BJPsych | The British Journal of Psychiatry (2020)
217, 593–594. doi: 10.1192/bjp.2020.155

CrossMark

Kaleidoscope

Derek K. Tracy, Dan W. Joyce,
Dawn N. Albertson, Sukhwinder S. Shergill

Measurement matters in depression, but 'average global changes' in depression scales can mask important individual symptom improvements. Besides this, there is a more fundamental question of what such scales should explore in the first place. Chevance et al undertook international sampling across 52 countries to determine which outcomes mattered most to patients ($n = 1912$), caregivers ($n = 464$) and professionals ($n = 627$).[1] There is inevitably a selection bias in such work, but the open-ended online answers clustered into domains of: symptoms ($n = 64$) including mental pain and motivation; functioning ($n = 16$), including social isolation; and other domains ($n = 57$) indirectly related to treatment benefits, including diverse areas of intervention safety, improved service access and reducing stigma. Symptom improvement matters, but is often the clinicians' and researchers' primary focus, whereas social functioning can lag a long way behind. The authors note how seldom the issues identified here are included in the most commonly used depression scales in clinical trials; yet they matter to those with the actual condition, and those who care for them. It is fascinating that at the tail end of 2020, this is the first large-scale study on the topic.

Genes matter in depression, but symptom permutations in current diagnostic symptoms mean there are 10 000 potential 'depressions' – is it any wonder progress with genome-wide association studies (GWAS) has been slow? Cai et al note the very many ways depression can appear clinically, and the variations in developmental timings, comorbidities and environmental factors that can contribute to the heterogeneity.[2] GWAS studies have increased in size from about 16 000 individuals in 2013 (producing no hits) to one in 2020 with more than a million individuals producing 223 results. Taking this back to 'the depressions', the question is what the genetic targets *mean* in terms of disease entities. The authors stratify three sources of heterogeneity – operational (the construct definition and diagnostic criteria), manifestation (symptoms, severity, onset, comorbidities), and aetiology (the gene × environment interface). The authors conclude that 'depression' likely reflects several distinct phenotypes that a single construct is unlikely to ever capture, so studying at the diagnostic level might not be optimal. They reason that future genetics work might optimally 'split', looking at 'more granular phenotypes with higher validity and reliability such as individual symptoms' or 'lump' and move beyond depression to transdiagnostic features that cross psychiatric conditions.

Inflammation matters in depression, but it's been hard to specify when and in whom. The issue of causality is particularly vexing in this area: does depression lead to inflammatory changes, or vice versa? A particular problem is that changes in cytokines and other markers, such as interleukins and tumour necrosis factor, are seen in some, but not all. One way around this is to provoke an immune response: lipopolysaccharides (LPS) are bacterial cell membranes that can be used in this way. van Eeden et al tested the longitudinal association between basal and LPS-induced inflammatory markers in individuals with major depressive disorders, across the course of up to 9 years.[3] Participants came from the Netherlands Study of Depression and Anxiety. Both basal and LPS-induced inflammatory markers were strongly associated with 'sickness behaviour' – but not 'non-sickness behaviour' symptoms – as well as sympathetic nervous system arousal at the 9-year time point. 'Sickness behaviour' is a broad umbrella term for a range of evolutionary responses to illnesses that preserve energy, promote healing, and protect the organism and its recovery (think of how you act when you have the flu). In terms of depression, one can conceptualise it overlapping with anhedonia, poor energy, concentration, appetite and libido. This is interesting, as sickness behaviour would fit with a construct of biological, immune-mediated processes. This also links nicely with the aforementioned genetics piece, with inflammation not correlating to 'depression' *per se*, but with the presence and trajectory of specific symptoms. Tracking forward to treatment, it would appear that any anti-inflammatory interventions should map accordingly. These first three studies strongly suggest that depression is dead, long live the depressions.

