Perspective

Personalized Medicine

Antidepressant-induced suicidality: how translational epidemiology incorporating pharmacogenetics into controlled trials can improve clinical care

David Healy[1], Kim Bechthold[2] & Peter Tolias*[,3]

[1]Department of Psychological Medicine, Cardiff University, Wales, UK
[2]Chief Executive Officer, Sundance DX, Inc., 1630A 30th Street, Suite 378, Boulder, CO 80301, USA
[3]Center for Healthcare Innovation, Stevens Institute of Technology, 507 River Street, Room 515, Hoboken, NJ 07030, USA
*Author for correspondence:
Tel.: +1 201 216 8253
Fax: +1 201 216 8196
ptolias@stevens.edu

Randomized controlled trials (RCTs) have been a staple of the drug development process for several decades. Here, we review the origins of RCTs and their adoption within drug development, highlighting shortcomings that tend to be ignored and possible solutions offered from personalized medicine. While RCTs play an important role in development of therapeutics, we underscore how if used indiscriminately, their adverse effects may outweigh the benefits. As an example, we focus on the development of antidepressants and how a severe adverse drug response – suicidal ideation – can be overlooked. We conclude with a discussion of how pharmacogenetics may address some of the deficiencies of RCTs, bringing the focus of drug response back to the individual patient rather than the population, using as an example the discovery of genetic markers associated with antidepressant-induced suicidal ideation.

Keywords: antidepressants • drug response • pharmacogenetics • randomized controlled trials • suicidal ideation • translational epidemiology

The origin of randomized trials

Ronald Fisher created the modern randomized controlled trial (RCT) in the 1920s, when investigating the effect of fertilizers. Many factors can confound fertilizer studies such as differences in soil drainage, exposure to wind or sunlight, and a myriad of soil elements. The known factors can be controlled for, but Fisher's insight was that he could control both known and unknown confounders by randomizing the fertilizer to alternate soil patches.

Fisher tied significance testing to randomization. If we got the same result every time, we had designed a good experiment. There was a *quod erat demonstrandum* quality to this – shave a bit off one side of a coin and you can expect heads to come up 19 times out of 20. Randomization was about leaving nothing to chance. This insight on what Fisher meant by an RCT has slipped out of view [1–3].

Fisher's statistical significance (SS 1) indicated that the experimenter knew what they were doing. The statistically significant (SS 2) findings cited in most drug trials indicate no such thing [4].

Randomization was first used in a treatment trial of streptomycin in tuberculosis by Bradford Hill (MRC 1948) [5]. Earlier nonrandomized trials had established all that is known about streptomycin for tuberculosis – that it works in the short term but that the germ becomes resistant and treatment comes with a significant risk of ototoxicity [6]. Thus, the MRC trial put randomization rather than streptomycin on the map; it demonstrated the efficacy of trials but did not establish their effectiveness, where efficacy means that trials do something while effectiveness means that they work for the intended purpose.

Early doubts about RCTs

While clinical trials are now thought of as a way to contain pharmaceutical company claims, by the mid-1960s Bradford Hill noted that drug company salesmen could be heard deploying RCT evidence to encourage doctors to use their company's products [7]. However, in contrast to the current enthusiasm for

Future Medicine part of fsg

RCTs, Hill [7] noted that if RCTs ever became the only way to evaluate drugs, that the pendulum would not just have swung too far, it would have come off its hook.

It was recognized in the 1950s that the philosophical basis of RCTs was uncertain; there was no agreement on the meaning of statistical significance, with Neyman and Pearson differing from Fisher; no logical basis for randomization had been elaborated then or now [6,8].

The primacy of RCTs today, as a method of evaluation, stems not from greater rational or logical coherence, but from events centering on the thalidomide crisis. Thalidomide created a political imperative to be seen to do something to make patients safer. As a result in 1962, a change to the provisions of the Food and Drugs Act required companies to demonstrate the 'effectiveness' of new compounds, with an understanding that this would be done through placebo-controlled RCTs.

As of 1962, RCTs were a novel evaluation method whose suitability for the task at hand was uncertain [9]. There is no better symbol of this uncertainty than the fact that as of 1962, only one drug had demonstrated effectiveness and safety through a placebo-controlled RCT prior to marketing – thalidomide [10]. By contrast, a placebo-controlled RCT of imipramine [11] failed to demonstrate effectiveness.

Finally, there is a crisis within drug development today that sits poorly with claims that RCTs are an effective evaluation method. As Table 1 shows, most major drug groups were introduced in the 1950s without the benefit of RCTs, and the drugs that were introduced then remain more effective than treatments that have come to the market since through RCTs. Empirically, therefore, it appears that RCTs are not necessary and may not be helpful in developing an effective drug arsenal.

There is a need to distinguish this from another crisis concerning the conduct of RCTs linked to the use of surrogate outcomes, in trials of inadequate duration, against a regulatory background that will license products on the basis of two positive RCTs even if there are ten or more negative studies, with almost all publications ghost-written and all trial data withheld [9]. This second crisis obscures the role of RCTs in the drug development crisis.

