

## Kaleidoscope

Derek K. Tracy, Dan W. Joyce,  
Sukhwinder S. Shergill**‘One in four’ will have a ‘mental health problem’ at some point in their lives; in one in six of the population this will be depression.**

Estimates to be sure, but conveying important health information to a wider audience. However, do we have sufficiently accurate epidemiological data for common mental illnesses? Levis *et al* challenge us reporting on 69 recent meta-analyses that defined a pooled prevalence of depression.<sup>1</sup> Only 10% of these studies based their data exclusively on research that used diagnostic interviews; the remainder used screening or rating tools or a combination of methods. Similarly, in the 2094 underpinning primary studies, only 13% using validated diagnostic interviews, and the majority based their findings on screening tools, unstructured interviews, medical records and so forth. Screening tools have an important role, but their typically high false-positive nature makes them inflators of true underlying data; in this work 14% higher than those solely using diagnostic interviews. This matters given its impact on policy and resource. The authors cite the example of a recent news-making review reporting that 27% of medical students had ‘depression or depressive symptoms’, but the one study reviewed that relied on actual diagnostic interviews determined a prevalence of 9%, which is equivalent to that seen in age-matched cohorts.

This links with the controversy – and tabloid headlines – that ‘too many’ people are on antidepressants. Hafferty *et al* explored prevalence, incidence, adherence and predictors of antidepressant use between 2009 and 2016 in the Generation Scotland cohort of over 11 000 individuals.<sup>2</sup> Just under 30% were prescribed at least one antidepressant across a 5-year period, with a 36% increase in annual prevalence from 2010 to 2016. By 2016, an estimated 17.3% of the adult population was using an antidepressant. Selective serotonin reuptake inhibitors remain the most commonly used, but there has been a growth in serotonin–noradrenaline reuptake inhibitors and mirtazapine. Importantly, most episodes of care were for more than 9 months (half being for over 15 months), in line with guidelines, and, pleasingly, medication adherence was relatively high at just under 70%. Crucially, the incidence remained stable across time, at about 2.4% per year; what this tells us is that the ‘growth’ is generated secondary to increased (and appropriate) longer-term use in those taking the medication, not that ‘everybody’ is being put on antidepressants.

**Predicting suicide is a core goal in mental health; we recognise clinical and sociodemographic risk factors, but equally that these are very common and non-specific.**

Given the increasingly large healthcare data-sets we have, and advances in statistical modelling, is the positive-predictive-value tide turning? Belsher *et al* evaluated the diagnostic accuracy of multiple suicide prediction models, and simulated the impact of implementing them at a population level.<sup>3</sup> They identified 17 cohort studies suitable for inclusion, covering 64 unique prediction models, five countries and 14 million participants. So, no lack of data and the authors note the research quality was high and classification accuracy was good. However... the predictive validity for subsequent mortality was ‘extremely low (<0.01 in most models)’, suggesting little use applying them prospectively to large populations. To contextualise this, the authors note that assuming a standard suicide mortality rate of 200 per 1 000 000, a 95th risk percentile classification model would identify 58 true positives – and 49 942 false positives. Their conclusion is rather damning, highlighting that ‘their accuracy of predicting a future event is near to 0’.

**Which takes us nicely to the wider field of predictive analytics proposed to solve the big challenges facing humanity.**

There is a widespread enthusiasm for technological solutions to the outstanding issues in medicine; initially focused on removing from clinicians tedious tasks that could be done more efficiently through technology, and more recently focused on improving service provision. This improvement can be efficiency in requiring less clinical input, or in superior diagnostic or prognostic performance. An example might be an app or tool helping clinicians make decisions. One aspect that is overlooked in the general drive to encourage these solutions is the ‘tabloid test’: when one of these tools gets it wrong (for example for a diagnosis or prognosis), who is to blame for the failure? If it is the clinician, one can envisage very few doctors wanting to buy into these tools; if it is the organisation who built the tool (or a hospital using it), one can imagine a need for regulation similar to pharmaceuticals or medical devices is required. In the UK, the Medicines and Healthcare Regulatory Agency have already issued guidance on such software applications offering to ‘tell you that you have a medical condition or disease or give you an individual percentage risk score of having one’. Broadly speaking, statutory frameworks similar to those for replacement joints and new pharmaceuticals apply equally to predictive analytics tools because of how they are deployed.

