

THE CREDENTIALS OF SCIENTIFIC EVIDENCE: REBALANCING THE EPISTEMOLOGICAL SCALES

9.1 Overview of series

Abstract

The relative value of different kinds of scientific evidence appropriate for investigating medical science, including drug investigation and comparison studies, is discussed in this series of commentaries. An understanding of Pearl's new causal inference theory is a requisite for a reassessment of the epistemological and practical value of RCTs. Extensive problems with evidence-based medicine, meta-analysis, randomised control trials, and the guidelines they give rise to, are examined from the point of view of their limited epistemological validity, or ability to elucidate the key question of causes and mechanisms, or to adequately address many key issues confronting clinical science. The necessity of using other methods is introduced, with special reference to the work of Thomas Bayes and Judea Pearl. It is argued that RCTs have achieved an unwarranted hegemony in medical science and that this has handicapped the progress of scientific investigation into causes and mechanisms. Furthermore, it is pointed out that eminent commentators on this subject, from the 1960s onward, have noted this excessive reliance on RCTs (and P values) — despite that the FDA has continued to rely heavily on RCTs¹. How the total reliance on P values was influenced by the eminence of the statistician Sir Ronald Fisher, and how persistent and widespread abuse of them has misdirected science, and how secret war-work influenced the delay in the adoption of Bayesian techniques, is discussed. I propound the argument that RCTs are not the 'Gold standard', but are closer to 'Fool's Gold' because they are epistemologically weak, and are not good tools for investigative scientific experimentation.

The succeeding commentaries in this series deal in more detail with each problem in turn: 9.2.1 external validity; 9.2.2 Simpson's paradox; 9.2.3 probability; 9.2.4 rating scales reliability; 9.2.5 the placebo response, 9.2.6 Fisher, Bayes, & more; and finally to Judea Pearl's causal inference theories 9.2.7 Pearl, causal inference.

This introductory overview paints the picture of the history and influences over EBM, guidelines, and RCTs, their inherent epistemological problems, and the attendant biases and influences that shape their presentation and prominence. It also outlines the groundwork for an understanding of how the ubiquity of RCTs have profoundly retarded the progress of investigative clinical science and how application of Pearl's causal inference methods offers a practical path forwards.

¹ As I was finalizing this commentary in January 26 the FDA issued a statement which can be seen here https://www.fda.gov/news-events/press-announcements/fda-issues-guidance-modernizing-statistical-methods-clinical-trials?utm_medium=email&utm_source=govdelivery

Background: The influence of RCTs, EBM, and guidelines

This juggler would think to charm my judgment, as mine
 eyes, obtruding false rules pranked in reason's garb
 Milton, *Comus, a masque*;
 Presented at Ludlow castle, in 1634

The notion of **Evidence Based Medicine (EBM)** dominates medical practice in most jurisdictions. EBM is largely based on the **supposed 'Gold standard' of randomised controlled trials (RCTs)** [1], of which 90% are industry sponsored — resultantly, as but one of many negative sequelae, out-of-patent drugs have few RCTs and thus less 'evidence' to gain a place in guidelines, meta-analyses, or reviews² — this represents and perpetuates a process of circular illogical reasoning³. As Professor Sir Michael Rawlins has argued, **'The notion that evidence can be reliably placed in hierarchies [as all guidelines do] is illusory'** [2];

I am sure he would have agreed with Milton that RCTs are indeed **'false rules pranked in reason's garb'**⁴.

One of these negative results is that many older drugs, across medical specialties, have receded into insignificance (e.g., of the psychotropics: clozapine, lithium, MAOIs, clomipramine, sodium valproate, L-Tryptophan). One is reminded of Oscar Wilde's words:

There is only one thing in life worse than being talked about,
 and that is not being talked about
 Oscar Wilde

This multipart commentary considers the relative merits of the epistemological credentials of RCTs, not only as a methodology for investigating drug efficacy, but, more importantly, as a tool for investigating the questions of causes and mechanisms — about these RCTs are mute. These are scientific questions that have received insufficient attention (see 9.2.7). RCTs are contrasted with emerging methodologies, particularly **the theory of causal inference of Judea Pearl — a Nobel prize winner**⁵.

It further outlines the groundwork for an understanding of how the ubiquitous hegemony of RCTs has retarded the progress of investigative clinical science. It would be helpful for there to be a greater emphasis on investigating causes and mechanisms, rather than comparisons of one drug with another, which has, historically, been a singularly unedifying process.

Material in these commentaries may be seen by some as outside the mainstream of scientific thinking; however, it should be noted that the key references given supporting central aspects of these commentaries are papers that have been cited more frequently than 99.9% of all papers. A paper that has been cited more than 100 times is in (approximately) the top .01% of all papers ever published, **many of the key citations given have been made not just hundreds of times, but thousands of times**. The citation statistics are given next to some of the papers when cited

²EBM, RCTs, meta-analyses, and treatment guidelines may be considered to overlap to an extent which means that they are semi-synonymous in much of this document

³One is reminded of the war-time incident where mechanics were reinforcing those parts of the aircraft that had returned from bombing raids, prioritising the areas where there were the most bullet-holes — until someone pointed out that they should reinforce the areas where there were not any bullet-holes — that is where the holes were in the planes that did not manage to come back.

⁴In Milton's time the word 'prank' meant, *'to dress up showily'*, and is also found in Shakespeare

⁵He in fact won the Turing prize, in 2011, which is the mathematics and IT equivalent of the Nobel prize, although most people are not familiar with that.

to emphasise this point. For example, the paper cited above by Sir Michael Rawlins, a distinguished pharmacologist, has been cited more than 700 times.

Many of these papers, e.g. [3–8], have been cited many hundreds, if not thousands, of times. This puts them in the top .01–0.001% of all papers ever published
We are evaluating the most consequential papers from eminent authors

First, some perspective and history

Even after sixty years the original purpose and precepts of EBM/RCTs have not been translated successfully into practical reality [3–8] — indeed, they often cannot be, because many questions in clinical medicine are not addressable using guidelines, rules, or RCT methodology.