This paper is not an addition to these critiques of the ways RCTs are conducted. It argues that RCTs have an important place in therapeutics, but if used indiscriminately their adverse effects may outweigh their benefits. Adapting Muir Gray's dictum that all screening is harmful, we might say that all RCTs are harmful but in some instances there are also benefits that warrant taking the unavoidable risks involved [12].

Mediculture or medicine?

The argument in brief is: humans and their diseases and the treatment of those diseases are not one-dimensional in the way Fisher's soil patches were and growing crops is. One set of problems deploying RCTs in medi-

Table 1. Drug effectiveness with and without randomized controlled trials.

Drug groups from the 1950s	Exemplars of 1950s medicines	Later medicines: more effective or not
Analgesics	Morphine, paracetamol	No
Antibiotics	Penicillins, tetracyclines	No
Anticonvulsants	Barbiturates, valproate phenytoin	Possible
Antidepressants	Tricyclics, MAOIs	No
Antihistamines	Chlorphenamine, diphenhydramine	No
Antihypertensives	Thiazides	No
Antipsychotics	Clozapine, haloperidol	No
Chemotherapies	Nitrogen mustards, cisplatin	Perhaps
Contraceptives	Second-generation COC	No
Diuretics	Furosemide	No
Hypoglycemics	Metformin	No
Steroids	Prednisone	No
Stimulants	Dexamphetamine methylphenidate	No
Tranquilizers	Diazepam	No
Vaccines	Polio, smallpox	No

COC: Combined oral contraceptives; MAOI: Monoamine oxidase inhibitors.

cine stems from the fact that transforming a chemical into a medicine is a different matter to demonstrating that a chemical is an effective fertilizer. Another set of problems stems from the conditions these medicines may be used to treat. These twin sets of problems introduce uncertainties into the evaluation of medicines that randomization may compound rather than control.

Medicines are not fertilizers. A fertilizer has only one action we need pay heed to, but drugs may have a hundred effects, all of which need attention. It is not problematic to designate a primary effect in an RCT of fertilizers and ignore others, but this is problematic in medicine, especially if the choice of effect is dictated by business advantage.

Randomization aims at eliminating sources of objective bias. Can there be a control of bias in SSRI trials; for instance, if the possible effects of these drugs on mood are designated as primary when these are less likely to happen than effects on sex and bowel function?

Medicine is not mediculture. Trials of fertilizers aim at establishing the average effects of the chemical. The fact that a small proportion of ears of corn might die prematurely because of the fertilizer is of no consequence. But medicine is critically concerned with the benefit to an individual patient and average effects are only useful in so far as they might be of help to the individual. Average effects that obscure harm to the individual patient entail risks that may not be worth taking. Clinical practice wants to discover heterogeneity, not obliterate it.

Randomization undertaken to manage unknown unknowns requires a focus on one effect of treatment. This focus generates an ignorance of ignorance regarding other effects. The process is akin to hypnosis, where holding a subject's attention to one focus can lead them to miss more important material out of focus, especially when for the sake of objectivity patients' reports are essentially ignored.

Placebo effect

RCTs of fertilizers are not controlled with placebos. The first RCTs in medicine were not placebo controlled and the first placebo-controlled trials in medicine were not RCTs. It seems so appropriate superficially to marry RCTs and placebos, in that both aim at controlling the bias of experimenters, that there has been little questioning of this relationship. The marriage of placebos and RCTs gives the impression that a further set of biases is being controlled – ironically that 'hypnotic' effects are being eliminated. However, placebo controls, *de facto*, introduce a systematic bias so that an active treatment simply needs to beat placebo on some dimension to be adjudged as working. This can be achieved for weaker and weaker agents by powering trials accordingly. Manipulations of this sort may be responsible in part for the fact that recent antihypertensives, hypoglycemics and antidepressants are in general weaker than treatments introduced without RCTs in the 1950s (Table 1).

It is far from clear whether the effects of placebos in RCTs control for, rather than introduce, confounders. It seems highly likely that nontreatments, placebos and active placebos will perform differently in most trials. It is clearly an assumption that the interaction between active treatment and placebo effects are the same across all trials, even across all trials of a particular drug group.

Antidepressants & suicide: a thought experiment

The differences between medicine and agriculture can be drawn out through two examples involving antidepressants and suicide in patients with major depressive disorder (MDD), but the lessons learned apply to all drug groups and all drug effects.

Depression is a major healthcare issue in the USA. The CDC reports that between 2005 and 2008, 11% of all Americans were taking antidepressants – the most frequently used medication in individuals aged 18–44 years [13]. MDD is associated with a suicide rate of 2–9% [14,15]. Antidepressant medications have demonstrated beneficial effects for MDD patients, but there is evidence that a subgroup of patients (6–13%) develop treatment-emergent suicidal ideation (TESI) and behavior in the early weeks following the initiation of therapy, dosage change or on withdrawal from medication [16,17]. In 2004 and 2005, regulatory agencies in the USA and internationally cautioned doctors and patients of the risk of TESI in their black box warning on all classes of antidepressant drugs.

