approach may be more efficient under certain circumstances. In randomised trials, however, both approaches can be assumed to provide the same underlying intervention effects if the trials are adequately randomised: the difference in mean endpoint scores will be the same on average as the difference in mean change scores. To our knowledge, the decision for using one measure over another is currently not resolved, with evidence suggesting that endpoint scores tend to produce more conservative estimates.[3] Cella & Preti further suggested applying a restricted maximum likelihood estimation method to the meta-analysis. Research synthesis methodology is a developing field and there is no consensus at present on the best approach for random-effects meta-analysis. Although valuable alternatives to random-effects estimation methods have been developed,[4] DerSimonian–Laird remains the most widely used approach and an adequate method in most scenarios.[5]

Finally, Cella & Preti raised two points about the way control conditions were handled in the review. We would like to reassure readers that we did not double-count studies by considering active control conditions for one analysis as active treatment conditions in another analysis. We further wish to clarify that treatment participants in all trials received the psychosocial intervention as an adjunct to treatment as usual (TAU), as doing otherwise would be considered unethical.

1 Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, 2011.

2 Elis O, Caponigro JM, Kring AM. Psychosocial treatments for negative symptoms in schizophrenia: current practices and future directions. *Clin Psychol Rev* 2013; **33**: 914–28.

3 Fu R, Holmer HK. Change score or follow-up score? Choice of mean difference estimates could impact meta-analysis conclusions. *J Clin Epidemiol* 2016; **76**: 108–17.

4 Cornell JE, Mulrow CD, Localio R, Stack CB, Meibohm AR, Guallar E, et al. Random-effects meta-analysis of inconsistent effects: a time for change. *Ann Intern Med* 2014; **160**: 267–70.

5 DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials* 2015; **45**(Pt A): 139–45.

Danyael Lutgens, MSc, Department of Psychiatry, McGill University, Douglas Mental Health University Institute, Montréal, Quebec; **Genevieve Gariepy**, PhD, McGill University, Institute for Health and Social Policy, Montréal, Quebec; **Ashok Malla**, MD, FRCPC, McGill University, Douglas Mental Health University Institute, Montréal, Quebec, Canada. Email: ashok.malla@mcgill.ca

doi: 10.1192/bjp.211.2.116a

The clinical utility of the ADI-R and ADOS in diagnosing autism

Larson *et al*[1] report on a major study on psychosis in autism, which is an important topic.[2] They point out that their sample is non-representative, but then use the Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS) for clinical diagnosis. This vastly increases the non-representativeness of the sample unnecessarily and takes it a very long way away from autism in the general population. The criteria they used to define autism are very narrow concepts of the disorder. Clinical diagnoses based on this narrow view tell us very little about autism as seen in routine clinical practice, where professionals throughout the world now accept that the broader autism phenotype. I see many parents who come to me in great distress knowing that their child has autism and that the school also observed this, but having been told that their child did not have autism according to the ADI-R. This instrument is not appropriate to making a sole diagnosis of autism in clinical practice. It not uncommonly misses high-functioning autism. In addition, Ventola *et al*[3] have shown that the ADI-R was significantly 'under-diagnosing toddlers'. How biased and unrepresentative the patients in this survey can be seen by Professor Gillian Baird's work on autism in the general population.[4] Indeed, using these narrow criteria gives a prevalence of autism of 25 per 10 000. When you use the broader autism spectrum, you get a truer rate of 116 per 10 000. One of the problems also is that the National Institute for Health and Care Excellence (NICE) guidelines on the diagnosis of autism,[5] which are accepted throughout the world, are not followed. These state that there is no specific instrument recommended for diagnosis of autism and that identification depends on a clinical diagnosis by an experienced clinician. Dorothy Bishop, Professor of Developmental Neuropsychology at the University of Cambridge, told Adam Feinstein that, 'If it could be shown that there were real benefits in accuracy of diagnosis from adopting this lengthy procedure, then I'd be happy to say: "Okay". But the originators of the instrument have never demonstrated [this] – it is really more an article of faith with them'.[6] Feinstein also reports that, at the prestigious International Meeting for Autism in London in 2009, senior autism researchers 'lambasted' these narrow instruments 'for missing many cases of autism'.