'I'll try anything once, twice if I like it, and three times just to be sure' said Mae West, and who are we to disagree – self-control is central to success. Discipline is seen to be related to all sorts of good outcomes; it's why we slip in that we are distance runners or competitive swimmers on job applications. Inherent in this is that pursuit of immediate pleasure is a failure of that self-control, and a threat to longer-term goals and positive outcomes. Bernecker & Becker created a measure of trait hedonic capacity, which combines the degree to which people feel pleasure in that short-term pursuit, with an inverse measure of intrusive thoughts regarding other conflicting long-term goals (pub with friends versus studying for an exam).[4] Interestingly, high hedonic capacity was independent of self-control, so not a failure at all; it was also associated with well-being and life satisfaction, with effect sizes over twice that seen for self-control. In the lab, those that were higher in hedonic capacity achieved greater relaxation, but if they had their competing goals activated experimentally beforehand they experienced the same amount of intrusive thoughts and diminished relaxation as the low capacity group. It appears that the key difference between groups is the spontaneous activation of conflicting goals, not the ability to inhibit the intrusive thoughts. In natural settings, those with the highest trait hedonic capacity were also highest in momentary pleasure ratings, and were judged to be enjoying it the most by others. High hedonic capacity proved stable over time and predicted more momentary enjoyment in everyday life and life satisfaction with a larger effect than self-control. The key to well-being seems to be the complementary and successful pursuit of both the more recognised long-term goals, as well as short-term pleasurable activities, making it not a failure of discipline but an integral part of self-regulation. The key is balance, supporting many goals at once. Mae West could have told us this years ago – 'you only live once, but if you do it right, once is enough'.

Deep learning (DL) and artificial neural networks (ANNs) – offering hope or hype for predicting individual-level treatment response? There is a lot of attention in the use of these computational models in helping classify different patient variables predicated on the argument that these more data-driven approaches may be superior to more conventional approaches. In a recent comprehensive review, Koppe et al argue that the success and performance of DL ANNs in other medical domains (such as radiology, drug discovery) might be helpful in psychiatry because they 'can implement very complicated, and in principle arbitrary predictor-response mappings efficiently' and therefore, might offer hope for the holy grail of predicting treatment response at the level of individual patients.[5] Most relevant to psychiatry is their argument that our relative paucity of very large data-sets (usually needed for DL applications) might not be as critical as once thought.

Classical ANNs (feedforward and recurrent kind) typically have an input and output layer of simulated neurons (nodes) as well as an

https://doi.org/10.1192/bjp.2020.155 Published online by Cambridge University Press

intervening 'hidden layer'. Data flow from the input to the hidden layer via weights or parameters – depending on their number and computations – the hidden layer provides for either an expansion or compression of information, feeding weighted data forward to the output layer. Thus, in these systems, learning is the process of mapping inputs onto outputs by incrementally adjusting the weights between layers so that the output layer has some desired property. DL networks differ from these more classical networks by having an architecture where there are multiple hidden layers: their training algorithms can be more complex, but inherit the same basic principles.

Take classification of an image as an example; a modest digital photograph of 256 × 256 pixels results in each image representing 65 536 input variables. If we want to classify a bunch of images as either a 'cat' or a 'dog', our ANN must construct or learn a mapping that reliably discriminate between the target output of 'cat' or 'dog'. Let's say our 65 536 input nodes have eight hidden nodes and one output (decision) node: that is at least 524 496 weights or parameters to learn. This highlights one controversy around ANN/DL: flexible mapping with so many parameters could arguably fit anything by learning only the specifics of the sample it was trained with – 'overfitting' – with the consequence that it would be likely to fail to generalise on new 'unseen' data. But ANN/DL architectures have delivered impressive results on high-dimensional data (huge numbers of inputs, far beyond 256 × 256 pixels) especially in image processing applications; the trade-off is that ANN/DLs usually require *huge* sample sizes of tens-of-thousands, way beyond classical statistical techniques like regression.

Koppe *et al* appeal to the oft-cited notion that DL networks can model highly non-linear relationships and decipher rare, but reliable, predictors from vast numbers of inputs. Performance in ANN is measured by validating on new data *not seen* by the network during training and recording the error. Koppe et al cite the appeal of big data 'as our sample increases in size, the variance (standard error) of estimated model parameters will decrease. As a consequence, we can afford more complex models which come with lower bias'. Embedded in this statement is the assumption that useful signal increases proportional to sample size because other factors (noise and measurement error) are effectively 'drowned out'. For prediction tasks in high-dimensional data (with thousands of inputs) noise accumulates as a function of the number of inputs, which means we are obliged to select a sparse set of relevant inputs (feature selection) to include in our prediction model. But feature selection in high-dimensional data is itself challenging because of spurious correlation; it becomes hard to establish if, for example, 50 out of 1000 candidate features, all having some small correlation with the output do not simply represent chance associations. This is problematic because one of the *appeals* of DL is that, embedded in a cornucopia of variables, are some nuggets that each reliably contribute a small amount of information to the task of classifying or predicting.