Imipramine, the first antidepressant, was launched in 1958 without RCTs, which came later and were not uniformly positive. A year later in 1959, a meeting of psychiatrists was convened in Cambridge to discuss its effects [18]. Imipramine and related tricyclic antidepressants, such as amitriptyline and clomipramine, are serotonin reuptake inhibitors but are more potent antidepressants than drugs that were developed later. They produce significant responses in patients with melancholia (or raised cortisol levels), where later drugs do not, and 'beat' later drugs in more severely depressed clinical populations [19]. Melancholic patients are 80-times more likely to commit suicide than mildly depressed patients [20]. Accordingly, comparing imipramine and placebo in a RCT of melancholic patients would likely show

fewer suicides and suicidal acts on imipramine than on placebo. The relative risk might be as low as 0.5 (Figure 1). Imipramine in this assay system protects against suicide.

At the meeting in Cambridge in 1959 [18], several clinicians noted on the basis of the Christmas TreeLight-Bulb test (Challenge–Dechallenge–Rechallenge) that, wonderful though imipramine was for many of their patients, it could trigger suicidal and homicidal ideation in some. When Christmas Trees had light bulbs, after a year laid up, there was an annual drama when the lights failed to work. Unscrewing them sequentially would lead to one which when unscrewed lit the set up. Screwing that bulb back in turned them off again. Jettisoning this bulb fixed the problem. This is at least as convincing a demonstration of cause and effect as statistical significance: as advocated by Fisher and more convincing that the statistical significance; and testing deployed in clinical trials today.

The findings of our imipramine thought experiment are represented schematically in Figure 1. These are the RCT findings for a drug that causes suicide. By contrast, Figure 2 shows the results of FDA's meta-analysis of findings for suicides and suicidal acts in selective serotonin-reuptake inhibitor (SSRI) and post-SSRI trials [21]. We get a relative risk that the drugs will cause suicide and suicidal acts of 2.0. This outcome comes about in part because these later drugs are weaker than imipramine, and so were tested in people who were at much less risk of suicide; as a result, the rate of suicidal acts on placebo falls and any effect of the drug to cause suicide becomes more noticeable.

As a matter of historical record, when a rough doubling of the relative risk of suicidal acts on antidepressants over placebo became statistically significant, regulators stated that there was now evidence that SSRIs caused suicide and came out with a warning to that effect.

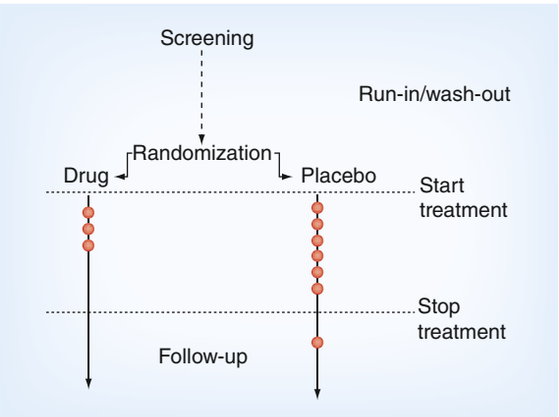

Figure 1. Randomized controlled trials. Imipramine in melancholia suicidal acts.

The regulatory position was wrong. The data as of 2006 show that in these assays, the drugs produce an excess of suicides and suicidal acts over lives saved. They say nothing about causality except in so far as there could not be an excess of suicides and suicidal acts if the drugs don't cause suicide. In the case of the SSRIs and suicide, the first description of this problem came from Teicher *et al.* It is now clear that the Christmas Tree light bulb descriptions in this paper on Prozac and suicide provided better evidence about its capacity to cause suicidality than came from any of the clinical trials run by companies [22].

The possibility that a treatment that can cause an effect would in controlled trials sometimes give rise to exactly the opposite outcome has previously been noted [23–25]. The possibility tends to be subsumed under the heading of confounding by indication. However, rather than the illness confounding things, the dose of both the illness and the drug can change the outcome, and can be active at the same time. No one appears to have addressed the question as to whether effects of this type compromise the capacity of randomized trials to come up with the right answer.

There are analytic techniques to manage effect modification but there is more involved here. Even if only effect modification is involved, we need an explanation for why such techniques failed when it came to establishing whether antidepressants could cause suicide. There is a further complexity. Imipramine and other antidepressants have multiple other and differing effects aside from their actions on the serotonin system, which there is good reason to believe can trigger or mitigate suicidality in their own right, and these effects kick in at different doses and potentially at different doses in different populations. We return to this point later and introduce how the underlying genetics of trial participants can affect outcomes.

Implications for RCTs

It is common to hear claims that RCTs demonstrate cause and effect. This was certainly the intention in Fisher's trials. However, if the trial is not designed to look at an issue, as in this case of antidepressants and suicide, it is clear that the findings in RCTs do not show cause and effect. Challenge–Dechallenge–Rechallenge relationships in contrast can show cause and effect in individual cases and physiological effects, including genetic markers, are likely to be of considerable importance in finally establishing cause of effect.