Given the increase in the volume of technological solutions, especially allied to artificial intelligence, Parikh *et al* argue that there are few exemplars of these predictive tools being put through their paces prospectively under robust conditions such as in a clinical trial.<sup>4</sup> What is the solution? They propose that the existing Food and Drug Administration (FDA) framework for biomarkers and novel diagnostics can be modified and applied to artificial intelligence-based systems. First, they state that there must be clinically meaningful end-points (similarly to arguments about surrogate end-points in trials) to evaluate the performance of systems. They cite an FDA-approved system called WAVE for identifying hospitalised patients at risk of vital-sign instability; the system signalled instability in patients on average 6.3 h before traditional documentation recorded the same change. When deployed in a hospital (i.e. when clinicians act on the system’s alerts) the pre- and post-deployment performance data showed the change in clinical behaviour was a reduction in a nurse-led time to response from 16 to 7 min. They also propose that analytics tools audit and specify interventions resulting from the tools’ deployment that would improve care – most published systems achieve impressive performance on data in isolation, but it is not clear that they improve care by directing clinicians to act in a certain way. Similarly, under the banner of appropriate benchmarks, the authors identify a central idea that is largely missing in the literature – the current focus on superhuman performance (in a machine-versus-human tournament) should instead be replaced by a machine-plus-human comparator against human-alone. They then move to interoperability citing predictive tools derived from scraping data from electronic health record (EHR) platforms as an example; EHRs are notoriously idiosyncratic to both the hospital and the platform’s developer – how can one be sure that a predictive tool developed on one EHR would have any value when applied to another institutions EHR (unless the inputs and data dictionary necessary to use the predictive tool are well defined with high concordance between different EHRs). Parikh *et al* conclude ‘Many developers may decry overregulation and standardization of a poorly understood field’, but that is the point: they are poorly understood because we have ignored the context of deployment and have yet to establish a culture that shows these tools really work in clinical practice.

**What indicates success in academia? Undertaking research that transforms our knowledge and has an impact on practice, producing highly cited papers or gaining prestigious grants from major**

**funderson?** All of them, we suppose, but how related are these phenomena? There are surprisingly few data looking at the output and success of large grant holders. Stavropoulou *et al* searched UK biomedical authors published between 2006 and 2018 who had over 1000 citations in Scopus, and then looked at whether these individuals had received a grant from the National Institute for Health Research, the Medical Research Council or the Wellcome Trust.<sup>5</sup> Of the 164 such first/last authors still working in UK academic institutions, only 36% currently held (and 48% held over the 2006–2017 period) one of these grants. Conversely, almost 70% of board members of these organisations held an active grant with one of the funders, despite only 1.1% having such a first/last authored highly cited paper profile. Highly cited UK authors are not winning the grants their work would seem to merit, and conversely, what is happening to the output of successful candidates? Of course, like impact factor, one can debate the meaning and value of citation number, and it is understandable that awarding bodies have sitting members who have been successful at and understand the application process, but can the process be improved?