Before assessing the extent to which diminishing the value of RCTs might risk ‘throwing the baby out with the bath water’ it is important to first develop an understanding of Pearl’s causal inference theory and the ‘*do*-operator’ [4], because that is a method that makes it possible to simulate an RCT using only observational data, and to extract causal connections — even at an individual-case ($n=1$) level. Pearl’s ground breaking theories are discussed in more detail in section 9.2.7 and **until Pearl’s theory is understood and incorporated into clinical investigations it will be difficult to have an informed and balanced view of their relative practical and epistemological values.**

As a prelude to enumerating and dissecting the refractory problems of RCTs it is necessary to remind ourselves that there are other scientific methods that can be used (discussed in the other commentaries in this series), which may bypass the difficulties encountered by RCTs; *scil.*, unavoidable ethical dilemmas; being ‘unfit-for-purpose’; being too time-consuming; being too costly; and being incapable of addressing many types of questions. The continuing dominance of, and preoccupation with, RCTs is an accident of history, powerfully driven by convenience, not by science. They are for the purpose of getting drugs approved by regulatory agencies and subsequently manufacturing evidence of their supposed superiority over competing treatments⁶.

RCTs have many crippling problems: external validity, ethical dilemmas, impractical, too time-consuming, too costly; unable to address many types of questions

How scientifically useful are RCTs in depression studies?

You have to know the past to understand the present
 Carl Sagan [there are innumerable versions of this notion,
 Sagan’s being the tersest]

The first antidepressants (MAOIs and tricyclics) were discovered serendipitously without the need for clinical trials [9], as were many other drugs; RCT-type trials were unnecessary because of the obvious therapeutic benefit (robust effect size) of those early ADs, and because they were used for treating more severe (melancholic) depressions [10, 11]. **Louis Lasagna, a famous clinical pharmacologist⁷ of that era and an influential consultant to the FDA in the 1960s, thought they were relying too much on RCTs** [12, 13]. RCT methodologies have subsequently added little or nothing substantive to questions of how to use ADs,

⁶ Extensive pharmaceutical company financing of research efforts over many years means that researchers are heavily invested in the RCT mentality — it will be difficult for them to reassess their value objectively. The sunk cost fallacy will influence thinking.

⁷ He has been referred to as ‘the dean of American pharmacology’

on whom to use them, nor even which drugs to prefer — it is clinical science and pharmacology that has informed us on these questions (e.g. a drug's propensity to cause sedation related to greater affinity at H1 receptors). Indeed, the propensity for that effect is so markedly different between TCAs that controls and randomisation are quite unnecessary (e.g., doxepin vs desipramine).

Additionally, RCTs have persistently misdirected clinicians⁸ by suggesting that, *inter alia*, the TCAs, SSRIs, and newer ADs, are each and all therapeutically equivalent [15]⁹, something that pharmacological science and clinical practice tells us is most certainly not the case. Since Anderson's work cited above, we have been subjected to a plague of meta-analyses, such that the eminent biostatistician from Stanford, [John Ioannidis \(H-index 278\)](#), has stated, and [this paper has been cited 1,600 times](#), [16]:

The production of systematic reviews and meta-analyses has reached epidemic proportions. Possibly, the large majority of produced systematic reviews and meta-analyses are unnecessary, misleading, and/or conflicted.

When these meta-analyses are compiled, they characteristically reject a large proportion of the RCTs that could be included, on the basis that they are unsound, in one way or another, and depending on who is making that judgement. Which studies are left out effects the conclusions. This exclusion process illustrates that a substantial proportion of the studies published are of 'unusable' quality. As I have pointed out previously, [it is possible to assemble a series of studies that are the equivalent of 'Penrose stairs' with drugs, where it can be shown that A>B>C>A](#).

These RCT-trials have now been going on for more than 60 years — it would be expected that the usefulness of them would now be obvious. That is not so in the field of psychopharmacology. The fact that it is not obvious has been exemplified by the weak conclusions¹⁰ contained in Cipriani's seminal meta-analysis [17]. Furthermore, there are serious questions about whether the billions of dollars these RCTs have cost, and the hundreds of thousands of patients who have been subjected to frequently inadequate treatment, is justified or ethical. As the **Declaration of Helsinki (DoH)** states, bad research can never be ethical¹¹.

A recent and troubling example of this is contained in the **myalgia encephalitis (ME)** story — this was a spectacular misdirection of public health policy endorsed by the UK Apex body called NICE: some have referred to that as the most egregious health-related disaster of this century [18]. Poor RCTs determined UK national guidelines and recommendations about ME treatment and supported misdirected policies [19-23]. The papers and influence of Professor Sir Simon Wessely advocated ineffective cognitive behaviour therapy and graded exercise therapy as interventions [21, 23-26]. NICE, but only after a decade, performed a *volte-face* and finally rejected the 'Wessely-camp' view, not without reactionary protest [27] — this demonstrates how RCTs, even in the hands of ostensibly respected and lauded researchers, working in 'apex' academic

⁸ The most egregious example of this is the studies on Myalgic encephalitis about which I have commented in some detail elsewhere [14].

⁹ The absurdity of this notion should be readily apparent because it is implausible that ADs with different structures and mechanisms of action could all be therapeutically equivalent — it is like claiming that all antibiotics equally effective against all types of infection

¹⁰ Conclusions which give little useful guidance about first line treatment, and do not relate to specialist practice, making them inconsequential.

¹¹ DoH 'bad medical research violates ethical standards and puts vulnerable groups at risk'.

centers (Maudsley, London¹²), are still producing catastrophically misleading results.

J K Galbraith said, *'if all else fails, immortality can always be assured by spectacular error'*. I said in an editorial [28], parodying Galbraith:

Individual experts can be wrong, but it takes a committee of experts to be spectacularly wrong

RCTs defined, but shape-shifting

The definition of an RCT is understood, or assumed to mean, different things in different circumstances. The homogeneous industry-style regulatory-approval trial dominates the literature, being about 80% of all trials [29], and probably a greater proportion for antidepressants.

Various modifications and variations described under the rubric of RCTs have either been proposed or tested¹³, and the external validation problem is being tackled by 'real-world' trials, these being referred to as pragmatic-controlled-trials, PCTs and PRCTs, e.g., [30-32]. The double-blind part of the methodology may or may not be present, indeed it may be an aspiration rather than an achievement — but it is particularly important for trials involving small effect sizes and subjective outcomes (viz. depression trials) — therein lies one indication of the difficulties that arise.

Even the control group of an RCT can be misleading, since drugs used as controls may themselves have less-than-clearly established efficacy (e.g. trazodone, mirtazapine, doxepin, moclobemide).

