DOI:
10.1111/jcpp.12461

Document Version
Peer reviewed version

"Threats to the validity of child psychiatry and psychology"

By

Michael Rutter and Andrew Pickles*

***Equal joint first authorship**

Our paper is dedicated to David Sackett (b. 17.11.34, d. 13.5.15) who pioneered evidence based medicine and led the way in using and implementing therapeutic trials. His textbook (Sackett, Haynes, & Tugwell, 1985) was the first to deal systematically with evidence based medicine (although the term was not coined until 1991). The book went through several editions and is rightly regarded as a classic.

Corresponding author:

Michael Rutter

PO Box 80, SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience.

De Crespigny Park, Denmark Hill, London SE5 8AF

michael.rutter@kcl.ac.uk

Background

Suggestions have been made that many claims concern false positive findings.

Methods

The literature was searched for concepts and findings on the validity of child psychiatry and psychology.

Results

Substantial progress has been made in some, but not all, areas and considerable challenges remain in all.

Conclusions

The two major threats to validity concern the inability to examine brain tissues in life and the evidence that there is a high overlap among disorders. We emphasize the need to follow published guidelines on pre-planned analyses and we note the dangers associated with unregulated flexibility in data analysis. We note the very important clinical and

developmental findings that have been ignored, perhaps partly because of an excessive focus on technologies. Nevertheless, we are positive about both the accomplishments and the ways in which challenges are being met.

Keywords: classification, diagnosis, genetics, epigenetics, bioinformatics, brain imaging, biomarkers, neural networks.

Introduction

Before turning to the perceived threats to the scientific validity of child psychiatry and psychology we take stock of the accomplishments of the last fifty years or so, noting the challenges and opportunities that lie ahead. The book edited by Bloch et al., (2014) provides a useful discussion of some (but not all) of the key issues and we refer to several chapters in our overview.

Accomplishment, challenges and opportunities

It is appropriate to begin by noting the early pioneering establishment of both epidemiological studies of sizeable samples – such as in the Isle of Wight studies beginning in the 1960s in which their value for service planning was combined with their use for identifying risk factors for mental disorder (see Rutter, Tizard & Whitmore, 1970) - and, similarly, the British national cohorts studies beginning in 1946 (Wadsworth, 1991) which were later followed by, among others, the highly productive longitudinal studies in Christchurch and Dunedin in New Zealand (Fergusson et al., 1996; Moffitt et al., 2001). Again, creative use was made of their value for tackling important scientific questions. The challenge lies in the fact that so many other reports are still based on small underpowered cross-sectional samples (see Kapur, Phillips & Insel, 2012: Rietveld et al., 2014). We discuss this worrying issue more fully in a separate section of this paper.

A further accomplishment lies in the recognition of both the need to study causal mechanisms and the development of guidelines for doing this - as best exemplified by Bradford Hill's 1965 paper, recently republished (Bradford Hill, 2015). Of course, this early work was not concerned with child psychiatry or psychology, but the Baron and Kenny (1986) introduction of the concept and statistical method of mediation analyses was. We consider some of the statistical issues later in this paper but, for analyzing treatment effects, problems remain in convincing practitioners of the need to use manual-guided structured treatments (see Weisz & Kazdin, 2010 for an excellent review of both the scientific and practical, as well as the conceptual issues). There are similar needs in the consideration of moderation effects – that is the variation among subgroups in either the degree or pattern of response (see Fonagy et al., 2015).

The systematic evaluation of both pharmacological and psychological interventions constituted an important advance – particularly with respect to the use of randomized controlled trials (RCTs) in the testing of treatment efficacy. With respect to RCTs, despite their having been regarded as providing the 'gold standard' for the testing of efficacy, three main areas of uncertainty remain.

First, as Cartwright (Cartwright & Munro, 2010) has argued, although RCTs provide the best internal validity of efficacy, they may have limited external validity (that is the finding may not be applicable to the populations intended to receive the intervention). Rawlins (2008), too, has noted the limitations stemming from the complexity of the treatment, key moderators, and issues of safety for use in children.

