EDITORIAL

What does it mean? Looking beyond group differences in clinical research

David Coghill

It is usual practice for scientific papers to focus on the description of groups of subjects in terms of their mean scores on a particular measure and to assess differences between these groups by comparing these means using standard statistical tests. When we read that, "compared to a control group, those with a particular disorder or problem scored significantly higher on a symptom questionnaire", or "performed less well on a particular neuropsychological task", or that "those patients treated with an active treatment had lower symptom scores at the end of the trial than those treated with a placebo", we are able to make certain assumptions about that disorder or treatment. If we are interested in the strength of an effect we will often look to the effect size for an answer. There is no doubt that incorporating information derived from such comparisons is not only appropriate but forms the basis of evidence-based clinical practice. However, sometimes it is informative to look beyond the means and consider the heterogeneity and inter-individual variability that is often seen across many levels of analysis including symptoms, cognition and more basic causal processes.

Andres and colleagues (this issue) compared cognitive functioning in a group of adolescents with anorexia nervosa to a matched group of healthy controls. Interestingly, despite measuring performance across a wide range of tasks assessing multiple cognitive domains, the two groups differed on only one task, the Rey Complex figure Task, but not on any of the other measures. From this one could conclude that adolescents with anorexia nervosa have only

very limited cognitive impairments. However, in addition to comparing the two groups using mean scores Andres and colleagues also identified those individuals who were cognitively impaired on two or more tasks. From this perspective they found that significantly more of those with anorexia had significant cognitive impairment compared to the healthy controls (30 vs. 7%). While this still means that most of the anorexic subjects were not impaired, there were clearly a significant minority who had significant problems.

One great strength of this study, and something that makes these findings even more relevant, is the way that the authors worked very hard to recruit a homogeneous patient group. Previous studies in this field included a much broader group of individuals with a wide range of eating disorders. This, of course, makes it difficult to know whether differences between individuals are a consequence of diagnostic heterogeneity rather than heterogeneity *within* the clinical phenotype. However, here we have a well-characterised patient cohort with a short duration of illness all of whom were in the acute phase of their illness. This clearly indicates that the neuropsychological heterogeneity, which was found in this group of adolescents and which is similar to that previously demonstrated in less well-defined cohorts of adults with anorexia nervosa is related to the disorder itself rather than to some broader associated factors.

Analogous findings have been reported across several other disorders. Perhaps the best researched example within child and adolescent mental health is attention deficit/hyperactivity disorder (ADHD) where several groups of authors have highlighted the considerable cognitive heterogeneity found within ADHD samples. Nigg et al. [4] described data from three independent sites that suggested while many neuropsychological tasks can differentiate those with ADHD from healthy controls, only a

D. Coghill (✉)
Division of Neuroscience, Medical Research Institute,
College of Medicine, Dentistry and Nursing,
Ninewells Hospital and Medical School, Dundee, UK
e-mail: d.r.coghill@dundee.ac.uk

subgroup of individuals with ADHD does show a deficit on any one particular task. A significant proportion of those with ADHD did not appear to have a deficit on any of the included measures. Nigg and colleagues focused exclusively on "executive functioning", however subsequent work has suggested that a similar picture is found when other aspects of cognitive functioning are included [5]. Most studies to date utilized the DSM-IV ADHD phenotype to describe cases; it is therefore possible that these findings were a reflection of the rather broad nature of this diagnostic category and its various subtypes. In an attempt to control for this possibility Coghill et al. [2] investigated neuropsychological heterogeneity within the narrower ICD-10 hyperkinetic disorder phenotype. Intriguingly, they found a very similar pattern of neuropsychological heterogeneity within this more tightly defined clinical group that was identified by the previous studies.

Another, somewhat related, example of the limitation of relying solely on effect sizes in the presence of significant within group heterogeneity is demonstrated by comparisons of the effect sizes associated with symptom scores and those associated with underlying neuropsychological processes. For example, a common criticism of neuropsychological theories of ADHD is that when one compares cases with controls the effect sizes for the neuropsychological deficits, usually fall between 0.4 and 0.6. This is considerably smaller than those for ADHD symptoms (between 2.5 and 4.0). It is therefore argued that this implies that the cognitive deficits cannot be causal. However, this conclusion may be rather premature. It is true that such findings against the notion that all cases of ADHD share the same underlying neuropsychological deficit as proposed by single cause theories of ADHD such as that put forward by Barkley [1]. However, this does not necessarily mean that they are not causal. If neuropsychological impairments are heterogeneous across those with ADHD (i.e. not every person with ADHD has any one particular deficit) then the effect size for any particular deficit will be smaller than that for the behavioural symptoms, which are the defining characteristic of ADHD and by definition occur in all of those with ADHD. While the overall effect for the group is relatively small the effect for a particular individual may be much higher.

