

Kaleidoscope

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

Perceptions 1. #BlackLivesMatter has reinforced the depth of societal inequities; minority ethnic status is linked with psychosis, but disentangling discrimination from other factors is challenging. Bardol et al included 17 studies ($n = 33\ 211$) in their meta-analysis demonstrating that perceived ethnic discrimination had a statistically significant odds ratio for the occurrence of psychotic symptoms (1.77) and experiences (1.94).¹ The findings held across different ethnic groups, and did not vary by country of origin; although ethnic identity, collective self-esteem and social support buffered the findings, this was a weak influence. Interestingly, given previous research, there was no support that ethnic density moderated the results. These findings are contrary to the suggestions that the association with discrimination could be entirely explained by socioeconomic and other confounders, although psychotic symptoms and experiences are not the same as a clinical diagnosis of psychosis. Social exclusion and ‘social defeat’ have been linked previously with the aetiology of psychoses; there are echoes of qualitative reports from many individuals, showing the profound adverse impact even a single racist comment can have on a person’s life.

Perceptions 2. Childhood maltreatment is linked with adverse later life outcomes – but how much does one’s subjective experiences contribute relative to more ‘objective’ markers? Danese & Spatz Widom compared court-documented evidence of maltreatment of children with individuals’ subjective recall of events and psychiatric histories as adults.² The authors accessed a unique sample of just over 900 official records from the USA to find that individuals’ perceptions were far more important: the risk of subsequent psychopathology was minimal, even in the presence of severe objective reports of trauma, without confirmatory subjective accounts. Conversely, the risk of psychopathology was high in subjective recall of significant events, whatever the objective data showed. This court-report data offers a greater level of confidence in objectivity, given the legal standards required, surpassing that more commonly obtained from single source corroborators such as parents and teachers. The authors note the findings were ‘remarkably invariant’ across different forms of maltreatment, genders and ethnic backgrounds – subjective experience is what drives any subsequent psychopathology. The findings are important at several levels: they inform future research in terms of how best one may stratify individuals in terms of risk; they may help better understand the pathogenesis and mechanisms through which trauma has an impact on mental health; and most importantly, it offers up direction to therapeutic intervention about where clinical efforts might best be focused.

Perceptions 3. Which factors influence the perception of treatment of depression as being helpful? Harris et al examined a coordinated series of community epidemiological surveys of over 80 000 individuals with a lifetime history of treatment for major depressive disorder across 16 low-, middle- and high-income countries.³ An interesting component of this work is that it included all forms of ‘treating professionals’, including: spiritual advisors, herbalists, acupuncturists, and other forms of ‘alternative medicine’. Factors that were linked with perceiving treatment as more helpful were: older age at first treatment, higher educational level, shorter delay in instigating treatment and medication from a mental health specialist.

The authors argue that the likelihood of receiving effective treatment is the mathematical product of two factors: the probability of perceiving a treating individual as helpful multiplied by that of persisting in help-seeking after any failure. It is notable that the combination of ‘mental health specialist plus medication’ was most strongly associated with perceived helpfulness: qualifications and experience matter, although of course they are no guarantee of success. Multifaceted supports in an individual’s life must be encouraged, but these data back up the science – folk with major depressive disorder need to see mental health professionals, not least as many who receive early unhelpful treatment will stop asking for more help.

Interpreting data 1: Gelman & Loken have asked why empirical research was failing on the scientific standard of reproducibility.

They proposed that researchers are focused on computing and presenting a single hypothesis test, but observed that a different test would have been used had the data presented differently. This can be examined using the example of assessing reaction times in two groups of people where we have an *a priori* hypothesis that one group will be slower. The group distributions show ceiling effects for some participants – they pause for long enough to have extreme reaction times, so they are excluded – we do not think they are trying. Then, although the means appear similar, the two distributions are quite skewed, making one group mean misleading as a measure of central tendency. So, a transformation is applied to the data to make the distributions better behaved. Satisfied the data are ‘cleaned’, we use a *t*-test to compare the means of the two distributions. It turns out there is a group difference in means, and the *t*-value exceeds the critical threshold to declare $P < 0.05$. We publish the result as a significant difference confirming our hypothesis that reaction times on the task differ between the two groups. Had the data looked different, we might have chosen different exclusion thresholds, transformation and statistical tests. The single hypothesis test presented is a function of the decisions made (the path) from data to presented result.

