BJPsych | The British Journal of Psychiatry (2018)
212, 389–390. doi: 10.1192/bjp.2018.95

Kaleidoscope

Derek K. Tracy, Dan W. Joyce,
Sukhwinder S. Shergill

We start with two excellent editorials with cautionary tales for both psychological and pharmacological studies. Comprehensive reporting of side-effects and financial conflict of interests (COI) beleaguer pharmacological work, and were the main criticisms levelled against the recent network meta-analysis by Cipriani *et al*.[1] But are these exclusively issues for medication trials? Crawford noted how psychological therapies can also cause harm,[2] but this is not always reported, and Cristea & Ioannidis comment[3] on the lack of reporting guidance on COIs in psychosocial interventions – including in the International Committee of Medical Journal Editors' standard disclosure form. Further, there is an issue that psychosocial COIs may not map particularly well onto the pharmaceutical model that typically looks at patents and corporate research investment, instead having potential COIs of co-owning a for-profit company or spin-off ventures, royalties from intervention materials, academic funding and so forth. More challenging examples discussed include instances of charitable organisations – associated with specific interventions and academics – lobbying policymakers for their wider dissemination. There is currently little guidance for professionals. Step one is acknowledging there is a problem, that it has parallels but differs from pharmaceutical work and building a more relevant taxonomy of relevant COIs.

Blessed are the national registry studies. Well, obviously it is not meant to be taken literally; it refers to any large data-sets. At least that has been the take from recent Kaleidoscope columns that have lauded the results from large pharmacological research of national registers that overcome the small *n*, short-term, selective cohort problems of the typical randomised controlled trial (RCT). However, Leucht & Davis remind us that these trials carry their own set of limitations.[4] They cannot control for 'confounding by indication' – namely we cannot unpick the non-random decision-making that drives the professionals' clinical actions. For example, patients who are less well might be treated more often with drug X – as clinicians perceive it to be more efficacious – but its use in this poorer outcome cohort may lead a registry study to show it as being less effective. Similarly, another drug Y may not be used in patients with relevant medical conditions, which would result in that medication 'producing' fewer side-effects than expected. Furthermore, replication is impossible and as clinical practice varies from country to country, comparison of outcomes between different jurisdictions becomes hugely problematic. This is not to say that these limitations undermine such work – like all research it has a place and a value. These studies may now require better standardisation of methods and reporting, but not quite 'Registries eunt domus' yet.

More psychological–pharmacological debate: when recovered from depression, would you plump for antidepressant continuation or replacing it with preventative cognitive therapy (PCT)? Of course, the real world is not always that binary, but equally, many patients will request guidance on medication discontinuation. Bockting and colleagues test this, reporting on a single-blind RCT[5] of individuals on such medication (for at least 6 months) but in remission after a history of at least two such episodes. About 100 patients were randomised to each of three arms: medication continuation, medication discontinuation and PCT or receiving both antidepressants and PCT. Neither medication continuation nor medication discontinuation and PCT was more effective than the other in terms of relapse or recurrence of depression in the subsequent 2 years, but the combination of both led to a 41% reduction in relative risk. The challenge is how to ensure that we have a service that can add the eight sessions of PCT to antidepressant treatment upon recovery to optimise outcomes.

All right, but apart from treating depression, anxiety, post-traumatic stress disorder, eating disorders and obsessive–compulsive disorder, what have antidepressants ever done for us? Although of course, in depression they are only *really* of use in moderate-severe episodes, right? The issue around lack of antidepressant efficacy in 'mild' depression may be confounded by a large primary care cohort that can include patients with social and emotional distress but who do not actually have depression. Studies have shown that up to three in four patients treated with antidepressants may not actually have depression, although conversely, three in four with clinical depression get no treatment at all. Furukawa *et al*[6] throw their hat in the ring. Evaluating individual-participant-level data ($n = 2464$) from 11 placebo-controlled double-blind Japanese RCTs of what they label 'new generation' antidepressants (duloxetine, escitalopram, mirtazapine, paroxetine, bupropion), they modelled symptom modification through what they describe as six increasingly complex competing mixed-effects models for repeated measures. Illness severity did not predict acute-phase outcome, and those with mild depression benefitted as much as those with severe variants. This is particularly interesting, and unexpected, as it runs contrary to recent data-sets showing severity as a strong modifier of outcome in psychosis and bipolar affective disorder, something the authors speculate may be the result of the greater heterogeneity seen in depression. They also note that a similar type of analysis found that depression severity did not predict cognitive–behavioural therapy effectiveness. However, an issue with these results is that of course the samples are highly selected and do not represent typical primary care cohorts. So, sit back, the debate might not be settled just yet.