Another implication is that RCTs do not give reliable data on frequency. Even where a study is designed to look at antidepressants and suicide, we cannot in fact infer from RCTs how often antidepressants in clinical practice might trigger suicidality.

Based on what is now known about these issues, a range of relative risks from 0.5 to 2.0 could be produced by judicious choice of drug and patient. The issues are well enough understood at this point so that, in line with Fisher's original intention, randomization could be used to produce close to whatever result was wanted. By contrast, RCTs were deployed in the 1990s to obscure rather than reveal such treatment effects. The key point is that this finding is not an inconvenience that stems from some oddity to do with antidepressants or suicide. It is intrinsic to RCTs within medicine. It can be expected every time a treatment and an illness produce, at least superficially, similar outcomes – whether a benefit or a harm.

These problems happen as often with the benefits of treatments as with their harms. What are termed benefits are often simply effects of treatment, in some instances happening with less frequency than some effects designated adverse. The number of living participants at the end of a trial is arguably at least as appropriate a measure of effectiveness as anything else. In RCTs of antidepressants, there have in fact been more dead bodies in the active treatment arms of trials than on placebo [21].

The verdict of RCTs is often pitted against clinical judgement. But in fact both clinicians and patients, if asked, can often distinguish between depression-induced suicidality and drug-induced suicidality. Where a trial may not be able to show that a drug causes suicidality, the exercise of clinical judgement within a trial can do so. Patients can also distinguish between the beneficial effect of a drug and the effect of that benefit on overall outcome, as for instance when they make it clear that an SSRI is producing a useful emotional numbing but that this not leading to a recovery from their clinical syndrome. This information is important if we want to make a decision as to whether to introduce another drug with a different mode of action into the mix or whether we want to stop the original drug and start another.

The implication for the interpretation of RCT results is important. Is it correct in this case to say in patients who show little clinical improvement on a rating scale that the drug is not working? Benzodiazepines in comparable trial designs have been shown to work in mild to moderate depressions. In the patients who do not show a rating scale benefit but who report being less anxious would we say the benzodiazepine was not working?

As things stand, RCTs are used *de facto* to obscure the specific effects of quite different therapeutic principles. In the case of the antidepressants, very diverse drugs acting selectively on different brain systems end up looking exactly the same in trials using the outcome measures currently adopted. In contrast to these clinical examples, trials of a drug taken by healthy volunteers reveal drug effects, unconfounded by a clinical condition.

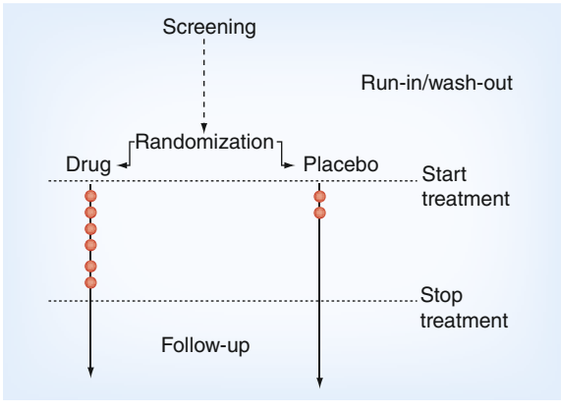

Figure 2. Randomized controlled trials. Selective serotonin reuptake inhibitors in mild-to-moderate depression suicidal acts.

Paroxetine & suicide: actual experiments

In the late 1980s, Eli Lilly undertook a trial of fluoxetine in a group of patients with recurrent brief depressive disorder (RBDD). These patients engage in suicidal acts more often than MDD patients do. In this trial, placebo was sweepingly statistically superior to fluoxetine. The published study was shorn of its key data [26].

In the early 1990s, SmithKline Beecham undertook study 106 of paroxetine in RBDD patients in the same hospital center, possibly with some of the same patients who had been in the fluoxetine trial. This study terminated early. The results were never published. The rate of suicidal acts on paroxetine was threefold higher than on placebo [Unpublished Data]. A few years later, SmithKline undertook study 057 in a similar group of patients. There are multiple datasets in the public domain from this study [27].

In April 2006, in a press release GlaxoSmithKline showed the following data for patients in their paroxetine MDD trials (Table 2). The MDD patients show a significant increase in suicidal act risk on paroxetine.

Table 3 uses a publication released by GlaxoSmithKline in April 2006 covering the suicidal act rate in their depression trials [28]. Despite the fact that studies 106 and 057 do not support using paroxetine for RBDD, the data from these studies when added can surprisingly cause the risk from paroxetine to vanish (Table 3).

One could add 16 more suicidal acts to the paroxetine RBBD column in Table 3, increasing the relative risk of an adverse event on paroxetine to 1.4, raising the combined paroxetine suicidal act number to 59, and still get the same apparently protec-

Table 2. Suicidal acts in major depressive disorder trials.

MDD trials	Paroxetine	Placebo	Relative risk
Suicidal acts/patients	11/2943	0/1671	Inf (95% CI: 1.3–inf)

Inf: Infinity; MDD: Major depressive disorder.

tive outcome overall. This paradoxical outcome is not a quirk of antidepressants and suicide. It is predictable and obvious. Knowing what a drug can do, you can often design studies that use a problem the drug causes to hide that same problem. In this particular instance, there is clearly a poor meta-analytic technique involved, but the example also points to a deeper problem. Something comparable can happen by accident in clinical trials carried out in every illness we don't fully understand – from back pain to Parkinson's disease.