Practicing doctors tend to think of drug trials as comparing an active treatment with a control, which is the most common manifestation of the RCT methodology that they encounter — the classic parallel-group RCT. **However, the heart of science is about elucidating causes and mechanisms.** RCTs, as they are usually constructed and carried out in medicine, including in psychiatric illnesses, are not suited for that purpose¹⁴. An 'ideal' RCT might achieve a suggestion of causality, but it rarely does, thus making any assumptions about causal connections tenuous [33]. However, recent research on causality by the **Turing prize winner, Judea Pearl**, heralds a new approach to clinical experimentation based on powerful strategies that can inform research and **a new sort of formal real-world trial better able to reveal causes and mechanisms** [34].

Science is about elucidating causes and mechanisms — RCTs are rarely adequate for that task

Some might argue that investigating causes and mechanisms is not necessary to show that a drug works. A Bayesian will readily appreciate that a reasonable hypothesis, tapping into knowledge of mechanisms and basic pharmacology, will alter presumptions related to a prior probability (here is an explanation of [prior probability](#)), which will in turn modify the interpretation of statistical values such as the P value. A shift in thinking about the usefulness of randomised frequentist designs, *à la Fisher*, compared to Bayesian and other possible designs, is occurring (see 9.2.6); for example, Bayesian adaptive trials showed advantages in testing

¹² Widely considered the most prestigious psychiatric hospital in the Anglosphere and a premier institution for mental health research and training.

¹³ One might ask why it has taken 60 years for it to be understood that modifications are necessary.

¹⁴ This is a key issue; to the extent that RCTs may occasionally produce evidence of causality it is not because of the randomisation or the control, it is because of the rapid and definite the change in outcome — *scil.*, antibiotics for pneumonia — the objectively measured outcomes of body temperature normalizes in 48 hours and one does not need randomisation or controls to demonstrate that. There is no 'placebo' effect either, because the outcomes are objective. The rapid and definitive result constitutes the causality component — that is quite independent of any randomisation or control.

Covid treatments [35]¹⁵. At this juncture, we might also note that Fisher had no interest in clinical trials in medicine and never involved himself in them even peripherally; Armitage reported [36]:

Fisher, for his part, seems to have taken little interest in clinical medicine — I know of no written comment by him on clinical trials...

Bayesian trials are the first step along the path towards improved reasoning and methodology using causal inference and Pearl’s *do*-operator (viz. ‘strangling the rooster/cockerel’), which is discussed further below. One might note that Sir Michael Rawlings¹⁶ in his Harveian oration [2] discusses Hill’s causation ideas in the context of discussing trials, emphasising his view of the importance of mechanism and causation, over and above statistics.

Consequently, RCTs are often called the gold standard for demonstrating (or refuting) the benefits of a particular intervention. Yet the technique has important limitations of which **four are particularly troublesome**: the null hypothesis, probability, generalisability, and resource implications... Clinical practice guidelines apply in general, but each doctor must apply them to each particular patient, taking into account all of that patient’s circumstances and other relevant considerations. ... **Hierarchies of evidence should be replaced by accepting—indeed embracing—a diversity of approaches.**

If a randomised controlled trial at P 0.05 claimed that capsules of freeze-dried bullshit had an antidepressant effect, then ‘Bayesians’ would assume that there was something amiss — ‘prior probability’ changes one’s assessment of an outcome.

Bayes’ and Pearl’s ideas represent the mathematics of common sense

As the eminent philosopher Cartwright points out [38]:

RCTs, touted as the best source of evidence on effectiveness, can do so little for us... *it is surely a good idea also to use other methods that allow us to draw causal conclusions... Causal Bayes Nets methods [] derive new causal information about a population from available causal and probabilistic information from that population*

Most discoveries in clinical medicine come from clinical experimentation and serendipity, not RCTs (cf. Lasagna, p 4) — the hegemony of RCTs in recent decades has served to implicitly denigrate the value of the clinical judgement aspect of medicine, to the detriment of originality and personalised patient care. RCTs and EBM encourage a blinkered clinical approach¹⁷ — the phrase I hear during my international consultations, distressingly often, as a reason for not using a treatment, is, ‘but it is not in the guidelines’¹⁸.

Informed clinical experience and experiment already ‘unconsciously’ utilise not only Bayes’ theorem, but also Pearl’s causal inference and ‘do-operator’, which explains why clinical practice can be of greater epistemological and scientific validity compared with RCTs — indeed, it possesses the particular advantage of being able to elucidate causes and mechanisms by sequential experimentation precisely because, like common sense, it utilizes Bayesian thinking and causal inference

¹⁵ Since the initial draft of this commentary the FDA have issued the statement about accepting Bayesian methods in drug approval studies, so progress is occurring — better late than never.

¹⁶ Died. Jan 2023. I read the available obituaries about Sir Michael which were stuffy and anodyne, except for Kirby who recounted, ‘True to form to the end, as the ambulance arrived to take him to hospital for the last time, he was taken from his home on a stretcher with a cigar in his mouth and a tumbler of whiskey in his hand!’ [37] — no other obituary included this information, which I find interesting. I would have got on well with Sir Michael!

¹⁷ This excuse has two components; fear of being criticised or even censured, and laziness

¹⁸ The irony being that there are so many different guidelines that someone has written a guideline about guidelines — so how are we to decide which guideline to use?

(see section about Judea Pearl's theory of causal inference, with examples, in 9.2.6 & 7, or buy the *Book of Why*) — 'strangle the bloody rooster'¹⁹.

Pearl's theory of causal inference has been described as the mathematics of common sense

There is a convincing argument that RCTs do not, as is usually stated and claimed, elucidate or substantiate causes or mechanisms — they may sometimes establish a possibility of a causal relationship; they do not inform us about that relationship — and that is the all-important 'sciencey' component! One can also argue that in so far as they may suggest causality that is only in instances where the effect is so clear that controls and randomisation are unnecessary (cf. Hill).

Causes and mechanisms are not considered relevant or important for drug-comparison trials; nevertheless, those trials could be designed to explore such questions: however, there is no incentive to do that in industry-funded RCT trials (see 9.2.7). The myriad of RCTs done with antidepressant drugs have added nothing, or at best little, to the key questions of causes or mechanisms, nor have they added to the question of relative effectiveness (cf. the 'seminal' Cipriani AD meta-analysis [17]).

