion, and causality. *Ethical Human Psychology and Psychiatry*, *25*(1), 29–37. https://doi.org/10.1891/EHPP-2023-0001

Healy, D., & Mangin, D. (2019). Clinical judgments, not algorithms, are key to patient safety. *BMJ (Clinical Research Ed.)*, *367*, l5777. https://doi.org/10.1136/bmj.l5777

Hill, A. B.. (1966). Reflections on the controlled trial. *Annals of the Rheumatic Diseases*, *25*(2), 107–113. https://doi.org/10.1136/ard.25.2.107

Hill, A. B.. (1984). *Letter to Louis Lasagna*.

Hinsaw, H. C., & Feldman, W. H. (1944). Evaluation of chemotherapeutic agents in clinical tuberculosis. *American Review of Tuberculosis*, 50, 202–203. https://doi.org/10.1016/S0025-7125(16)35924-7

Hudson, I. (2000). *Deposition of Ian Hudson in Tobin et al, plaintiff v SmithKline Beecham, defendant In the United States District Court for the Eastern District of Pennsylvania, No: 00CV0025, December 15, 2000*.

Jachuk, S. J., Brierley, H., Jachuk, S., & Willcox, P. M. (1982). The effect of hypotensive drugs on the quality of life. *J Royal College General Practitioners*, 32, 103–105.

Johnstone, D. J., Barnard, G. A., & Lindley, D. V. (1986). Tests of significance in theory and practice. *The Statistician*, *35*(5), 491. https://doi.org/10.2307/2987965

Katz, R. & FDA. (2002). Letter from R Katz food and Drug Administration to T Kline, regulatory affairs, Glaxosmithkline, October 21, 2002. US Regulatory Affairs Achive. https://www.industrydocumentslibrary.ucsf.edu/drug/docs/rnbn0225

Keller, M. B., Ryan, N. D., Strober, M., Klein, R. G., Kutcher, S. P., Birmaher, B., Hagino, O. R., Koplewicz, H., Carlson, G. A., Clarke, G. N., Emslie, G. J., Feinberg, D., Geller, B., Kusumakar, V., Papatheodorou, G., Sack, W. H., Sweeney, M., Wagner, K. D., Weller, E. B., & McCafferty, J. P. (2001). Efficacy of paroxetine in the treatment of adolescent major depression: A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(7), 762–772. https://doi.org/10.1097/00004583-200107000-00010

Kim, T. (2018). *Goldman Sachs ask in biotech report: Is curing patients a sustainable business model*. *CNBC*. www.cnbc.com/2018/04/11/goldman-asks-is-curing-patients-a-sustainable-businessmodel

Lasagna, L. (1960). Thalidomide—a new nonbarbiturate sleep—inducing drug. *Journal of Chronic Diseases*, *11*(6), 627–631. https://doi.org/10.1016/0021-9681(60)90061-8

Lasagna, L. (1983). Discovering adverse drug reactions. *Journal of the American Medical Association*, *249*(16), 2224–2225. https://doi.org/10.1001/jama.1983.03330400070028

Lasagna, L. (1998). Back to the future: Evaluation and drug development. In D. Healy (Eds.), The Psychopharmacologists, volume 2 (pp. 135–166). Arnold, London.

Lasagna, L., & Wardell, W. (1975). FDA, politics and the public. *Journal of the American Medical Association*, *232*(2), 141. https://doi.org/10.1001/jama.1975.03250020015015

Laughren, T. P. (2001). The scientific and ethical basis for placebo-controlled trials in depression and schizophrenia: An FDA perspective. *European Psychiatry*, *16*(7), 418–423. https://doi.org/10.1016/s0924-9338(01)00600-9

Leber, P. (1996). The role of the regulator in the evaluation of the acceptability of new drug products. In D. Healy & D. Doogan (Eds.), *Psychotropic drug development* (pp. 69–77). Chapman and Hall.

Leber, P. (1990). *FDA psychopharmacologic drugs advisory committee* [Thirty-third meeting] [Thirty-third meeting] (pp. 90–91).