Just as diverse sets of pain patients or Parkinson's disease patients can meet criteria for their respective illnesses, so also RBDD patients will often meet criteria for MDD. Provided there is more than one RBDD patient entered into MDD trials randomization will ensure these patients will hide the effect of an SSRI on suicidal acts, just as back pains of one type will mask what may be beneficial treatment effects on another type of pain. Similar outcomes are almost certain for at least some of the many effects of every drug in irreducibly heterogeneous clinical populations. The only way to overcome this bias and get the kind of result that would allow Fisher to agree demonstrates that we know what we are doing is in fact to understand what we are doing – that is to have a relatively complete understanding of the clinical condition we are treating and of the effects of the drug we are using. But at this point, RCTs take on the quality of a demonstration.

Since TESI and behavior is life threatening, the physician must weigh the benefit of medication alongside the significant risk. There have been no clinical/patient indicators of TESI risk with which to guide physicians in medication management to date.

Pharmagnosia

Unlike fertilizers used in agriculture, drugs used in medicine have 100 or more effects. When we design an experiment employing randomization to manage the unidentified unknowns for one of these effects, we risk generating unawareness about ignorance regarding most of what the drug does. This is as true for statins, antibiotics and other drugs as it is for antidepressants.

In the case of the SSRI antidepressants, the design of trials was dictated by business considerations. This meant powering studies to produce a statistically significant outcome on a series of rating scales that measure changes in clinical syndromes in a very rough fashion. The effect on these scales was designated as the primary effect, even though these drugs have a much more convincing effect on sexual functioning than on any clinical syndrome. However, because of the focus on Hamilton Depression Rating (HAMD) scores, data on sexual functioning and the other many effects these drugs have was either not collected or was poorly collected, allowing companies to claim afterwards that less than 5% of those taking SSRIs had a disturbance of sexual functioning on treatment. Trial design in the case of the SSRIs has inevitably generated an agnosia of most of the effects of these drugs. While some problems were inevitable, this agnosia has been compounded by a rhetoric that gives the impression that since these drugs have been through RCTs, most of what needs to be known about them is known.

The original design of RCTs was to test the null hypothesis. Strictly speaking, this only allows the conclusions that either this drug has not been shown to be of benefit or that we cannot say it is of no benefit. Sticking to conclusions like these would make it clear that trials reveal relatively little of what a drug in fact does, and would make pharmagnosia less likely.

Interim summary: adverse effects of RCTs

All RCTs do harm. Some do good as well and, of these, some do more good than harm at a reasonable cost.

Table 3 Suicidal acts in major depressive disorder and recurrent brief depressive disorder trials.

Trials	Paroxetine	Placebo	Relative risk
MDD trials acts/patients (n)	11/2943	0/1671	Inf (1.3, inf)
RBDD trials acts/patients (n)	32/147	35/151	0.9
Combined acts/patients (n)	43/3090	35/1822	0.7

Inf: Infinity; MDD: Major depressive disorder; RBDD: Recurrent brief depressive disorder.

RCTs can do harm in many different ways. First, when the null hypothesis is confirmed, they risk rejecting as without benefit, some treatment that in fact has benefits.

Second, when the null hypothesis is rejected, if a good medicine is a chemical that comes with good information, then a reliance on RCTs only as a means of evaluating medicines will lead to pharmagnosia and a consequent degradation of our medical arsenal. Pharmagnosia is worth risking when there are grounds to think a claimed benefit does not hold water and if an ineffective treatment is widely used, vulnerable patients are likely to be harmed. This is close to the original use of RCTs, which was a method to deal with the claims of hucksters and charlatans.

Third, there is a public health risk. There are far more MDD than RBDD patients in the example above. Knowing this, it becomes clear that the results above suggest paroxetine harms more people than not. But not understanding the clinical condition and just going on the data, public health officials or regulators are likely to maintain that the treatment has a favorable risk:benefit ratio. By this they mean that paroxetine produces more benefits than harms on a population basis, although once the data are understood, it suggests the treatment is likely to harm more patients than it benefits.

It is a moot point just what proportion of such pronouncements by regulators is lacking in appropriate supportive evidence.

Fourth, one of the consequences of the primacy now put on RCTs, and their incorporation into the regulatory apparatus, as the gateway through which drugs are licensed, is that pharmaceutical companies have been handed the perfect way to market drugs analogous to snake oil. If an effect that might be construed as in some way beneficial can be demonstrated, companies are able to market their product and all but make its prescribing compulsory, as it appears almost unethical not to prescribe agents whose risk–benefit ratio is favorable. Moreover, we have no obvious brake on the risk that patients may be put on ten or more drugs, all on the basis that each has been shown to 'work'. There are of course no RCT data for any of these combinations.