Second, because of these considerations, and the possible situation specificity of effects, Fonagy et al., (2015) argued for a more unusual approach. It is usually misleading, they argued, to imply that a particular form of treatment is, or is not, effective, regardless of

circumstances. We agree. They also suggested that in some instances, observational studies (provided they are based on large, appropriately chosen samples and provided adequate attention has been paid to possible alternative explanations) may be informative on efficacy and more especially on side-effects. However, Fonagy et al., (2015) also argued that a more relaxed approach was needed when fewer evaluations were available. Whilst we agree on the value of considering a range of research strategies, we emphasize the major danger of misleading inferences from observational studies because of the biasing influence of confounding variables (Academy of Medical Sciences, 2007). When RCTs are not available, it may be particularly dangerous to rely on observational studies, and there should be no compromises with respect to the quality of the research.

Third, it has been argued by Howe et al., (2002) that, provided strict criteria are followed (and often this may be difficult) prevention trials using RCTs can test theories of causation. Emsley and colleagues (2010) have gone further in arguing that statistical developments sometimes allow confounding between mediators and outcome to be removed. That provides a methodological improvement but it cannot deal with the key issue that the element that is operative in treatment may not be operative in the earlier process that led to the disorder being studied. Thus, ECT may be effective in the treatment of severe psychotic depression but the depression will not have been caused by a lack of electric shocks. Of course, prevention trials can play a major role in the study of mechanisms (see, for example, Green et al., 2015) but the problem remains that finding that a particular treatment is effective does not necessarily show what the prior causative process involved.

The setting up of the National Institute for Health and Care Excellence (NICE) in the UK to review the scientific evidence on the efficacy of different treatments (see Layard and Clark, 2014) constituted an important advance. Initially, their reports focused on the use of

pharmacological treatments in adults but coverage has been extended to children and to a wide range of psychological therapies. These have had an important clinical impact because, for the first time, clinicians were provided with evidence-based recommendations for what worked best for what problems (although we need to appreciate that their recommendations take costs into consideration) (see Sackett et al., 1996). With respect to psychological therapies for adult mental disorders, it is clear that cognitive behavior therapy (CBT) has the most evidence for its efficacy for most disorders (see Beck & Dozois, 2014). In striking contrast to the development of pharmacological treatments, the development of CBT has been based on a substantial theoretical and empirical framework.

With respect to CBT for children's mental disorders, several points need emphasis. First, not everyone responds to treatment. Second, one small study found that with respect to outcomes the training of the therapist matters (see , for example, Ginzburg et al., 2012). Unfortunately many of the professionals providing psychological services have not been adequately trained in the methods they use. Third, it also appears from the findings on relapse that children, perhaps rather more than adults, often need booster sessions in order to maintain the gains (Layard & Clark, 2014). Finally, there is a major problem in the UK of both a limited access of children to services and rationing in terms of the number of sessions that are permitted. Although the situation is vastly better than it was half a century ago, it is far from satisfactory.

The use of stimulants and then tranquilizers to treat disruptive behavior developed by Bradley (1937/8) constitutes one of the earliest studies of pharmacotherapy in childhood. Many of the early studies did not use RCTs but their use had become established by the 1960s (Lipman, 1974). However, what was striking about the drug studies in child psychiatry/psychology was that they involved attention to important issues beyond efficacy. Thus, Rapoport and colleagues (1978) showed that stimulants did not have a

'paradoxical' effect in attention deficit hyperactivity disorder (ADHD). Rather the effects in typically developing individuals were qualitatively similar to those found in people with ADHD. Another example is Eisenberg et al.'s RCT finding that the use of medication to treat disturbed behavior in a training school for delinquent boys (see Eisenberg et al., 1963; Molling et al., 1962) led to the observation that the exclusion of some boys from the study (because they were not part of the RCT) led to a worsening of behavior whereas use of a placebo did not. Those outside the RCT knew that they were not being given any treatment whereas those in the RCT knew that they were getting treatment, although they did not know whether they were receiving the active drug or a placebo. The paper was not explicit on the reason why some people were excluded from the RCT, thus leaving open the possibility that selection biases might have played some part. However, it seems dubious whether it would be possible to select a group with worse results than with a placebo.