Functional magnetic resonance imaging (fMRI) experiments have complex data pre-processing pipelines representing just these kinds of forking paths from raw data to hypothesis tests. A study by Botvinik-Nezer et al gave 70 different teams identical access to the same fMRI data-set to analyse and test nine pre-defined hypotheses.⁴ The teams were told to analyse the data as they would routinely for their own research and to report a ‘yes’ or ‘no’ answer to each hypothesis. Of the nine hypotheses, across the 70 teams, one hypothesis received 84% support for a significant finding; three hypotheses were consistently reported as non-significant and for the remaining five hypothesis, rates of reported significance varied from 21% to 37%. The authors examined which aspects of the analysis pipelines contributed to most of this variation. In fMRI research, the data are routinely ‘smoothed’ to reduce artefacts. They found the amount of smoothing was strongly associated with significant hypothesis outcomes. Further, they note the imaging data ‘maps’ produced at intermediate steps of the analysis pipeline also varied substantially between teams. They suggest this represents different teams’ decisions about how to correct for multiple comparisons as well as differing specifications of the regions of anatomical interest used in the analyses. Consistent with Gelman & Loken, they conclude that ‘Our findings highlight the fact that it is hard to estimate the reproducibility of single studies that are performed using a single analysis pipeline’. This suggests that consistency in results will be improved by using different analytic pipelines.

Interpreting data 2: what do effect sizes and clinical significance in antidepressant trials mean? Most people will ‘accept’ the classic meta-analytical finding of a standardised mean difference of about

0.3 in favour of antidepressants over placebo on the Hamilton Rating Scale for Depression (HRSD); what that *means* in terms of clinical significance remains more contentious. Hieronymus et al note the limitations of using group averages, and the lack of a clear scientific basis for the proposed HRSD standardised mean difference cut-off of 0.875 for ‘minimum clinical significance’ (a seven-point HDRS difference) – higher than most comparable treatments across medicine.⁵ Indeed, they argue that there is no agreed delineator for determining clinically significant from insignificant. They suggest that research should include measures that directly reflect functioning and well-being, as well as the current focus on capturing changes to depression psychopathology via summed scores on scales. We are reminded that there is a research bias that studies utilise the HRSD because past studies did so too, but it is quite an old tool measuring somatic symptoms such as gastrointestinal and sleep disturbances that can be *caused* by antidepressants in the acute phase. It is increasingly recognised that when one isolates the core symptom of low mood, antidepressants separate quite well from placebo.

Interpreting data 3: might there be validity and utility to a general ‘p’ measure of overarching psychopathology? Most will be familiar with the ‘g’ factor of general intelligence, which, despite obvious broad brushstrokes downfalls, has some utility in predicting individuals’ longer-term outcomes. Using ‘g’ as a comparison, Pettersson and colleagues investigated the magnitude and predictive validity of ‘p’ using Swedish national registers that included data from: over 1 million adults, 2000 adolescents and their parents, and parent data on 14 000 children.⁶ Predictive validity of ‘g’ was established from Swedish military conscription data from over 400 000 men born between 1980 and 1992. For ‘p’, these data were linked to the National Patient Register to identify any diagnoses before military service. Follow-up occurred on average 10.9 + 3.3 years after military enrolment, and measures included ‘highest income’ and ‘educational level obtained’, as well as Swedish Scholastic Aptitude Test results. In addition, 16 adverse outcomes were gathered from national databases, including: overdoses, suicide, psychiatric medication prescription, crime convictions and the use of social benefits.