Fifth, in order to run a treatment trial in clinical conditions, where mortality or return to work are not available as outcomes, rating scales or other surrogate outcomes are used. In the case of female sexual dysfunction, the rating scales used include items such as clitoral sensitivity. In this case, the clinical trial process forces women to attend to aspects of functioning that they would not ordinarily attend to. Were any of these treatments ever to come on the market, the marketing of the effects of drugs on clitoral sensitivity risk seriously affecting the experience of and understanding of lovemaking.

The clinical encounter is a relationship, and good care involves sensitivity to the dynamics of the relationship. In a rather similar way to the way RCTs risk changing love-making, clinical trials have in fact affected the clinical encounter. The doctor has been numbed to the reality of the patient who has become increasingly invisible. Clinical encounters have become an industrial process, like agriculture, that aims at implementing impersonal algorithms and guidelines. The effects are in fact worse than in agriculture, in that in medicine the guidelines are based on miscoded data in ghost-written publications from trials not designed to detect or incapable of detecting many of the significant effects of treatment. These issues do not just apply to adverse effects. We will reduce the rate of discovery of new drugs if doctors and patients are trained to ignore the full range of effects a treatment may have.

Future perspective: genetic testing may be able identify risk to the patient that is uninformed by RCTs

Suicidal ideation and behavior is perhaps the most serious of adverse drug responses. In spite of the challenges associated with RCTs discussed earlier, we are hopeful that advances in the field of genetics and the practice of personalized medicine will provide a solution quickly to this most serious of the many challenges by restoring the patient to the center of the medical stage. The following view of the genetic discoveries that are reported to be associated with antidepressant-induced suicidal ideation provides a perspective of an emerging area that may help investigators extract better knowledge from RCT data and that may lead to predictive tests that assess relative risk for each patient of developing this most serious of adverse drug responses.

To date, researchers have reported on at least nine studies associating nearly 100 genetic markers with TESI [29–37]. A review of articles published prior to 2010 [38] underscored that in 3231 unique subjects, 424 (13.1%) showed increases in suicidal ideation, eight (0.25%) attempted suicide and four (0.12%) completed suicide.

Three of the published reports employed genome-wide association studies. In the STAR*D trial, variants within the genetic loci encoding papilin (*PAPLN*) and the IL-28α receptor (*IL28RA*) were discovered [32]. In the GENDEP study, a genetic marker in the vicinity of the *GDA* gene was associated with emergent or worsening of suicidal ideation [33]. Finally, in the MARS sample, 79 additional markers were identified [37].

The remaining studies focused on a genetic assessment of genes suspected to operate in neurological pathways. Genetic associations with TESI were found in the STAR*D cohort, with genetic markers within

Executive summary

The origin of randomized trials

- Randomized controlled trials (RTCs) date back to the 1920s when RA Fisher studied the effect of fertilizers on seed growth.
- Bradford Hill used randomization for the first time in 1947 in a study of the efficacy of streptomycin in tuberculosis.

Early doubts about RCTs

- A change in 1962 to the Food and Drugs Act required the demonstration of 'effectiveness' of new compounds through placebo-controlled RCTs.
- Most drugs introduced in the 1950s, without RCTs, remain more effective than treatments that have come to market through RCTs.
- Short term placebo-controlled RCTs rarely demonstrate effectiveness.
- While RCTs have an important place in therapeutics, if used indiscriminately, their adverse effects may outweigh their benefits.

Mediculture or medicine?

- Randomization undertaken to manage unidentified unknowns requires a focus on one effect of treatment, which generates an unawareness of ignorance regarding other effects.

Placebo effect

- Placebo controls introduce a systematic bias so that an active treatment simply needs to beat the placebo on some dimension to be judged as working, which can be achieved for weaker agents by powering trials accordingly.

Antidepressants & suicide: a thought experiment

- A subgroup of patients (6–13%) develop treatment-emergent suicidal ideation and behavior in the early weeks following the initiation of therapy, dosage change or upon withdrawal from antidepressants.
- The initial RCTs failed to establish that antidepressants could cause suicide.

Implications for RCTs

- RCTs may obscure rather than reveal treatment effects.
- If the trial is not designed to look at an issue, as in this case of antidepressants and suicide, the RCTs results will not show cause and effect.
- Even if a study is designed to examine antidepressants and suicide, RCTs cannot predict how often in clinical practice the drugs might trigger suicidality.
- In RCTs of antidepressants, there have been more dead bodies in the active treatment arms of trials than on placebo.
- Where a trial may not be able to show a drug causes suicidality, the exercise of clinical judgement within a trial can do so.

Paroxetine & suicide: actual experiments

- Despite the fact that two different studies do not support using paroxetine for recurrent brief depressive disorder (RBDD) or major depressive disorder (MDD), the data from these studies when added can surprisingly cause the risk from paroxetine to vanish.
- When RBDD patients are entered into MDD trials, randomization will ensure these patients will hide the effect of an selective serotonin reuptake inhibitor (SSRI) on suicidal acts.