The use of stimulants to treat ADHD has been long established and the use of selective serotonin reuptake inhibitors (SSRIs) for the treatment of depression and anxiety and obsessive compulsive disorders (OCD) is now common. With respect to validity, several points need to be made. First, with very few exceptions, the drugs shown to be effective were not derived from biological research but rather from serendipitous observations. Second, it has proved much more difficult than expected to discover new drugs. As a consequence, many of the major drug companies have stopped their involvement in the development of drugs for mental disorders. Third, even when drugs have been shown to be effective, it is unusual for there to be much understanding of their multiple pharmacological actions and of the features that bring benefits. Whilst there is no doubt that the availability of effective drug treatments has brought immense benefits, it is striking that the comparative efficacy of different drugs has led to little understanding of mechanisms.

The clinical study of different syndromes and diseases has yielded some gains and left many unanswered questions. Autism provides an instructive illustration. Kanner's identification of infantile autism in 1943 showed what can be achieved by the careful observations of an unusually astute clinician. The follow-up studies undertaken subsequently with Eisenberg (Kanner & Eisenberg, 1956) did much to increase understanding of the nature and qualities of this behavioral pattern. The experimental studies undertaken first by Hermelin and O'Connor (1970) and later by Baron-Cohen, Leslie and Frith (1985) took things forward by their demonstration that mentalizing deficits were likely to play a major role in the liability to autism. Twin studies showed that there was a substantial genetic liability and that this extended beyond the traditional diagnosis to a broader phenotype (see Rutter and Thapar, 2014) providing a further step towards validation. Brain imaging studies (Frith, C., 2003) were crucial in showing that autism did not stem from a localized abnormality in one part of the brain but probably from a lack of normal connectivity across brain systems. Whilst none of this adequately validates the diagnosis of autism, it certainly went well down that road. There is general agreement that autism is likely to be heterogeneous but an agreement on how the subdivision should be done has so far not been established.

Differentiation of Rett's syndrome from autism first shown by Rett in 1966 but more firmly put on the map by Hagberg and his colleagues in 1983 was followed by the demonstration that this was due to a mutant gene. With a really good mouse model it was established that genetic modification enabled apparently irreversible neurodevelopmental deficits to be altered (see Rutter and Thapar, 2014). This provides a clear example of a well-validated syndrome in child psychiatry even though some questions remain on the details.

More recently, research by Viding & McCrory (2015) has done much to demonstrate the validity of the differentiation between psychopathy and conduct disorders in terms of

genetic, cognitive, emotional, biological, environmental and personality features (Frick et al., 2014). However, these reviews also bring out questions on the clinical identification of the syndrome and on the need for research in more appropriate samples. Thus, for example, psychopathy differed from conduct disorder in terms of a higher heritability, and of involving some specific genetic influences. Magnetic resonance imaging findings have also shown that psychopathy is associated with atypical patterns of brain structure and function (Viding & McCrory, 2015).

Whilst there has been quite a lot of progress in the validation of some mental disorders in childhood, validation of others has been much less satisfactory (see relevant chapters in Bloch et al., 2014). This applies particularly in the fields of forensic psychiatry, addictions and personality disorders. Possibly this has been a function of disagreements over whether these are 'true' disorders (whatever that may mean).

In the remainder of this review, we consider the strength and weakness of important areas of research, and some methodological concerns and opportunities that require us to be consistently pressing for the highest standards.

Structure of review

We start with classification and diagnosis because they have usually been the focus for initial discussions of validity. Because of the extensive overlap among diagnostic categories, we consider that the present classification systems constitute a degree of threat to the validity of child psychiatry and psychology. We put it cautiously because there is evidence of some diagnostic validity (see Rutter & Pine, 2015) and because it may be better to consider dimensions rather than categories. Nevertheless, their validity would have to be examined. Throughout this review, we consider biological findings as providing possible

validating evidence. That is not because we consider that they constitute the only sources of validation but rather because it would be foolish to ignore them.

However, before assessing the major sources of biological evidence, we consider the quite different issue of the phenomena that present major challenges but which have received very little biological attention up to now. These are discussed under the headings of clinical findings and developmental influences on mental functioning.

We then turn to a discussion of the possible validating value of genetics, neuroimaging, and biomarkers. In each case we try to balance the strengths and the limitations of the area. We highlight the example of abuse as a cause of mental disorder as one that necessitates attention to the biological embedding of experiences. Because most parts of our review need to consider whether causal inferences are justified, the next section deals with that issue, under the headings of the role of 'natural experiments' and animal models.