Maybe the *real* common enemy of psychiatrists and psychologists are computers and artificial intelligence (AI): the Judean People's Front of mental health (splitters). Recent work described a machine learning tool for classifying people into three groups ('suicidal', 'mentally ill but not suicidal' and 'controls'), 'with up to 85% accuracy'. An editorial in *Nature*[7] explores the 'hype-fail cycle' in AI, where even limited successes of certain methods are promoted to the exclusion of competitor methods, and where one negative ('epic fail') can lead to the sudden death of research endeavours. Historically, the hype-fail cycle is exemplified in the so-called 'AI winter' of the 1970s and 1980s where there was an almost complete cessation of research in artificial neural networks – the same networks that are today the 'go to' machine learning technique manifest as deep-learning networks. That hiatus is attributed to Minsky & Papert's book *Perceptrons* that demonstrated that single-layered neural networks (the precursor to modern multilayered neural networks of which deep-learning networks are descendants) could implement associative learning, but only solve simple – so-called 'linearly separable' – classification problems.[8] How to deploy a trained AI method in clinical settings?[7] There appears little evidence of an 'epic fail' moment for current machine learning, but there are warnings related to their deployment and particularly in clinical decisions.

Judea Pearl argues[9] that proponents have been looking on the bright side of life, and even the most state-of-the-art machine learning techniques learn associations that can be inferred from the data alone. Pearl contends that although these approaches can capture complex relationships we require a facility for reasoning about intervention (answering queries of the kind 'what if I do X?') and counter-factuals ('Was it X that caused Y?' and 'What if I had

acted differently?') to make these systems fully capable. These are the kinds of questions that expert epidemiologists and trialists attempt to answer from observational and controlled trial data. We need a move beyond modelling associations (however complex) and classification to embedding these models into decision-making applications that use relevant domain expertise and typical patterns of clinical inference (for example, 'what is the risk associated with discharging a patient who may complete suicide in the next 7 days given management plan A or B ?'). To see why, imagine deploying a suicide classification system with 85% accuracy in a clinical environment – we can expect 15% of individuals will be misclassified. If these misclassified ones are always false positives – predicted to be suicidal but are not – then the classifier is appropriately risk-averse (setting aside the costs associated with overinvestigating). If, however, any of these 15% are in fact suicidal, but are incorrectly classified as 'mentally ill but not suicidal' or 'control' then we are in trouble – these false negatives could lead to loss of life. This asymmetry – the cost or utility of correct *v.* incorrect classification – is rarely captured both in the way classifier algorithms are penalised for errors as they learn from the data, or, in the validation of the trained classifier's accuracy reported in the literature. Classification algorithms most often assume symmetrical cost or utility, so the penalty is the same if the classifier proposes a suicidal individual is a control and vice versa. The biostatistician Frank Harell has blogged on this issue: classification is a forced, discrete and mutually exclusive choice among classes (for example suicidal or not suicidal) induced by a threshold on an underlying continuous variable. If, for example, the probability of a being a 'suicidal case' changes from 0.001 above the threshold for 'suicidal' to 0.001 below the threshold then the accuracy measure (i.e. the true/false positives) changes by $1/n$ (where n is the number of cases being classified) – even when the underlying continuous probability of being suicidal has shifted only a vanishingly small amount that is unlikely to translate to a significant shift in clinical risk. As Pearl concludes, 'Data science is only as much of a science as it facilitates the interpretation of data – a two-body problem, connecting data to reality. Data alone are hardly a science, regardless how big they get and how skilfully they are manipulated'.[9] We need to translate and evaluate machine learning/AI methods in the evidence-based framework we would apply to any other medical device.