A shift of focus away from group differences and towards inter individual differences can also be informative when interpreting from the results of clinical trials. All clinicians will be aware that even where very strong group level treatment effects have been shown across multiple clinical trials it is rarely, if ever, the case that one treatment is found to be universally effective. There is always a proportion patients who either fail to respond, or do not tolerate a particular treatment. While it will be reassuring for the patient to know that the treatment offered was consistently more effective at a group level than either a placebo or an alternative treatment, it might be even more important for them how likely they, as an individual, are to respond. While data on response rates are now much more routinely reported than it was in the past, it is still often the case that the definition of response is somewhat arbitrary. Thus, it would be very helpful if standardised definitions of response such as those described by Jacobson and Truax [3] were routinely incorporated into clinical trials.

Individual differences are also extremely important when considering adverse treatment effects. Adverse effects on pulse and blood pressure associated with ADHD medications illustrate this notion quite well. Stimulant medication is recognised to cause a small increase in heart rate averaging 1–2 beats per minute. However, these reports of mean changes hide a small proportion of cases where the increment is larger—up to 50 beats per minute [6]. Similarly the average increases in both systolic and diastolic blood of 3 to 4 mm Hg, typically seen in clinical trials of ADHD medications, are often described statistically but not clinically significant. However, these small average changes mask a significant minority of patients where the rise is potentially harmful. Categorical data suggest that 6.8% of those treated with atomoxetine will shift into the hypertensive range after starting treatment. Although similar data on stimulants is lacking it seems likely that the picture is similar. Instead of only considering effect sizes, clinicians should also consider the Number Needed to Treat (NNT) and its converse Number Needed to Harm (NNH). These parameters refer to the average number of patients one needs to treat in order to make one person better (NNT) or suffer a particular adverse event (NNH). These numbers are easy to calculate, when the appropriate data are made available, and, easy to interpret once one is familiar with the concept. Several familiar reference points are available (e.g. NNT for stimulants and atomoxetine in ADHD ≈ 4, SSRI antidepressants for child and adolescent depression ≈ 9, CBT for child and adolescent depression ≈ 12, NNH for fluoxetine in adolescent depression for significant adverse effects ≈ 21). Having this type of information to hand, within the clinic, is extremely helpful when discussing the size of any potential risks with patients. Unfortunately many trials fail to report these data, particularly that relating to adverse effects. It is, for example, rather unfortunate that data on rates of hypertension and tachycardia associated with stimulant treatment are not readily available.

It is certainly not my intention to suggest that neither researchers should abandon the search for differences between large groups of individual nor the journal readers should stop being interested in the effect sizes associated with such differences. It is, however important that clinicians think about the implication of these differences, or

lack of difference, on their patients. It is also important that that researchers ensure that their data are presented in a such a way that that the clinician has the best opportunity to make informed judgements about the diagnosis and treatment and discuss relative risks associated with treatment with their patients.

References

1. Barkley RA (1997) Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 121:65–94
2. Coghill DR, Rhodes SM, Matthews K (2007) The neuropsychological effects of chronic methylphenidate on drug-naive boys with attention-deficit/hyperactivity disorder. Biol Psychiatry 62:954–962
3. Jacobson NS, Truax P (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 59:12–19
4. Nigg JT, Willcutt EG, Doyle AE, Sonuga-Barke EJ (2005) Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaired subtypes? Biol Psychiatry 57:1224–1230
5. Sonuga-Barke E, Bitsakou P, Thompson M (2010) Beyond the dual pathway model: evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 49:345–355
6. Wernicke JF, Faries D, Girod D, Brown J, Gao H, Kelsey D, Quintana H, Lipetz R, Michelson D, Heiligenstein J (2003) Cardiovascular effects of atomoxetine in children, adolescents, and adults. Drug Safety Int J Med Toxicol Drug Exp 26:729–740