The measures of intelligence and psychopathology were found to have similar magnitudes across age groups, indicating a utility of the single score approach. Within the army conscription group, ‘p’ predicted all negative register-based outcomes a decade later, even when controlling for intelligence with a strength equivalent to how ‘g’ predicted positive educational and financial outcomes. Of course, scepticism is warranted. Like intelligence, it is difficult to say what ‘p’ is actually measuring, although it may be a useful summary with value for clinicians and researchers nonetheless. As a proxy for prognosis, ‘p’ may have some potential for supplementing management plan decisions as to who may need access to care or additional support. This raises ethical concerns where there is potential for some individuals being perceived to have a very poor prognosis.

Finally, perceptions of interpreted data: hard to imagine the popular press ever get it wrong, but what if – sometimes – they

were sensationalist on the science...? Pathak et al evaluated online articles on antidepressants and psychotherapy in five UK national newspapers: *The Sun*, *Daily Mirror*, *Daily Mail*, *Daily Express* and *The Guardian*.⁷ The data are clear that in terms of treating depression, both drugs and psychotherapy are approximately equally effective – and often optimal when combined – so did the press track this? Headlines and content of 221 articles published between 2013 and 2018 were independently rated and in the main articles, antidepressants were portrayed negatively 37% of the time, positively 15% of the time and neutrally 48% of the time; for psychotherapy, the figures were 36% positive, 2% negative and 61% neutral. Antidepressants got more coverage, featuring in 184 articles, compared with 132 for psychotherapy, although clearly here all publicity is not good publicity. Two of the more disturbing headlines noted were ‘Mother says antidepressant drugs turned her son into a “psychotic killer”’ and ‘Antidepressant drugs are “immensely harmful” and responsible for thousands of deaths, claims leading scientist’. Media coverage of psychiatrists has been shown to be more negative than that towards other medical specialities; this is extended to one of our most common interventions. *The Daily Mail* was by far the leading offender – something we are sure will come as a shock to you. The impact of this bias in reporting on public perception remains unclear, in terms of stigma, misinformation and perhaps most concerning in influencing decisions to access or utilise healthcare.

References

- 1 Bardol O, Grot S, Oh H, Poulet E, Zeroug-Vial H, Brunelin J, et al. Perceived ethnic discrimination as a risk factor for psychotic symptoms: a systematic review and meta-analysis. *Psychol Med* 2020; **50**: 1077–89.
- 2 Danese A and Spatz Widom C. Objective and subjective experiences of child maltreatment and their relationships with psychopathology. *Nature Hum Behav* [Epub ahead of print] 18 May 2020. Available from: <https://doi.org/10.1038/s41562-020-0880-3>.
- 3 Harris MG, Kazdin AE, Tat Chiu W, Sampson NA, Aguilar-Gaxiola S, Al-Hamzawi A, et al. Findings from World Mental Health Surveys of the perceived helpfulness of treatment for patients with major depressive disorder. *JAMA* [Epub ahead of print] 20 May 2020. Available from: <https://doi.org/10.1001/jamapsychiatry.2020.1107>.
- 4 Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 2020; **582**: 84–8.
- 5 Hieronymus F, Jauhar S, Dinesen Østergaard S, Young AH. One (effect) size does not fit at all: interpreting clinical significance and effect sizes in depression treatment trials. *J Psychopharmacol* [Epub ahead of print] 25 May 2020. Available from: <https://doi.org/10.1177/0269881120922950>.
- 6 Pettersson E, Larsson H, D’Onofrio BM, Bølte S, Lichtenstein P. The general factor of psychopathology: a comparison with the general factor of intelligence with respect to magnitude and predictive validity. *World Psychiatry* 2020; **19**: 206–13.
- 7 Pathak A, Lim E, Lawrie S. Needless controversy: the reporting of pharmacological and psychotherapy for the treatment of depression in the UK media. *Psychol Med* [Epub ahead of print] 29 May 2020. Available from: <https://doi.org/10.1017/S0033291720001427>.