Furthermore, it is clear from extensive writing about the subject that theory concerning RCTs, whatever merits one perceives them to possess or lack, is distinct from the implementation of RCTs in the real world²⁰ [7] — clinical experiment/experience is still frequently more informative on many questions than the results of RCTs [13]. The argument that clinical practice is full of 'folklore' that is wrong is true in specific instances (it could hardly be otherwise), but it does not alter the fact that, epistemologically, clinical experience is often of superior value and validity to RCTs. Remember, even for already established risk factors, RCTs can also be plain wrong, e.g. (*inter alia*) the ME trials, and they have proved hopelessly unreliable for studying the relationship between many variables in clinical medicine; e.g. diet and disease [39].

Tangentially relevant, a recent comment in *Nature* confirms that **faked and flawed RCTs continue to be common** [40], as Ioannidis had previously shown in one of his many highly cited papers criticising science research [41]. *Nature* also reports [that 10,000 papers were retracted](#) during 2023 alone — were this reported from a source less respected than *Nature* one would hardly believe it.

The theoretical qualities and capabilities of RCTs are quite distinct from the practicalities of their implementation in the real world

It is important to recognise these problems because most practising clinicians will be unable to distinguish the studies that are useful and reliable from the majority that are not; and they are certainly most unlikely to detect those that are fraudulent. Indeed, it is clear that seasoned researchers have trouble critically analysing studies, as demonstrated by the percentage that have got through the refereeing process and into journals, when one might expect, or hope, that the refereeing process would winnow out the substantial proportion of those with flaws [4, 7, 42, 43].

A consequentially misleading result produced regarding MAOI antidepressant drugs was from an 'apex' body, full of eminent experts, the Clinical Psychiatry

¹⁹ The old Aristotelian trope of the supposed logical fallacy of *post hoc, ergo propter hoc* is not a fallacy, as can be illustrated by the simple experiment of strangling the rooster (Pearl's Do-operator), which does not affect the rising of the sun.

²⁰ As demonstrated when meta-analyses reject most of the RCTs that might be considered, on the basis that they are of inadequate quality

Committee of the Medical Research Council, *Clinical Trial of the Treatment of Depressive Illness* [44] in 1965, which concluded, to the astonishment of many working clinicians, that phenelzine was less effective than placebo in melancholic depression²¹. That seemingly authoritative conclusion has misdirected generations of clinicians and deprived innumerable patients of potentially effective MAOI treatment — **misdirection cannot get much worse than that**²².

We are never going to learn how to treat depression in an MRC
statistician's office, William Sargant

It is thus difficult to sustain the viewpoint that RCTs have inherent superiority of any sort, either in theory, or practice, as the eminent Australian, Professor Parker, argued 20 years ago [45]. That is especially the case because few trials are validated by replication — **if it is not replicated, it is not established science** [41, 46, 47].

The magnitude of the practical and epistemological problems inherent, both in reliance on the use of RCTs, and the Fisherian statistics used to evaluate them (cf. Tukey's trenchant comments about Fisher below and 9.2.6/7), are not sufficiently recognised or understood in general medical circles — hence the continued publication of such large numbers of poor quality RCTs, that, **as Ioannidis has stated, constitute 'an epidemic of false claims'** [8, 43, 48]. Soon after their uptake (1950s) many eminent scientists and researchers in the field of epistemology and clinical research were stating that RCTs were overvalued and overused, and that other methodologies and clinical experience were underused and undervalued. This series of commentaries provide ample substance to the view that **this 'Gold standard' assertion about RCTs is both a huge overstatement and an implicit denigration of the value and validity of clinical-practice knowledge, experiment, and experience** (cf. Bernard²³).

There is no qualitative epistemic difference between RCTs and
clinical experience and experimentation [49]

The entirety of the evidence requires critical re-evaluation in the light of the arguments marshalled and outlined in this series of commentaries, and a re-analysis of the relative merits of various other kinds of evidence is overdue if we are to progress in rebalancing the assessment of treatments for severe depression, for example MAOIs and ECT, and the various types of TMS.

A modified approach to investigations and trials, drawing on the understanding generated by **Pearl's theory of causal inference** may offer a practical and ethical path forward²⁴ **to elucidate key questions, not just about effectiveness, but, more importantly, about causes and mechanisms** [34, 50-53]. **The 'how and why'.**

High-quality, reliable, and useful RCTs are not common in the real world (cf. Ioannidis [41, 48, 54]). As history has now amply demonstrated, the non-ideal RCTs are of little use, partly because to make them practical and manageable external-validity often needs to be sacrificed to an extent that makes extrapolation, to real-world situations, problematic or invalid [55-58]. Indeed, estimates are that **75% of those being treated for major depression are excluded from trials** — because they do not meet the strict inclusion criteria [59].

²¹ Tony Hill was a co-author! A gaggle of experts, but they could not even get the dose of PLZ right.

²² Perhaps this is what provoked Sir William Sargant to make the perspicacious comment that, 'we are never going to learn how to treat depression in an MRC statistician's office': that powerful and profound statement is an important truth

²³ Bernard said: a physician observing a disease in different circumstances reasoning about the influence of these circumstances and deducing consequences which are controlled by other observations — this physician reasons experimentally even though he makes no experiments.

²⁴ RCTs are impractical and unethical in many situations

If the balance of evidence from RCTs for newer drugs is concluded to be less weighty (than is generally perceived), then what about the other side of the scales? How much do other methodological approaches contribute and **to what extent does an updated consideration of epistemology, and especially cause-effect relationships, change our assessment of the value of clinical practice methodology and experience**: good clinical practice is always an experiment because patients are individuals.

Is the evidence for MAOIs greater than has thus far been generally appreciated? Clinical practice experience indicates that it is. One might first note that there is extensive international support for the practical reality of the effectiveness of MAOIs amongst experienced clinicians²⁵, just as there is for ECT; that is a fact that Lasagna and many of his successors would opine cannot be dismissed. The recently formed **International MAOI Expert Group** is a focus for this and represents the view that clinical experience is informative and valuable. **Second, there are also recent meta-analyses and reviews of significant trial data in the literature [60-65] reminding us of the weight of evidence that exists supporting the effectiveness of MAOIs in depression.** It is also relevant to remind ourselves that this evidence is founded on an increased understanding of pharmacology and patho-physiology, which provides the conclusions with the added Bayesian impetus of a greater prior probability when subjected to real-world testing.