Finally we turn to the most important issue of the reliability and validity of both measurement and analysis. We address the problems of undisclosed flexibility in data collection and analysis and note the importance of following CONSORT and STROBE guidelines. Our review concludes with a section, Ways Forward, outlining our recommendations.

Classification and diagnosis

The autism and psychopathy studies were informative on diagnostic classification but DSM-5 (American Psychiatric Association, 2013) was expected to deal more broadly with the issues. Ordinarily, it might be assumed that the question of validity of child psychiatry/psychology should begin with a consideration of diagnostic differentiations but that would not be appropriate now. That is because of the mass of evidence that diagnostic

categories overlap to a much greater extent than assumed in the past (Rutter and Pine, 2015). Thus, in child psychiatry, DSM-IV did not allow both autism and ADHD to be coded because they were thought to be so distinct. It became clear that, to the contrary, a varied mixture of the two is common. This is problematic because it is common for findings to be first reported on the basis of one diagnosis but yet for another paper to report the same for a different diagnosis. The research into copy number variations is a case in point (see section on genetics below). In addition, most diagnoses are, to a considerable degree, dimensional in character. Also, it may well be that it makes more sense to focus on broader groupings. On the other hand, a degree of validation is evident from epidemiological and clinical evidence, as well as biology, indicating that there are meaningful differences among diagnoses. Thus, for example, although autism and ADHD overlap, only autism is associated with the onset of seizures in late adolescence and only ADHD responds to stimulant medication. But sometimes too, validation concerns finer, rather than broader differentiations. Thus, the COMT gene valine/methionine polymorphism is not associated with either antisocial behavior or ADHD, but it is associated with antisocial behavior occurring in individuals with ADHD (Caspi et al., 2008). As Kapur et al., (2012) put it, the diagnostic systems were not designed to facilitate biological differences and they do not; but, equally the biological studies have not led to a clinically viable alternative. Similarly, Hyman (2014) described diagnostic categories as 'transiently useful fictions'. Kendler (2014) who chaired the scientific review committees for DSM-5 broadly agreed but made two additional points. He stated that no major biological breakthrough should be expected; rather a modest step by step approach should be followed. Second, the way DSM criteria are used in practice creates a major reification problem. The criteria may be useful to index disorders but they should not be viewed as direct measures of any diagnosis.

Many had hoped that DSM-5 would resolve the main problems but, sadly, it failed to do so (Rutter and Uher, 2012). Thus, there is still an enormously long list of diagnoses (as there was in both DSM-IV and ICD-10 despite many being rarely, if ever, used). There is little to be gained from a discussion of its limitations. Instead, we seek to adopt a problem-solving approach in discussing the ways forward. As Collins (2010) argued, the task is not to predict the future, but to enable it.

Neglected areas of research

Clinical findings

Both funders and researchers in recent times have moved away from clinical research to hypothesis-free designs driven by technology. This coincided with a relative neglect of very important clinical findings that are relevant to the validity issue but on which we have very little understanding of the biological basis. For example, Howlin et al., (2009) showed that about a third of individuals with autism had exceptional cognitive or savant skills. This was a much higher rate than had been previously considered. What we do not know is why these appear to be more frequent in individuals with autism and what is the biological basis for the phenomenon, including why autism is associated with both severe general cognitive impairment and also these positive splinter skills. Further research needs to be undertaken but it needs to be with a firm focus on the biology.

Much the same applies to the frequency of temporary developmental regression involving language loss, which Pickles et al., (2009) showed applied to over a quarter of individuals with autism but to only a tiny handful of children with specific language impairment associated with other causes. Again, the need is to investigate the biology underlying this phenomenon and the reason why it is particularly associated with autism.

A third example concerns the finding by Bolton et al., (2011) that the age of onset of seizures in individuals with autism is at a mean age of onset of 13 years. The rate of autism is not particularly high in comparison with other neurodevelopmental disorders but the later age of onset does differ from that found with either individuals in the general population of normal intelligence or in individuals with severe cognitive impairment. That must have some neuropathological significance but it is unclear what it is.

A fourth example would be the lack of response of core symptoms of autism to medication (see McCracken, 2011; Mohiuddin and Ghaziuddin, 2013). Although the same may apply to eating disorders, it is unusual among multifactorial mental disorders. Thus, many (but not all) cases of schizophrenia respond to psychotropic medication, depression responds to antidepressants, obsessive-compulsive disorder responds to clomipramine and SSRIs. Why is autism different? Does this mean that autism does not arise on the basis of any kind of neurotransmitter function, or have the wrong drugs been tried so far?