Attention-deficit hyperactivity disorder (ADHD) attracts debate, from diagnostic rates to stimulant medication in children. What of the *misuse* of stimulant medications? Weight loss, obtaining a 'high' and nootropic cognitive enhancement are the main potential reasons for stimulant misuse. It has been a concern for some time, but we have lacked good data, especially in adult populations and interestingly enough, US prescriptions for adult ADHD have now surpassed their use in adolescents. Compton and colleagues report[10] on a nationally representative US population survey of over 100 000 adults. A total of 6.6% had used prescription stimulants over the past year, with about 2% doing so without a medical reason. Misusers were more likely to be male, have other substance misuse problems and come from lower socioeconomic groups. The most common rationale (56.3%) for misuse was to increase alertness or improve concentration, with getting a 'high' and weight loss accounting for about 15% and 4%, respectively. Over half the time these were obtained for free from friends or relatives, which raises the second challenge that some are not misusing the drugs themselves, but are passing at least some of their prescribed medications to others to do so. Beyond the obvious ethical problems, these medications can cause harm, including insomnia, tachycardia, a range of mental health problems and indeed death. Turning statistics in people's lives, an estimated five million Americans are misusing stimulant medications, 400 000 to the point of having a use-disorder problem.

Finally, 'Prevent' has been one of the most controversial and vehemently contested UK governmental programmes. Do mental health services *really* have a role identifying those at risk of radicalisation? Campelo *et al*[11] review if there are psychological and social profiles that predict this in European adolescents. They cluster their findings into a three-level model: individual risk factors, the microenvironment and societal issues. The first included early experiences of abandonment, personal uncertainty and perceptions of injustice; microenvironmental aspects were family dysfunction and friendship with those already radicalised; the wider issues were geopolitical trends. It was noted how recruiters could tap into these by isolating and dehumanising potential subjects, and offering them what seemed a new societal model. Assimilating the evidence, the authors come out in favour of mental health professionals having roles in identifying, understanding and – where appropriate – treating such issues. Educative support, therapeutic groups, family therapy and appropriate mentoring are proposed, although they accept that the evidence base remains weak and how, problematically, their factors are common to much adolescent psychopathology. Durkheim's concept of 'anomie' – the mismatch between individual or in-group standards and those of wider society – is invoked. Obvious concerns relate to the production of enormously elevated rates of false positives, not to mention the risks of increasing alienation and disenfranchisement of specific communities inevitably profiled and targeted, as has been observed with respect to many in Muslim communities.

References

1 Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. 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**: 1357–66.

2 Crawford MJ, Thana L, Farquharson L, Palmer L. Patient experience of negative effects of psychological treatment: results of a national survey. *Br J Psychiatry* 2016; **208**: 260–5.

3 Cristea I-A, Ioannidis JPA. Improving disclosure of financial conflicts of interest for research on psychosocial interventions. *JAMA Psychiatry* 11 April 2018 (doi:10.1001/jamapsychiatry.2018.0382).

4 Leucht S, Davis JM. Enthusiasm and skepticism about using national registers to analyze psychotropic drug outcomes. *JAMA Psychiatry* 2018; **75**:314–5.

5 Bockting CLH, Klein NS, Elgersma HJ, van Rijsbergen GD, Slofstra C, Ormel J, et al. Effectiveness of preventive cognitive therapy while tapering antidepressants versus maintenance antidepressant treatment versus their combination in prevention of depressive relapse or recurrence (DRD study): a three-group, multicentre, randomised controlled trial. *Lancet Psychiatry* 2018; **5**: 401–10.

6 Furukawa TA, Maruo K, Noma H, Tanaka S, Imai H, Shinohara K, et al. Initial severity of major depression and efficacy of new generation antidepressants: individual-participant data meta-analysis. *Acta Psychiatr Scand* 3 April 2018 (doi: 10.1111/acps.12886).

7 AI diagnostics need attention. *Nature* 2018; **555**: 285–6.

8 Minsky M, Papert SA. *Perceptrons: An Introduction to Computational Geometry*. MIT, 1969.

9 Pearl J. Theoretical impediments to machine learning with seven sparks from the causal revolution. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, California, 5–9 February 2018*. (https://arxiv.org/abs/1801.04016).

10 Compton WM, Han B, Blanco C, Johnson K, Jones CM. Prevalence and correlates of prescription stimulant use, misuse, use disorders, and motivations for misuse among adults in the USA. *Am J Psychiatry* 16 April 2018 (doi: 10.1176/appi.ajp.2018.17091048).

11 Campelo N, Oppetit A, Neau F, Cohen D, Bronsard G. Who are the European youths willing to engage in radicalisation? A multidisciplinary review of their psychological and social profiles. *Eur Psychiatry* 2018; **52**: 1–14.

https://doi.org/10.1192/bjp.2018.95 Published online by Cambridge University Press