The comments below highlight that well-conducted (non-RCT) clinical research and practice-experience (some of which can be referred to as expert opinion) can produce evidence as good as, or better than, that generated by RCTs. There is no qualitative epistemological difference in the validity of the evidence thus produced, although each may be suitable in different circumstances. RCTs can mislead, clinical experience and expert opinion can mislead, but there is no qualitative difference between the two.

Historical comments by academics of note

Introduction

Progress, far from consisting in change, depends on retentiveness. ... Those who cannot remember the past are condemned to repeat it
Santayana [66]

The comments below are not an ‘appeal to authority’ but rather a reflection and a summary of historical research and opinion about matters of vital importance to the scientific endeavour. **These scientists have published serious papers (most of them in the top 0.1-1% of all published papers)** elaborating the views expressed below, and it will be to our detriment if we ignore or forget these historical contributions to epistemology.

Comments

Sir Austin Bradford Hill²⁶ [67, 69] made a point of endorsing [Claude Bernard](#)’s view that:

²⁵ It is not likely that generations of clinicians, speaking different languages, in different cultures have all come to that opinion erroneously. If that were the case we would be better off gathering herbs from the fields guided by the ‘doctrine of similars’ that guided people in the Middle Ages (it was espoused by Paracelsus).

²⁶ It was Tony Hill, as he was usually called — his name is not double-barrelled, although the literature refers to him as Austin Bradford Hill — and Sir Richard Doll who became famous for their work connecting smoking with cancer. Hill produced an early discussion of cause-effect relationships, as part of that work [67]. That paper presaged the ideas developed later by Judea Pearl, as I detailed in my review of neuroleptic malignant syndrome [68]

| there is no qualitative epistemic difference between experiment [RCTs] and observation
| [clinical science/experience] ...

Cromie [70] echoed that:

| little or no credence is now given to clinical observations even by experienced
| investigators ... while there is a blind acceptance of double-blind trials without a critical
| evaluation of their short-comings and their ability to mislead

Louis Lasagna (consultant to the FDA in the 1960s, when DESI was instituted [13, 71]) came to that view; by the mid 1960s Lasagna was saying the pendulum has swung too far towards the RCTs:

| In contrast to my role [in the FDA, and with DESI] in the 1950s, which was trying to
| convince people to do controlled trials, now I find myself telling people that it is not the
| only way to truth ... **most knowledge comes from naturalistic observations by smart
| physicians using their past knowledge and experience as control**

and Sir Austin Bradford Hill [69] went a step further in his Heberden oration²⁷ when he added:

| Any belief that the controlled trial is the only way would mean not that the pendulum
| had swung too far but that it had come right off its hook. ... no randomisation or
| statistics are needed if the results are clear...

Paul Leber, FDA head 1980s & '90s, echoed Hill, and told Professor Healy in an interview in 2008:

| If a drug really works you don't need statistics

But 'in public' Leber said RCTs were needed.

William Sargant stated, with his characteristic forthrightness, cited by Hill [69]:

| We are never going to learn how to treat depressions in an M.R.C. statistician's office

Lord Rutherford DSc. Nobel prize winner, and the father of nuclear physics, (cited in [72]) uttered a simple dictum in this terse observation:

| If your experiment needs statistics, you ought to have done a better experiment

More recent comments

The following respected authors have expressed similar reservations and criticisms:

Ashcroft [73] RCTs are:

| ...autonomous of the basic sciences...blind to mechanisms of explanation and causation
| [cf. Pearl below]

Solomon [74]:

| Emphasis on EBM has eclipsed other necessary research methods in medicine

Berwick [75]:

| We have overshot the mark with EBM

Professor Sir Michael Rawlins (a distinguished clinical pharmacologist who held many important posts in relation to drug regulation and approval) in his Harveian Oration [2], cited 700 times, argued that:

²⁷ Many current clinical trialists could learn something by re-reading this oration

The notion that evidence can be reliably placed in hierarchies [as all guidelines do] is illusory ... [RCTs have] important limitations of which **four are particularly troublesome**: the null hypothesis, probability, generalisability, and resource implications ... Consequently, RCTs are often called the gold standard for demonstrating (or refuting) the benefits of a particular intervention. Yet the technique has important limitations of which four are particularly troublesome: the null hypothesis, probability, generalisability, and resource implications... Clinical practice guidelines apply in general, but each doctor must apply them to each particular patient, taking into account all of that patient's circumstances and other relevant considerations. ... Hierarchies of evidence should be replaced by accepting—indeed embracing—a diversity of approaches.

Judea Pearl, Turing prize winner and leading researcher in the field of causality, [34]:

It is critical to realize that data are profoundly dumb about causal relationships
 Data do not understand causes and effects; humans do
 Where causation is concerned, a grain of wise subjectivity tells us more about the real world than any amount of objectivity
 Fisher's methods assume that the experimenter begins with no prior knowledge of, or opinions about, the hypothesis to be tested. They impose ignorance on the scientist
 Deep learning has instead given us machines with truly impressive abilities but no intelligence. The difference is profound and lies in the absence of a model of reality
 Despite heroic efforts by the geneticist Sewall Wright (1889–1988), causal vocabulary was virtually prohibited for more than half a century. And when you prohibit speech, you prohibit thought and stifle principles, methods, and tools
 Without the ability to envision alternate realities and contrast them with the currently existing reality, a machine...cannot answer the most basic question that makes us human: "Why?"
 A mantra most scientists can recite in their sleep is that correlation does not imply causation; but they do not grasp the depth of it²⁸ ... **Causality is the key: there is no way of doing science without causality, it is the *sine qua non* for all understanding and progress**

Which explains why Fisher's frequentist statistics are so often of restricted utility, or even useless.

Causality is the key: there is no way of doing science without causality, it is the *sine qua non* for all understanding and progress

Nancy Cartwright, a well-known philosopher, from her much-cited paper *Are RCTs the gold standard? (cited nearly 1,000 times)*, and see also her more recent paper with Deaton **(cited 2,500 times)**. The abstract of sums up their point well [38, 56].

External validity for RCTs is hard to justify ... RCTs can play a role in building scientific knowledge and useful predictions but **they can only do so as part of a cumulative program, combining with other methods, including conceptual and theoretical development, to discover not 'what works', but 'why things work'**.

Statisticians rarely if ever use the word 'cause', thanks to Fisher's influence²⁹

Should we, must we, will we, now concede at last that the 'Gold-standard' is nothing but fool's gold?