**Following on from this, one rarely sees ‘single author’ papers anymore because science requires teams that aggregate together diverse skillsets.** However, a paper by Wu *et al* shows that it is small teams that disrupt and provide new ideas that echo down generations of research, whereas larger teams tend to be productive in taking current high-impact ideas and developing them further.<sup>6</sup> To describe a continuum between disruption and development Wu *et al* used a previously published measure that varies between –1 (development) and +1 (disruption). Of particular note are two exemplars Wu *et al* use to anchor this idea. In 1987 Bak, Tang and Wiesenfeld published a paper (the BTW model) on self-organised criticality, a principle that explains how the widely observed phenomena of ‘flicker noise’ arises from systems being moved slightly from equilibrium. That paper cites five other works in the literature that it attempts to unify under the BTW model. It has received approximately the same number of citations as a 1995 paper by Davis *et al* that provides empirical results that support the Bose–Einstein condensation proposed in two papers from 1924 and 1925 by (unsurprisingly) Bose and Einstein, respectively. Wu *et al* show that most research that followed from, and cited, the BTW model did not co-cite the references Bak, Tang and Wiesenfeld aimed to draw together in one explanatory framework. In contrast, citations for the empirical work in Davis *et al* tended to co-cite the original Bose and Einstein papers. They argue that the BTW model disrupted (i.e. led to whole branches of new work) whereas the Davis *et al* paper developed, or solved, previously proposed problems. When these two exemplars are plotted among over 25 million Web of Science articles, a distribution of the ‘disruption’ (disruption max value +1 to development max value –1) shows that (as expected) the Davis *et al* paper is close to –0.5, and the BTW model is located at 0.86 with review articles around 0 (neutral, neither disruptive or developmental). But what about team sizes? Using data from Web of Science, patents and GitHub and looking at the number of authors, they found that consistently, as teams

grow from 1 to 50 members, the team’s papers, patents and products (such as software tools) scored in lower percentiles of the disruption measure. They note that teams of ten are more likely to produce a high-impact paper (by citation count) but are more likely to be developmental than disruptive, whereas solo-authored papers are 72% more likely to be in the top 5% of highly disruptive papers. These results appear consistent across disciplines, with notable exceptions being engineering and computer science, where because Web of Science indexes only journal papers, it fails to capture their primary publication forum of conference proceedings. They also note a fascinating trend where work from smaller teams tends to cite earlier and less popular references, compared with their larger-team counterparts who focus on contemporary and highly cited work.

**Finally, Brexit. We write this copy a day after two million marched, and five million signed a petition, for a #peoplesvote on the issue; by the time you read this presumably the outcome of the final negotiations will be known.** We report on the first paper of which we are aware that looked at the cognitive styles of voters in the 2016 referendum to exit the European Union. Zmigrod *et al* looked at the association between voting behaviour and attitudes with performance on a cognitive battery that tested cognitive flexibility.<sup>7</sup> They found that subjective and objective cognitive inflexibility predicted conservatism, nationalism, authoritarianism, as well as support for Brexit and opposition to migrants and the European Union. Their model accounted for about 48% of the variance in support for Brexit. Their data show that it is not just an emotional response to an issue, but also a cognitive style, that informs decision-making. The authors suggest that simplistic political phrases such ‘Take Back Control’ (and presumably, ‘Make America Great Again’) utilised to appeal to nationalist identity appear to work best with those who have less flexible cognitive styles.

## References

- 1 Levis B, Yan XW, Sun Y, Benedetti A, Thombs BD. Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. *BMC Med* 2019; **17**: 65.
- 2 Hafferty JD, Wigmore EW, Howard DM, Adams MJ, Clarke TK, Campbell AI, et al. Pharmaco-epidemiology of antidepressant exposure in a UK cohort record-linkage study. *J Psychopharmacol* 2019; **33**: 482–93.
- 3 Belsher BE, Smolenski DJ, Pruit LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths. A systematic review and simulation. *JAMA Psychiatry* 13 Mar 2019 (doi:10.1001/jamapsychiatry.2019.0174).
- 4 Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. *Science* 2019; **363**: 810–2.
- 5 Stavropoulou C, Somai M, Ionnidis JPA. Most UK scientists who publish extremely highly-cited papers do not secure funding from major public and charity funders: a descriptive analysis. *PLoS One* 2019 **14**: e0211460.
- 6 Wu L, Wang D, Evans JA. Large teams develop and small teams disrupt science and technology. *Nature* 2019; **566**: 378.
- 7 Zmigrod L, Rentfrow P, Robbins TW. Cognitive underpinnings of nationalistic ideology in the context of Brexit. *Proc Natl Acad Sci U S A* 2018; **115**: E4532–40.