Koppe et al develop an intriguing idea; to harness the flexibility inherent in DL models by training on as many exemplar data as one can find, perhaps constraining the huge number of parameters using regularisation to deliver a 'pre-trained' network. With this network one would then further train on an individual patient's data to derive a 'personal' model for a given task such as predicting treatment response. If one can afford it, this might be an interesting future direction.

Finally, professionalism matters, to us and those we try help, but what does it mean? It is a softer aspect of practice, and one that varies across countries and cultures, and with time. Historically, its lack of objective criteria has meant it has been used as an exclusionary tactic – conscious or unconscious – against women and other marginalised groups, typically fitting in with what happens to be 'on brand' for white men. Social media have added a complexity, with multiple formats across which everyone – healthcare clinicians and academics, just like everyone else – can share their lives and experiences. We can simultaneously see the gains and potential challenges, as we all navigate the fact that our private lives our now far more public than ever before.

There was much recent negative attention on a paper in the *Journal of Vascular Surgery*, now retracted,[6] that explored the social media profiles of junior vascular surgeons. Among the more egregious aspects of the paper, three men on the academic team set up false profiles to view and 'rate' professionalism, 'scoring' individuals on a number of characteristics, including the photos they posted, and what they wore. Notably, women who had any pictures on their social media profiles that included them in swimwear had their 'professionalism' downgraded – by the three men. The backlash included the trending hashtag #MedBikini where (predominantly) women in healthcare posted pictures of themselves in swimwear noting their competencies and professionalism were not challenged by such antiquated, pejorative, and sexist 'opinions'. It all feeds into medicine's (and wider society's) broader patriarchal nature, with women and their bodies, objectified and damned whatever they do. The paper has opened up debate on professionalism, although scarcely in a way the authors of the original piece could have imagined. This included the question if #MedBikini itself potentially excluded some who might otherwise feel an ally to its primary message. In the UK, psychiatry trainees, as well as those in other specialties, are increasingly receiving training on 'managing' social media and the right to a private life in a highly visible world. Such conversations need to continue with everyone at the table, even if it's only to decide for ourselves what defines modern professionalism for our field. But using bad science to frame judging women on what they wear and how they look as objective, and defending it, especially by menu under the paternalistic guise of helping, needs to stop now.

References

1 Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. *Lancet Psychiatry* 2020; **7**: 692–702.

2 Cai N, Choi KW, Fried EI. Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. *Hum Mol Genet* 22 Jun 2020 [Epub ahead of print]. Available from: https://doi.org/10.1093/hmg/ddaa115.

3 van Eeden WA, van Hemert AM, Carlier IVE, Penninx BWJH, Lamers F, Fried EI, et al. Basal and LPS-stimulated inflammatory markers and the course of individual symptoms of depression. *Transl Psychiatry* 2020; **10**: 235.

4 Bernecker K, Becker D. Beyond self-control: mechanisms of hedonic goal pursuit and its relevance for well-being. *Pers Soc Psychol Bull* 26 Jul 2020 [Epub ahead of print]. Available from: https://doi.org/10.1177/0146167220941998.

5 Koppe G, Meyer-Lindenberg A, Durstewitz D. Deep learning for small and big data in psychiatry. Neuropsychopharmacology 15 Jul 2020 [Epub ahead of print]. Available from: https://doi.org/10.1038/s41386-020-0767-z.

6 Hardouin S, Cheng TW, Mitchell E, Raulli SJ, Jones DW, Siracuse JJ, et al. RETRACTED: Prevalence of unprofessional social media content among young vascular surgeons. *J Vasc Surg* 2020; **72**: 667–71.

https://doi.org/10.1192/bjp.2020.155 Published online by Cambridge University Press