The fifth example is so-called "quasi-autism" (see Rutter and Sonuga-Barke, 2010) which arises in a substantial minority of individuals experiencing profound institutional deprivation in the early years of life. When the children are young it seems similar to ordinary autism, but there is a more flexible social approach and a greater prominence of circumscribed interests. As the children get older some features become more like those associated with disinhibited attachment, but there is a high persistence of problems and a heavily reliance on service usage. It has been suggested that there are some parallels with congenital blindness which also shows somewhat similar patterns. Does this mean that autism can arise as a result of externally imposed sensory restriction, as well as inherent deficits in processing of stimuli?

The last autism example to be given concerns the finding from several different studies (for example, Lichtenstein et al., 2010; Pourcain et al., 2011) that ADHD and autism overlap in their features to a very considerable extent. The overlap is by no means restricted to these two syndromes, however. Simonoff et al., (2008) found multiple varied co-occurrences between autism and other forms of psychopathology in their population-derived sample.

A rather different clinical feature that requires study is the high frequency of sex differences in rates of disorder. On the whole, those showing a female preponderance concern disorders beginning in adolescence in which the main pattern is of either emotional disturbance or eating disorders. By contrast, those with a marked male preponderance are predominantly disorders beginning in early childhood in which neurodevelopmental impairment is a frequent feature (see Thapar and Rutter, 2015a). The importance of these major sex differences is well-established but their meaning is not well understood. Also, the early onset disorders found more often in males are by no means a homogeneous group of syndromes and their biological basis may well vary across diagnostic groups (see Thapar and Rutter 2015a). Similarly it cannot be assumed that emotional disturbance and eating disorders have the same biological basis.

The challenges look equally great if viewed from the other end. That is to say, why is there such non-diagnostic specificity in the disorders associated with large rare copy number variations (CNVs). These have been found to be associated with autism (Sebat et al., 2007), schizophrenia (International Schizophrenia Consortium, 2008) and ADHD (Williams et al., 2010) as well as intellectual disability. All of these groups differ from controls whereas this has not been found for dyslexia. For the most part the findings have been reported as if the main outcome was the specific syndrome but closer examination of

the findings indicates that there is not a standard phenotype. The question is why is it so varied?

Developmental influences on mental functioning

A close parallel to the clinical areas neglected in research concerns the equal, if not more marked, neglect of developmental effects. For example, Rapoport and colleagues (1980) noted the addictive effects of stimulants in adults where euphoria was often experienced. But this was rarely found in childhood where the experiences tended to be dysphoric. Children with severe ADHD benefited from the stimulants but they did not like the feelings that were engendered by the drug. More recently, it has been found that whereas depressed children do respond to SSRIs, they do not respond to tricyclics (Hazell et al., 1995). This is curious because other research has shown a substantial continuity between depression in children and adolescents and the occurrence of depressive disorders in the same individuals when they reach adult life (Fombonne et al., 2001; Harrington et al., 1990).

More recently, there has been a growing awareness of the serious metabolic side-effects associated with psychotropic drugs such as risperidone and olanzapine (Arango et al., 2014). Over a six-month period, people aged 4 to 17 years showed an average increase of 7 kilograms of weight following risperidone and 11.5 kilograms of weight with olanzapine. It seems that these serious metabolic side-effects of psychotropic medication are greater in children than in adults (Stigler et al., 2004; Fedorowicz & Fombonne, 2005). There is a resulting appropriate caution with the use of these drugs in children but the reason for this developmental moderation of drug responses is quite unknown and it much needs systematic study.

The cannabis effect of increasing the risk for schizophrenia has been found only with heavy use beginning before adult life (Arseneault et al., 2002; Caspi et al., 2005). That this is

likely to be a true causal effect is shown by its drug specificity (neither heroin or cocaine use led to schizophrenia despite their other harmful effects); by the finding that the experimental provision of cannabis induces psychotic experiences (Henquet et al., 2005, 2008); and by the finding that the risk for schizophrenia is evident only when the cannabis use begins before adulthood. Although an animal model of schizophrenia has not been developed, the effects of cannabis products on cognition have been studied in rodents. As in humans, the effects are much more marked in adolescent rats than in adult ones (Schneider and Koch, 2003; Trezza et al., 2008). As with the other findings, we do not understand the biological basis of this age moderation of cannabis risk effects.