1 Larson FV, Wagner AP, Jones PB, Tantam D, Meng-Chuan L, Baron-Cohen S, et al. Psychosis in autism: comparison of the features of both conditions in a dually affected cohort. *Br J Psychiatry* 2017; **210**: 269–75.

2 Fitzgerald M. Schizophrenia and autism/Aspergers syndrome: overlap and difference. *Clin Neuropsychiatry* 2012; **9**: 171–6.

3 Ventola PE, Kleinman J, Pandey P, Barton M, Allen S, Greene J, et al. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *J Autism Dev Disord* 2006; **36**: 839–47.

4 Baird G, Simonoff E, Pickles A, Chandler S, Loucast Meldrum D, Charman T. Prevalence of disorders of the autism spectrum disorder in a population cohort of children in South Thames: the Special Needs and Autism Project. *Lancet* 2006; **368**: 210–5.

5 National Institute for Health and Care Excellence. *Autism: Recognition, Referral, Diagnosis and Management of Adults on the Autism Spectrum*. British Psychological Society & Royal College of Psychiatrists, 2012.

6 Feinstein A. *A History of Autism*. Wiley–Blackwell, 2010.

Michael Fitzgerald, Psychiatrist and Adjunct Professor of Psychiatry, Department of Psychiatry, Trinity College Dublin. Email: profmichaelfitzgerald@gmail.com

doi: 10.1192/bjp.211.2.117

Author's reply: Professor Fitzgerald comments on the 'real-life' utility of the ADI-R and ADOS. It is his view that this may have limited the representativeness of our sample with respect to the autistic presentation. However, the ADI-R or ADOS were used only for inclusion/exclusion when we were referred participants who did not have an existing clinical diagnosis of autism spectrum disorder (ASD). This was a small number of participants ($n = 19$), and the number who were excluded because they did not meet cut-offs was even smaller ($n = 8$). It is interesting to note that our experience actually supports Fitzgerald's observations, in that participants in the research did differ significantly from a comparison sample of people with ASD in terms of their ADI-R scores, as discussed in our article.[1]

The wider question alluded to in Fitzgerald's letter is one of categorical diagnoses and the utility of boundaries. Obviously,

research requires a common language in order to facilitate discovery – I need to know that what I am measuring is equivalent to something of the same name measured in another country by another researcher at another time. Thus, instruments such as the ADI-R and ADOS (considered gold standards in ASD research) are vital. They also allow for meaningful comparison of groups – it is only when the cut-offs are applied rigidly that they become less useful, and this was not the case in our research.

However, a tremendous amount of heterogeneity exists among people with ASD (those who meet current categorical definitions), let alone those who have some symptoms but do not meet full diagnostic criteria (the broader autism phenotype). The challenge, then, is what to do with individuals who lie in different places along the spectra that comprise a standard ASD conceptualisation. As Fitzgerald rightly points out, it is those individuals who have independently learned or been supported to cope with their differences in a way that allows them to function in a 'neurotypical' society who are missed and excluded by the diagnostic categories that the ADOS and ADI-R conform to. However, simply because they appear to be functioning well does not mean that they are not experiencing difficulties – I agree with Fitzgerald. It was my experience conducting this research that, for many participants, it was actually because of the pressure of coping, or because they were not recognised as struggling, that many individuals got into difficult circumstances that precipitated the onset of psychosis or other serious mental health problems – an observation that is unsurprising for many clinicians, I'm sure.

The challenge for research and clinical practice, then, is to find a way to bridge the gap between rigid diagnostic categories and representative samples. This is a problem for psychiatry as a whole, not just those interested in certain conditions, which makes innovations such as the research domain criteria initiative from the National Institute for Mental Health so relevant and interesting.[2]

1 Larson FV, Wagner AP, Jones PB, Tantam D, Meng-Chuan L, Baron-Cohen S, et al. Psychosis in autism: comparison of the features of both conditions in a dually affected cohort. *Br J Psychiatry* 2017; **210**: 269–75.