Leber, P. (1998). Managing uncertainty. In D. Healy (Ed.), The Psychopharmacologists volume 2 (pp. 555–588). Arnold London.

Le Noury, J., Nardo, J., Healy, D., Jureidini, J., Raven, M., Aba-Jaoude, E., & Tufanaru, C. (2015). Restoring study 329: A randomised, controlled trial of the efficacy and harms of paroxetine and Imipramine in the treatment of adolescent major depression. *BMJ (Clinical Research Ed.)*, 351, h4320. https://doi.org/10.1136/bmj.h4320

Medical Research Council. (1948). Streptomycin treatment of pulmonary tuberculosis. *BMJ (Clinical Research Ed.)*, 769–782. https://doi.org/10.1136/bmj.2.4582.769

Oakes, M. (1986). *Statistical inference*. John Wiley and Sons.

Poole, C. (1998). Confidence intervals exclude nothing. *Am. J. Public Health*, 77, 492–493. https://doi.org/10.2105/ajph.77.4.492

Puliyel, J., & Naik, P. (2018). Revised World Health Organization (WHO)'s causality assessment of adverse events following immunization—A critique. *F1000 Research*, *7*, 243. https://doi.org/10.12688/f1000research.13694.2

Ray, K. K., Wright, R. S., Kallend, D., Koenig, W., Leiter, L. A., Raal, F. J., Bisch, J. A., Richardson, T., Jaros, M., Wijngaard, P. L. J., Kastelein, J. J. P., & ORION-10 and ORION-11 Investigators. (2020). Two phase 3 trials of Inclisiran in patients with elevated LDL cholesterol. *The New England Journal of Medicine*, *382*(16), 1507–1519. https://doi.org/10.1056/NEJMoa1912387

Ross, D. B. (2023). The decline of science at the FDA has become unmanageable. *BMJ (Clinical Research Ed.)*, *381*, 1061. https://doi.org/10.1136/bmj.p1061

Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (p. 194). Lippincott-Raven.

Rozeboom, W. W. (1997). Good science is not abductive, but hypothetico-deductive. In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 335–391). Lawrence Erlbaum.

Salsburg, D. S. (1986). *Statistics for toxicologists*. Marcel Dekker, Inc.

Sterne, J. A., & Davey Smith, G. (2001). Sifting the evidence—What's wrong with significance tests? *BMJ (Clinical Research Ed.)*, *322*(7280), 226–231. https://doi.org/10.1136/bmj.322.7280.226

Stone, M., Laughren, T., Jones, M. L., Levenson, M., Holland, P. C., Hughes, A., Hammad, T. A., Temple, R., & Rochester, G. (2009). Risk of suicidality in clinical trials of antidepressants in adults: Analysis of proprietary data submitted to US food and drug administration. *BMJ (Clinical Research Ed.)*, *339*, b2880. https://doi.org/10.1136/bmj.b2880

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine*, *358*(3), 252–260. https://doi.org/10.1056/NEJMsa065779

Wardell, W., & Lasagna, L. (1975). *Regulation and drug development*. American Enterprise Institute Washington D.C.

Woloshin, S., & Kesselheim, A. S. (2022). What to know about the alzheimer drug aducanumab (aduhelm). *JAMA Internal Medicine*, *182*(8). https://doi.org/10.1001/jamainternmed.2022.1039

World Health Organization. (2018). Causality assessment of an adverse event following immunization (AEFI): user manual for the revised WHO classification (second edition). Geneva: World Health Organization. licence: CC BY-NC-SA 3.0 IGO.

Disclosure. The authors have no relevant financial interest or affiliations with any commercial interests related to the subjects discussed within this article.

Funding. The author(s) received no specific grant or financial support for the research, authorship, and/or publication of this article.

Correspondence regarding this article should be directed to David Healy, MD FRCPsych, McMaster University Faculty of Health Sciences, Hamilton, Canada. E-mail: healy54@google-mail.com