Lastly, there is the finding that aphasic-like symptoms do not arise following lesions in the dominant hemisphere in young children, whereas they are dominant in adolescents and adults (Goodman and Yude, 1997; Vargha-Khadem et al., 1992). At one time this was thought to mean that young children were less affected by brain lesions than adults – the so-called Kennard principle – but that is mistaken. The difference actually goes in the opposite direction. On the whole, children are more vulnerable to brain injuries than adults but they do not show a lateralized pattern of aphasic features. The biological basis has yet to be identified.

We appreciate that there is a lack of coherence in these developmental examples but therein lies the need for research to examine possible mechanisms. Berens and Nelson (2015) reviewed studies of serious early adversity and concluded that the evidence indicated some form of sensitive period and that this suggested the need for early intervention. Whether or not all the examples reflect a similar sensitive period effect is not known and the biology of sensitive periods has still to be established, although epigenetic mechanisms are likely to play a key role (Boyce & Kobor, 2015).

Validating possibilities from biological studies

Genetics

The field of genetics is rapidly changing and it would not be appropriate to attempt any overall discussion of the advances. Rather, we focus here on those that have particular implications for biological validation of child psychiatry/psychology. Most attention in the literature has been paid to the availability of DNA micro assays which allow the genotyping of millions of SNPs (single nucleotide polymorphisms) and their use in genome-wide association studies (GWAS). Unlike candidate gene analyses, these provide a hypothesis-free coverage of the whole genome. Because of this, they enable novel genes to be discovered. The findings on schizophrenia (based on a pooling of several consortia providing nearly 37,000 cases and over 113,000 controls) constitute an example of this research at its best (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). One hundred and twenty-eight genes reached genome-wide significance, and eighty-three of these had not been previously reported. The meaningfulness of the findings was indicated by the finding that the genes particularly involved not only glutamine (connected with schizophrenia by other research findings) but also immune functions. In keeping with GWAS research into other behavioral traits, the effect sizes for individual genes were tiny (Rietveld et al., 2014) and still quite small when the genes were combined to produce a polygenic score. The large sample enabled replication. The enthusiasm for GWAS has to be limited, however, because the research strategy provides no evidence on what the genes 'do' (with respect to effects on proteins). There has to be reliance on bioinformatics data on human and animal findings on protein effects – such as in the neuroscience database provided by Croning et al (2009). However, there is, as yet, no agreement on the requirements to specify what is meant by a biological pathway. The situation is likely to improve in the future but at present it is unsatisfactory.

The combination of genes to provide a composite polygenic score has been seen by some as a useful way forward for clinical prediction (eg Hodgson et al., 2014; Iyegbe et al., 2014). The idea is sound in that multifactorial disorders (this applies to most of child psychiatry) involve the operation of multiple causal influence – both genetic and environmental. However, the clinical value of polygenic scores has yet to be established. In addition, there is, from a biological pathways perspective, the huge problem that there is no way of telling which genes in the polygenic index are having particular effects. It cannot be assumed that they will all act on the same pathway. A further problem of all uses of GWAS is that findings are consistent in showing that most genes are pleiotropic – meaning that they have multiple, often diverse, effects. That may well increase the difficulty in identifying biological pathways.

Another important genetic advance was the detection of copy number variations (CNVs) – meaning submicroscopic deletions or substitutions of stretches of DNA that may span multiple genes (Thapar and Cooper, 2013; Malhotra and Sebat, 2012). CNVs may arise either de novo or be inherited. Three key findings need emphasis. First, although CNVs are significantly associated with mental disorders they have probabilistic, and not determinative, effects. That is although the rates of CNVs are higher in individuals with a disorder, they are also seen in individuals without psychopathology and may show incomplete penetrance. Second, the effects are diagnostically non-specific (being found for example in association with intellectual impairment, autism, ADHD and schizophrenia). Third, CNVs are particularly associated with disorders having major neurodevelopmental origins and CNVs seem to be associated with synaptic dysfunction. That is, the findings provide clues (but not proof) of biological pathways.