In conceding this, we can then focus on other methodologies that address the experimental science-based questions 'how' and 'why'.

²⁸ I would add that, whilst many recite that mantra, they conspicuously fail to abide by its implications, or modify their speculations and conclusions accordingly

²⁹ Indeed, Pearl has noted that the word 'cause' does not appear in the index of any book on statistics

A parody

Sigh no more, friends, sigh no more, RCTs were deceivers
 ever,
 One foot in SEs and one unsure, to science constant never
 [KG]

Sigh no more, ladies, sigh no more, men were deceivers ever,
 One foot in sea and one on shore, to one thing constant
 never [WS]

When Ashcroft stated that RCTs were ‘**autonomous of the basic sciences and blind to mechanisms of explanation and causation**’ he was making a statement that takes us to the heart of the epistemology of experimental science and the differences between ‘Fisherian’ frequentist approaches, basic research in science and biology looking at causes and mechanisms, Bayesian approaches, and Pearl’s innovative science of causation [34, 38, 50-53]

Whilst Fisher thought that statistics was the ‘**grammar of science**’, Judea Pearl has said something fundamentally different in this statement, ‘**it is impossible to deal with causal relationships with statistical language**’.

Causes and mechanisms are the heart of science, without which it is impotent

Albert Einstein:

The development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment

RCTs are not ‘systematic experiments’.

Crucially, RCT-based data are lacking for much of medical practice and supportive RCTs are never likely to be accomplished successfully, not just because of their scientific impotence and inapplicability, but because of the major practical, logistic, ethical, and financial problems of carrying out such trials. This leaves large uncertain areas for many conditions [76], where other methods must be utilised.

Causes and mechanisms, elucidated by systematic experiments, are both the *raison d’être* and sine qua non, of serious science; yet RCTs in psychiatry are not able to address those questions — their role is more for comparing one drug with another, especially in relation to short-term symptom changes and short-term side-effects — neither are they generally practical for establishing long-term treatment efficacy or long-term side-effects.

While the ideal RCT can suggest a cause-effect relationship, that situation, in so far as it theoretically exists, would *ipso facto* make that RCT redundant — if the result is sufficiently clear to establish a cause-effect relationship, then randomisation, controls, and statistical analysis would be unnecessary (cf. Hill). **It thus follows, with an elegant inevitability, that any causal evidence produced by an RCT is independent of randomisation and controls.**

If an RCT provides evidence of causality, it does that independently of randomisation, controls, and statistical analysis, thereby confirming its own pointlessness

Specific problems of RCTs: Prelude

In the seven subsequent commentaries in this series I discuss the many specific problematic facets of RCT methodology, *à la* Rawlins above (9.2.1 to 9.2.7).

There I emphasise that RCTs, like Fisherian frequentist statistics, exist divorced from real-world science, especially in the psychiatric arena.

This introductory commentary has focused on history and general epistemology, but before enumerating RCTs' specific epistemological problems we must acknowledge other overarching and serious issues that permeate this analysis. First, we must acknowledge that a proportion of medical researchers do not have an adequate training in epistemology, statistics, or even research methodologies; hence the large proportion of RCTs that are excluded from MAs: second, we must recognise the effect of influences outside science, political, and financial influences particularly, which contaminate the scientific endeavour to an undesirable and counter-productive extent. The example of this that echoes in my recollection is the lamentable episode where my old associate from our early days at Guys hospital, Dave Nutt, was ousted from his position heading the committee that advised to the government on the harms of drugs. He produced the first scientific and comprehensive analysis of the harms of drugs, but the results did not suit the political establishment, so they followed the standard playbook of the ethically and intellectually inadequate and shot the messenger.

A proportion of medical researchers do not have an adequate training in epistemology, or statistics, or even research methodologies

Anyone seeking scientific truth ignores at their peril recognition of the all-pervading influence of industry, and the corporatisation of the whole of the medical science, education, and publishing world. That is dealt with in detail in a [separate commentary](#) — some key recent references may be noted in passing [16, 41, 48, 77-80]. The retraction watch website will help keep you up-to-date with the avalanche of retracted papers³⁰. <https://retractionwatch.com/the-retraction-watch-leaderboard/>

Never has the admonition *caveat lector* been more germane. Indeed, a high degree of critical reading and thinking skills is now a requisite. **Remember, it is not established science until it is independently replicated, preferably more than once.**

Unfortunately, RCTs rarely meet Cartwright's standard requirement that they be **part of a cumulative program, combining with other methods, including conceptual and theoretical development, to discover not what works, but why things work.**

RCTs, if they are to be useful, must be part of a cumulative program, with a sound foundation in basic science, and combining with other methods, to discover not what works, but how and why it works

Why, why, why, have we not grasped the cardinal importance of why?

Buy Judea Pearl's [‘The Book of Why: The New Science of Cause and Effect’](#) !

References

1. Jones, D.S. and S.H. Podolsky, *The history and fate of the gold standard*. *Lancet*, 2015. **385**(9977): p. 1502-3.
2. Rawlins, M., *De testimonio: on the evidence for decisions about the use of therapeutic interventions*. *Lancet*, 2008. **372**(9656): p. 2152-61.