2 Sanislow CA, Pine DS, Quinn KJ, Kozak MJ, Garvey MA, Heinssen RK, et al. Developing constructs for psychopathology research: research domain criteria. *J Abnorm Psychol*, 2010; **119**: 631–9.

Felicity Larson, PhD, Department of Psychology (Clinical Psychology Section), Frankland Building, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. Email: fxl437@bham.ac.uk

doi: 10.1192/bjp.211.2.117a

The evidence base and readability of Freeman *et al* on virtual reality for treating delusions

I commend Freeman *et al*[1] for their innovative research using virtual reality in the treatment of persecutory delusions. The rather 'soft' finding - probably valid, but hardly surprising – is: 'Cognitive therapy using virtual reality could prove highly effective in treating delusions'.

I have concerns about the evidence base of the study, that is the design, the data, the sample and the statistical methodology. Each of these concerns interrogates the validity and reproducibility of the study.[2,3]

First, the sample size is extremely small – 30 participants. The consequences of this include overestimates of effect size and low reproducibility of results.

Second, 27 of the 30 participants were unemployed. There is little point to any research if one cannot extrapolate from one's sample to some broader reference population. For this, the sample should mimic the population in important ways. However, there is no discussion about the sampling, or the reference population to which extrapolation might be extended.

Third, neither the patients nor the researchers were masked to the randomisation allocations. This, surely, is a fundamental flaw of the experiment. A double-blind experiment should be used to ensure impartiality, and avoid bias, such as, for example, the Hawthorne effect.[4]

Fourth, the main outcomes – comparing the delusional conviction of the two groups at the beginning and end of testing, as well as their distress – were tested using ANCOVA. But in the results section the authors report: 'For ratings of conviction in paranoia, a gradual reduction across the scenarios for the threat belief testing group can be seen, whereas the conviction scores remain stable in the exposure group' (p. 64). This suggests that the two groups diverge over time, having different slopes, rather than the assumed homogeneous slopes in the ANCOVA model.

Fifth, the term 'repeated measures mixed model' covers a wide range of possible models, and leads one to expect a single model incorporating the repeated measures and random effects, not ten models as are presented in the online supplement. Further, none of the models is clearly articulated in mathematical form.

Sixth, there are no graphs to display the data or statistical results. Tay *et al*[5] propose the use of graphical descriptives to enhance research rigour, especially in psychology.

It appears that the article is written on two levels. The introduction and method sections, describing participants, design and virtual reality, are clear and lucid. By contrast, the evidence base of the article, discussing the data, models, analysis and results, is almost unintelligible. Further, the small sample size, sampling bias, lack of randomisation masking, lack of model specification and lack of statistical graphics, seriously undermine the study.

The phrase 'evidence-based research' has become popular in psychology. Thus, it is incumbent on readers, authors and journal editors to 'raise the bar' and demand higher standards of the evidence base of research studies.

1 Freeman D, Bradley J, Antley A, Bourke E, De Weever N, Evans N, et al. Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. *Br J Psychiatry* 2016; **209**: 62–7.

2 Gardenier JS, Resnik DB. The misuse of statistics: concepts, tools, and a research agenda. *Account Res* 2002; **9**: 65–74.

3 McNutt M. Reproducibility. *Science*, 2014; **343**: 229.

4 McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. *BMC Med Res Method* 2007; **7**: 30.

5 Tay L, Parrigon S, Huang Q, LeBreton JM. Graphical descriptives: a way to improve data transparency and methodological rigor in psychology. *Perspect Psychol Sci* 2016; **11**: 692–701.

John L. Fresen, Statistician, IT Services, University of Oxford, Oxford, UK. Email: john.fresen@it.ox.ac.uk

doi: 10.1192/bjp.211.2.118

Community treatment orders and capacity to consent

We welcome Newton-Howes & Ryan's plea for a more restrictive use of community treatment orders (CTOs).[1] They have a heavy