Some disorders have been associated with rare pathogenic mutations such as SHANK 3 and the neuroligins (Sakurai et al., 2011). These may act in a determinative

fashion but the particular rare genes vary across different individuals (Pinto et al., 2010) making it unlikely that they will aid biological understanding more generally.

The biological validation is slightly stronger in the case of gene-environment interaction (G x E) because of the availability of human and animal experimental findings (see Byrd and Manuck, 2014; Caspi et al., 2010; Dodge and Rutter, 2011; Karg et al., 2011; Rutter, 2008). The epidemiological findings have been criticized because of the apparently inconsistent replication, but Caspi et al., (2010) have pointed to the biased samples used in the critiques. It is noteworthy that the two studies using the Liptak-Stouffer test (LST) for pooling p values confirmed G x E with respect to both the 5HTTLR gene and the MAOA gene. The LST has the advantage over the more traditional pooled effect-size approach of coping better with heterogeneity (see Byrd et al., 2015) but the disadvantage of being unable to quantify the strength of the G x E. This is not a substantial limitation because, like heritability, G x E effects are population-specific and will vary with the range of environments and developmental change (see Kendler, Jaffee and Romer, 2011). Indeed, there will be no statistical G x E in situations in which there is little or no variation in E (as obvious in the case of phenylketonuria because phenylalanine is present in all usual diets). The Karg et al., (2011) findings were important in showing that the G x E applies to maltreatment but only very marginally, if at all, to acute life events. We conclude that the evidence for G x E is overwhelming (because of the human experiments and the animal models) but many questions remain including those as to the range of conditions under which it operates that may limit its relevance for public health and therapy.

Epidemiological studies such as the Caspi and Moffitt Dunedin longitudinal study provide several striking examples of G x E, but cannot either prove causation or identify the biological pathway. Experimental studies such as those by Weinberger, Hariri and Meyer-Lindenberg are informative on both, although so far they have not shown the biological

pathway(s) that bring G and E together. The imaging findings on G x E are also striking in their indication that the findings apply to individuals without psychopathology, and not just to patients. That makes any suggestion on their validation of child psychiatry/psychology very indirect. A further twist is provided by the growing evidence that the individuals most genetically vulnerable in the presence of adversity may also be most responsive to therapeutic interventions (differential susceptibility – see Ellis and Boyce, 2011).

It is evident that there have been spectacular advances in psychiatric genetics but it is equally apparent that they do not provide a biological validation of child psychiatry/psychology. Their greatest success has been in the field of cancer genetics where the genetic analysis of the tumor tissues has enabled differentiations that now shape both the subclassification of cancer and its treatment. Because brain tissues are not available in life, there has to be reliance on postmortem brains (see McGowan et al., 2008 for its use in examining methylation patterns). It is also relevant that the brain is a much more heterogeneous structure than other organs, and that function often depends on coordination between different regions. The contrast with cancer is extreme and therefore arguments based on analogy with it often misleading.

In recent times, there have been attempts to circumvent the problem by analyzing either blood products or buccal cells on the rationale that, compared with other tissues, expression is particularly high in the brain. Mostly, this has been tested by examining the reliability coefficients between expression findings in brain and blood (or buccal tissues). It is not likely, however, that the reliability will be high because it varies between different blood cells. Rather, we suggest that the peripheral tissues be treated as biomarkers in which the issue is whether the overall patterns found are similar to those found with the brain (and not with individual reliability). (See Boyce & Kobor, 2015).

Neuroimaging

Over the last two decades, there has been much excitement over the possibility that neuroimaging might do much to provide an understanding of the biological basis of behavioral traits or disorders (see Pelphrey et al., 2015). Whilst the potential is undoubtedly there, there are substantial problems that have to be recognized and dealt with. There are many different types of brain imaging and, as is the case with most areas of science, each has its own particular strengths and limitations. That means there is much to be gained by combining two or more neuroimaging approaches.

The possibilities of the gains in understanding produced by neuroimaging are well illustrated by the comparison of infant neural sensitivity to dynamic eye gaze in relation to behavioral correlates. Elsabbagh et al., (2012) hypothesized that neural sensitivity to eye gaze in early infancy would predict the later development of ASD. Strikingly, what they found was that evoked response potentials (ERPs) indexing face perception showed that ERPs recorded at age of 6-10 months did indeed predict the development of autism at 36 months, although behavioral measure of eye gaze did not.