³⁰ This avalanche is like an iceberg, we are only seeing the tip of it

3. Sackett, D.L., et al., *Evidence based medicine: what it is and what it isn't*. BMJ, 1996. **312**(7023): p. 71-2.
4. Ioannidis, J.P., *Why most published research findings are false*. Public Library of Science: Medicine, 2005. **2**(8): p. e124.
5. Feinstein, A.R., *Meta-analysis: statistical alchemy for the 21st century*. J Clin Epidemiol, 1995. **48**(1): p. 71-9.
6. Feinstein, A.R. and R.I. Horwitz, *Problems in the "evidence" of "evidence-based medicine"*. The American journal of medicine, 1997. **103**(6): p. 529-535.
7. Leichsenring, F., et al., *The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses*. World Psychiatry, 2022. **21**(1): p. 133-145.
8. Ioannidis, J.P., *Why Most Clinical Research Is Not Useful*. PLoS Med, 2016. **13**(6): p. e1002049.
9. Kuhn, R., *The treatment of depressive states with G 22355 (imipramine hydrochloride)*. Am J Psychiatry, 1958. **115**(5): p. 459-64.
10. Shorter, E., *Before prozac: the troubled history of mood disorders in psychiatry*. 2009: Oxford University Press.
11. Carroll, B.J., *Bringing back melancholia*. Bipolar Disord, 2012. **14**(1): p. 1-5.
12. Lasagna, L. and P. Meier, *Clinical evaluation of drugs*. Annu Rev Med, 1958. **9**: p. 347-54.
13. Lasagna, L., *Clinical trials of drugs from the viewpoint of the academic investigator (a satire)*. Clin Pharmacol Ther, 1975. **18**(5 Pt 2): p. 629-33.
14. Gillman, P.K., *Epistemology, pharmacology, knowledge sources, and legal responsibility for accurate information*. Drug Science, Policy and Law, 2026: p. in press.
15. Anderson, I.M., *SSRIS versus tricyclic antidepressants in depressed inpatients: a meta-analysis of efficacy and tolerability*. Depression and Anxiety, 1998. **7**(Suppl 1): p. 11-7.
16. Ioannidis, J.P., *The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses*. Milbank Q, 2016. **94**(3): p. 485-514.
17. Cipriani, A., T.A. Furukawa, and G. Salanti, *Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis*. Lancet, 2018. **391**(10128): p. 1357-1366
[http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(17\)32802-7/fulltext](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)32802-7/fulltext).
18. Monbiot, G., *First, Do No Harm*. www.monbiot.com, 2021: p. <http://www.monbiot.com/2024/03/27/first-do-no-harm/>.
19. Vink, M. and A. Vink-Niese *The Updated NICE Guidance Exposed the Serious Flaws in CBT and Graded Exercise Therapy Trials for ME/CFS*. Healthcare, 2022. **10**, 898 DOI: 10.3390/healthcare10050898.
20. Thoma, M., et al. *Why the Psychosomatic View on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Is Inconsistent with Current Evidence and Harmful to Patients*. Medicina, 2024. **60**, 83 DOI: 10.3390/medicina60010083.
21. Marks, D.F., *The Rise and Fall of the Psychosomatic Approach to Medically Unexplained Symptoms, Myalgic Encephalomyelitis and Chronic Fatigue Syndrome*. Archives of Epidemiology & Public Health Research, 2022. **1**(2).

22. Harvey, S.B., et al., *The relationship between prior psychiatric disorder and chronic fatigue: evidence from a national birth cohort study*. Psychol Med, 2008. **38**(7): p. 933-40.
23. White, P., et al., *Anomalies in the review process and interpretation of the evidence in the NICE guideline for chronic fatigue syndrome and myalgic encephalomyelitis*. J Neurol Neurosurg Psychiatry, 2023. **94**(12): p. 1056-1063.
24. Adamson, J., et al., *Cognitive behavioural therapy for chronic fatigue and chronic fatigue syndrome: outcomes from a specialist clinic in the UK*. J R Soc Med, 2020. **113**(10): p. 394-402.
25. Geraghty, K.J., 'PACE-Gate': *When clinical trial evidence meets open data access*. J Health Psychol, 2017. **22**(9): p. 1106-1112.
26. Ahmed, S.A., J.C. Mewes, and H. Vrijhoef, *Assessment of the scientific rigour of randomized controlled trials on the effectiveness of cognitive behavioural therapy and graded exercise therapy for patients with myalgic encephalomyelitis/chronic fatigue syndrome: A systematic review*. J Health Psychol, 2020. **25**(2): p. 240-255.
27. White, P., et al., *Anomalies in the review process and interpretation of the evidence in the NICE guideline for chronic fatigue syndrome and myalgic encephalomyelitis*. J Neurol Neurosurg Psychiatry, 2023. **94**(12): p. 1056-1063.
28. Gillman, P.K., *Monoamine oxidase inhibitors: A paradigm of poor science*. Journal of Psychopharmacology, 2025. **39**(12): p. 02698811251381762.
29. Hopewell, S., et al., *The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed*. Bmj, 2010. **340**: p. c723.
30. Jeong, N.-Y., et al., *Pragmatic Clinical Trials for Real-World Evidence: Concept and Implementation*. cpp, 2020. **2**(3): p. 85-98.
31. Leather, D.A., et al., *Real-World Data and Randomised Controlled Trials: The Salford Lung Study*. Adv Ther, 2020. **37**(3): p. 977-997.
32. Porzolt, F., et al., *Form follows function: pragmatic controlled trials (PCTs) have to answer different questions and require different designs than randomized controlled trials (RCTs)*. Journal of Public Health, 2013. **21**(3): p. 307-313.
33. Cartwright, N., *A philosopher's view of the long road from RCTs to effectiveness*. Lancet, 2011. **377**(9775): p. 1400-1.
34. Pearl, J. and E. Mackenzie, *The book of why: the new science of cause and effect*. 2018: Basic books.
35. Ryan, E.G., D.-L. Couturier, and S. Heritier, *Bayesian adaptive clinical trial designs for respiratory medicine*. Respirology, 2022. **27**(10): p. 834-843.
36. Armitage, P., *Fisher, Bradford Hill, and randomization*. Int J Epidemiol, 2003. **32**(6): p. 925-8; discussion 945-8.
37. Kirby, R., *Professor Sir Michael David Rawlins, 1941–2023*
Professor Sir Michael David Rawlins, 1941-2023. Trends in Urology & Men's Health, 2023. **14**(2): p. 37-37.
38. Cartwright, N., *What evidence should guidelines take note of?* J Eval Clin Pract, 2018. **24**(5): p. 1139-1144.
39. Temple, N.J., *How reliable are randomised controlled trials for studying the relationship between diet and disease? A narrative review*. British Journal of Nutrition, 2016. **116**(3): p. 381-389.
40. Van Noorden, R., *Medicine is plagued by untrustworthy clinical trials. How many studies are faked or flawed?* Nature, 2023. **619**(7970): p. 454-458.