Pharmagnosia

- Because of the focus on HAMD scores, data on sexual functioning and many other drug effects was either not collected or was poorly collected, allowing companies to claim that fewer than 5% of those taking SSRIs had a disturbance of sexual functioning on treatment.
- Trial design in the case of the SSRIs has inevitably generated an agnosia of most of the effects of these drugs.

Adverse effects of RCTs

- If an effect that might be construed as in some way beneficial can be demonstrated, drug companies are able to market their product and all but make its prescribing compulsory, as it appears almost unethical not to prescribe agents whose risk–benefit ratio is favorable.
- We have no obvious way to stop the risk that patients will be put on ten or more drugs, all on the basis that each has been shown to 'work', even though there are no RCT data for any of these combinations.

Future perspective

- Advances in the field of genetics and the practice of personalized medicine may provide solutions for the shortcomings of RCTs, particularly in the advent of antidepressant-induced suicidal ideation and behavior.
- Researchers have reported nine studies associating nearly 100 genetic markers with treatment-emergent suicidal ideation.
- A pharmacogenomic test may one day be available that can give physicians the opportunity to weigh the benefit of antidepressant treatment against a predicted risk and employ personalized treatment accordingly.

the genes encoding the glutamate receptors GRIK 2 and GRIA 3 [29], and in CREB1 [30]. In the GENDEP study, genes encoding BDNF, NTRK2 and ADRA2A were reported [33]. In the TORDIA study, markers within *FKBP5* were found [34]; in a sample of depressed outpatients, *FDBP5* and *MDR/TAP*, and *ABCB1* were identified [36].

Work remains to be done by these groups to follow-up their initial discoveries linking TESI to specific genetic markers by confirming their results in different cohorts. Partial replication has been reported for two glutamate receptor genes in the MARS project [31]. Of the 79 markers found in the MARS project, 14 markers were identified in a replication sample, and a discriminate analysis of the 79 markers revealed at 91% probability to classify TESI versus non-TESI correctly in the replication sample [37].

The reported results are promising and provide hope that genetic analysis of each RCT participant may provide more meaningful safety information for drug developers, regulators and physicians. They also suggest that a pharmacogenomic test may one day be available that can give physicians the opportunity to not only be aware of a particular patient's risk, but be able to weigh the benefit of antidepressant treatment against the predicted risk and against the many psychosocial factors affecting the individual patient. If heightened risk is reported by the pharmacogenomic test, the physician may choose from many treatment options: employ alternate pharmaceutical therapy; institute heightened vigilance, engage in frequent and fail-safe patient monitoring, refer to psychosocial therapy, engage the family and close friends in supervision, utilize compelling interactive media for young persons, and provide hospitalization in the case of very high patient risk until the patient has stabilized.

Financial & competing interests disclosure

D Healy is the CEO of RxISK.org, a patient adverse event reporting company. He has been and continues to be an expert witness in legal cases involving antidepressants and suicide. P Tolias and K Bechthold are cofounders of Sundance Dx, a diagnostic company developing a pharmacogenomic test for risk of medication-induced suicidal ideation. K Bechthold is the CEO of Sundance Dx., a diagnostic company developing a pharmacogenomic test for risk of medication-induced suicidal ideation. K Bechthold is the CEO of Sundance Dx. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:

- • of interest
- •• of considerable interest

1 Fisher RA. *The Design of Experiments* (9th Edition). Macmillan, London, UK (1935).

2 Savage LJ. On rereading RA Fisher. *Ann. Statistics* 4, 441–500 (1976).

3 Marks HM. The Progress of Experiment. Cambridge University Press, Cambridge, UK (2000).

4 Healy D. The antidepressant tale: figures signifying nothing? *Adv. Psychiatr. Treat.* 12, 320–328 (2006).

5 Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *Br. Med. J.* 2(4582), 769–782 (1948).

6 Toth B. Clinical trials in British medicine 1858–1948, with special reference to the development of the randomised controlled trial. Dissertation, University of Bristol, Bristol, UK (1998).

- • **Good review on the history of randomized controlled trials.**

7 Hill AB. Reflections on the controlled trial. *Ann. Rheum. Dis.* 25, 107–133 (1966).

8 The Empire of Chance. Gigerenzer G (Ed.). Cambridge University Press, Cambridge, UK (1990).

9 Healy D. *Pharmageddon.* University of California Press, CA, USA (2012).

- • **Comprehensive argument against the pharmaceuticalization of medicine, which tackles problems in healthcare that are leading to a growing number of deaths and disabilities.**

10 Lasagna L. Thalidomide – a new non-barbiturate sleep-inducing drug. *J. Chronic Dis.* 11, 627–631 (1960).

11 Hohn R, Gross GM, Gross M, Lasagna L. A double-blind comparison of placebo and imipramine in the treatment of depressed patients in a state hospital. *J. Psychiatr. Res.*1, 76–91 (1961).

12 Raffle A, Gray M. *Screening: Evidence and Practice*. Oxford University Press, Oxford, UK (2007).

13 Pratt LA, Brody DJ, Gu Q. Antidepressant Use in Persons Aged 12 and Over: United States, 2005–2008. NCHS Data Brief, No 76. National Center for Health Statistics, MD, USA (2011).