41. Ioannidis, J.P.A., *The Reproducibility Wars: Successful, Unsuccessful, Uninterpretable, Exact, Conceptual, Triangulated, Contested Replication*. Clin Chem, 2017. **63**(5): p. 943-945.
42. Ioannidis, J.P., *Contradicted and initially stronger effects in highly cited clinical research*. Jama, 2005. **294**(2): p. 218-28.
43. Ioannidis, J.P., *An epidemic of false claims. Competition and conflicts of interest distort too many medical findings*. Sci Am, 2011. **304**(6): p. 16.
44. Thiery, M., et al., *Clinical Trial of the Treatment of Depressive Illness. Report to the Medical Research Council by Its Clinical Psychiatry Committee*. Br Med J, 1965. **1**(5439): p. 881-6.
45. Parker, G., *Evaluating treatments for the mood disorders: time for the evidence to get real*. Australian and New Zealand Journal of Psychiatry, 2004. **38**(6): p. 408-14.
46. Leppink, J. and P. Pérez-Fuster, *What is science without replication?* Perspect Med Educ, 2016. **5**(6): p. 320-322.
47. Picho, K., L.A. Maggio, and A.R. Artino, Jr., *Science: the slow march of accumulating evidence*. Perspect Med Educ, 2016. **5**(6): p. 350-353.
48. Ioannidis, J.P.A., *Hundreds of thousands of zombie randomised trials circulate among us*. Anaesthesia, 2021. **76**(4): p. 444-447.
49. Bernard, C., *A N'etude de la medecin eexperimentalte*. Bailliere, Paris (Trans. H. C. Greene, 1927), 1865.
50. Pearl, J., *Comment: Understanding Simpson's Paradox*. American Statistical Association, 2014. **68**: p. 8-13.
51. Pearl, J., *On the Interpretation of do(x)*. Journal of Causal Inference, 2019: p. Feb.
52. Pearl, J., *Causal diagrams for empirical research (with discussions)*, in *Probabilistic and causal inference: The works of Judea Pearl*. 2022. p. 255-316.
53. Pearl, J., M. Glymour, and N.P. Jewell, *Causal inference in statistics: A primer*. 2016: John Wiley & Sons.
54. Flacco, M.E., et al., *Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor*. J Clin Epidemiol, 2015. **68**(7): p. 811-20.
55. Buckley, N.A., I.M. Whyte, and A.H. Dawson, *Diagnostic data in clinical toxicology--should we use a Bayesian approach?* J Toxicol Clin Toxicol, 2002. **40**(3): p. 213-22.
56. Deaton, A. and N. Cartwright, *Understanding and misunderstanding randomized controlled trials*. Soc Sci Med, 2018. **210**: p. 2-21.
57. Pearl, J. and E. Bareinboim, *External validity: From do-calculus to transportability across populations*, in *Probabilistic and causal inference: The works of Judea Pearl*. 2022. p. 451-482.
58. Huebschmann, A.G., I.M. Leavitt, and R.E. Glasgow, *Making health research matter: a call to increase attention to external validity*. Annual review of public health, 2019. **40**: p. 45-63.
59. Blanco, C., et al., *Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions*. J Clin Psychiatry, 2008. **69**(8): p. 1276-80.
60. Heijnen, W.T., et al., *Efficacy of Tranylypromine in Bipolar Depression: A Systematic Review*. J Clin Psychopharmacol, 2015. **35**(6): p. 700-5.
61. Bellos, A., *Multicriteria Decision-Making Methods for Optimal Treatment Selection in Network Meta-Analysis*. Medical Decision Making, 2023. **43**(1): p. 78-90.

62. Bahji, A., et al., *Comparative efficacy and tolerability of pharmacological treatments for the treatment of acute bipolar depression: A systematic review and network meta-analysis*. Journal of Affective Disorders, 2020. **269**: p. 154-184.
63. Ricken, R., et al., *Tranylcypromine in mind (Part II): Review of clinical pharmacology and meta-analysis of controlled studies in depression*. Eur Neuropsychopharmacol, 2017. **(8):Epub 2017 Jun 1.**: p. 714-731.
64. Suchting, R., et al., *Revisiting monoamine oxidase inhibitors for the treatment of depressive disorders: A systematic review and network meta-analysis*. J Affect Disord, 2021. **282**: p. 1153-1160.
65. Giménez-Palomo, A., et al., *Efficacy and tolerability of monoamine oxidase inhibitors for the treatment of depressive episodes in mood disorders: A systematic review and network meta-analysis*. Acta Psychiatrica Scandinavica, 2024: p. <https://onlinelibrary.wiley.com/doi/abs/10.1111/acps.13728>.
66. Santayana, G., *The Life of Reason*. 1905: http://www.gutenberg.org/catalog/world/readfile?fk_files=1498120.
67. Hill, A.B., *The environment and disease: association or causation? 1965*. J R Soc Med, 2015. **108**(1): p. 32-7.
68. Gillman, P.K., *Neuroleptic Malignant Syndrome: Mechanisms, Interactions and Causality*. Movement Disorders, 2010. **25**(12): p. 1780-1790.
69. Hill, A.B., *Reflections on controlled trial*. Ann Rheum Dis, 1966. **25**(2): p. 107-13.
70. Cromie, B.W., *The Feet of Clay of the Double-Blind Trial*. Lancet, 1963. **2**(7315): p. 994-7.
71. Lasagna, L., *Problems of Drug Development. The Government, the Drug Industry, the Universities, and the Medical Profession: Partners or Enemies?* Science, 1964. **145**(3630): p. 362-7.
72. Bailey, N. and E. Rutherford, *The mathematical approach to biology and medicine*. London: Wiley, 1967. **23**.
73. Ashcroft, R.E., *Current epistemological problems in evidence based medicine*. J Med Ethics, 2004. **30**(2): p. 131-5.
74. Solomon, M., *Just a paradigm: evidence-based medicine in epistemological context*. European Journal for Philosophy of Science, 2011. **1**(3): p. 451.
75. Berwick, D.M., *Broadening the view of evidence-based medicine*. Qual Saf Health Care, 2005. **14**(5): p. 315-6.
76. Frieden, T.R., *Evidence for Health Decision Making - Beyond Randomized, Controlled Trials*. N Engl J Med, 2017. **377**(5): p. 465-475.
77. Carlisle, J.B., *False individual patient data and zombie randomised controlled trials submitted to Anaesthesia*. Anaesthesia, 2021. **76**(4): p. 472-479.
78. Ioannidis, J.P.A., *Cochrane crisis: Secrecy, intolerance and evidence-based values*. Eur J Clin Invest, 2019. **49**(3): p. e13058.
79. Howick, J., et al., *Most healthcare interventions tested in Cochrane Reviews are not effective according to high quality evidence: a systematic review and meta-analysis*. J Clin Epidemiol, 2022. **148**: p. 160-169.
80. Carlisle, J.B., *Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals*. Anaesthesia, 2017. **72**(8): p. 944-952.