14 Bostwick JM, Pankratz VS. Affective disorders and suicide risk: a reexamination. *Am. J. Psychiatry* 157, 1925–1932 (2002).

- • **Early report linking suicidal ideation to antidepressant drug use in some of the treated patients.**

15 Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* 62, 593–602 (2005).

16 Perlis RH, Beasley CM , Wines JD Jr *et al.* Treatment-associated suicidal ideation and adverse effects in an open,

multicenter trial of fluoxetine for major depressive episodes. *Psychother. Psychosom.* 76, 40–46 (2007).

17 Seemuller F, Riedel M, Obermeier M ,The controversial link between antidepressants and suicidality risks in adults: data from a naturalistic study on a large sample of in-patients with a major depressive episode. *Int. J. Neuropsychopharmacol.* 12, 181–189 (2009).

18 Davies EB. Proceedings of the Symposium held at Cambridge, 22–26 September 1959. Clark D *et al.* 86–89 Cambridge University Press, Cambridge, UK (1962).

19 Healy D. *The Antidepressant Era*. Harvard University Press, MA, USA (1998).

20 Hagnell O, Lanke J, Rorsmart B. Suicide rates in the Lundby study: mental illness as a risk factor for suicide. *Neuropsychobiology* 7, 248–253 (1981).

21 Stone M, Jones L. Clinical review: relationship between antidepressant drugs and adult suicidality. US FDA, Washington, DC, USA (2006). www.fda.gov/ohrms/dockets/ac/06/briefing/2006--4272b1-index.htm

22 Teicher MH, Glod C, Cole JO. Suicidal preoccupation during fluoxetine treatment. *Am. J. Psychiatry* 147, 1380–1381 (1990).

23 Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int. J. Epidemiol.* 15, 412–418 (1988).

24 Lanes SF. Biological interpretation of relative risk. *Drug Safety* 21, 75–79 (1999).

25 Rothman KJ, Greenland S, Lash T. *Modern Epidemiology*. Lippincott, Williams & Wilkins, PA, USA (2008).

26 Montgomery DB, Roberts A, Green M, Bullock T, Baldwin D, Montgomery SA. Lack of efficacy of fluoxetine in recurrent brief depression and suicidal attempts. *Eur. Arch. Psych. Clin. Neurosci.* 244, 211–215 (1994).

27 der Mast RC Van, Hengeveld MW, Tuyl JP, Zwindermann AH, Van Kempen GM. Reduction by paroxetine of suicidal behavior in patients with repeated suicide attempts but not major depression. *Am. J. Psychiatry* 155, 543–547 (1998).

28 GlaxoSmithKline. Paroxetine adult suicidality analysis. GlaxoSmithKline (2006). www.gsk.com/media/paroxetine/briefing_doc.pdf

29 Laje G, Paddock S, Manji H. Genetic markers of suicidal ideation during citalopram treatment of major depression. *Am. J. Psychiatry* 164(10), 1530–1538 (2007).

30 Perlis RH, Purcell S, Fava M *et al.* Association between treatment-emergent suicidal ideation with citalopram and polymorphisms near cyclic adenosine monophosphate response element binding protein in the STAR*D study. *Arch. Gen. Psychiatry* 64(6), 689–697 (2007).

31 Menke A, Lucae S, Kloiber S *et al.* Genetic markers within glutamate receptors associated with antidepressant treatment-emergent suicidal ideation. *Am. J. Psychiatry* 165(7), 917–918 (2008).

32 Laje G, Allen AS, Akula N, Manji H, Rush A John, McMahon FJ. Genome-wide association study of suicidal ideation emerging during citalopram treatment of depressed outpatients. *Pharmacogenet. Genomics* 19(9), 666–667 (2009).

33 Perroud N, Aitchison KJ, Uher R *et al.* Genetic predictors of increase in suicidal ideation during antidepressant treatment in the GENDEP project. *Neuropsychopharmacology* 34(12), 2517–2528 (2009).

34 Brent D, Melhem N, Turecki G. Pharmacogenomics of suicidal events. *Pharmacogenomics* 11(6), 793–807 (2010).

•• **Pharmacogenetic studies conducted up to 2010 that identified specific mutations associated with drug-induced suicidal thoughts.**

35 Lucire Y, Crotty C. Antidepressant-induced akathisia-related homicides associated with diminishing mutations in metabolizing genes of the CYP450 family. *Pharmgenomics Pers. Med.* 4, 65–81 (2011).

36 Perroud N, Uher R, Ng MY *et al.* Genome-wide association study of increasing suicidal ideation during antidepressant treatment in the GENDEP project. *Pharmacogenomics J.* 12(1), 68–77 (2012).

37 Menke A, Domschke K, Czamara D *et al.* Genome-wide association study of antidepressant treatment-emergent suicidal ideation. *Neuropsychopharmacology* 37(3), 797–807 (2012).

38 Brent D, Melhem N , Ferrell R. Association of FKBP5 polymorphisms with suicidal events in the treatment of resistant depression in adolescents (TORDIA) study. *Am. J. Psychiatry* 167(2), 190–197 (2010).