Here we focus on challenges to be overcome. First, early claims with respect to fMRI argued that brain imaging showed the brain in action but it is clear that it does so only indirectly. For example, functional magnetic resonance imaging (fMRI) relies on the measurement of oxyhemoglobin measures but these only indirectly reflect brain activity (see Logothetis, 2008). Also, it will not reflect deactivation to stimuli. The second problem is that of multiple comparisons. They carry with them the strong likelihood of a chance finding. This has been systematically addressed by neuroimaging researchers with some success. Third, children are more likely to be restless and move about while being scanned and this can lead to motion artifacts and it is evident that this can lead to quite misleading

conclusions about the interconnectivity across brain regions (Power et al., 2012; van Dijk et al., 2012). Fourth, different brain imaging techniques vary in their temporal and spatial resolution. Thus, fMRI has good spatial resolution but poor temporal resolution whereas techniques based on the EEG, including magneto- encephalography (MEG) have good temporal resolution but poor spatial resolution. Also, the EEG is constrained by the focus on what is happening at the surface of the brain, rather than deep inside. Fifth, all functional imaging is highly dependent on appropriate experimental design and on the interpretation of findings. Pelphrey et al., give the example of regional differences in activation to faces. Although rarely dealt with in published papers it is necessary to ask what was the control condition and whether there is deactivation when external stimuli are presented. In that connection, it is necessary to be aware of the default mode network (see Snyder and Raichle, 2012) and the need to examine both activation and deactivation in response to stimuli.

Pfeifer and Peake (2012) have argued that new brain imaging approaches when used in conjunction with well understood paradigms and measures can do much to establish the validity of the empirical findings that they are designed to validate. We agree that brain imaging constitutes a valuable tool for validation but, so far, its achievement do not live up to the claims and its promise. The future probably lies in the combination of different forms of imaging and appropriate experimental design.

Biomarkers

The interest in the development of good biomarkers arose because both clinical diagnosis and risk characteristics for mental disorders rely on behavioral judgments that are open to biases and uncertainties over both reliability and validity. Biomarkers cannot, in themselves, define the pathophysiology of the conditions for which they are a marker but, if successful, they could enable a conceptualization that is much closer to the biological nature

of whatever condition is being considered, but not necessarily to the genetic liability (see Rutter 2014).

Much of the research on biomarkers has used straightforward intergroup comparisons without paying much attention to the assumptions that need to underlie the study of biomarkers. Obviously, strong connections between the biomarker and the phenotype are crucial and group comparisons will do that but there is a need for good data on the reliability and validity of the biomarker in relevant populations. Thus, if the focus is on differentiations among different forms of mental disorder, the comparisons must involve multiple disorders and not just comparison with a control group of some kind (see also Kapur et al., 2012). In addition, there is the crucial need to determine whether the inferences based on biomarker findings are robust and valid at an individual level. A crucial feature of multifactorial causation is that the specifics of the mix between multiple genetic and multiple environmental factors will vary from individual to individual suggesting many biomarkers may lack predictive power for an unknown proportion of individuals. From a practical point of view, it is essential to determine whether biomarkers do a better job in predicting the phenotype of interest as compared with traditional clinical measures. In addition, cross-validation is essential.

There may be advantages in making diagnoses on the basis of biological findings rather than symptom patterns because they might provide understanding. Nevertheless, we must return to the point that biomarkers cannot in themselves define the pattern of physiology for the conditions for which they are a marker.

Calhoun and Arhabshirani (2014), using schizophrenia as an example, argued for automatic diagnosis based on support vector machine learning. In essence this involves an iterative approach in which variables are retained or dropped from the algorithm. The

method constitutes a highly sophisticated form of discriminant function analysis, which means that there is a risk that it capitalizes on random error. We conclude that it is a potentially useful tool but its value in the field has still to be established.

Validity of abuse as a cause of mental disorder

One of the major changes that have taken place over the last few decades is in the increase in awareness of the clinical importance of physical and sexual abuse of children. With respect to the topic of this review, a key question has to be whether there has been validation of the important role of abuse in the liability to child psychopathology. That has been tackled firstly by testing the causal inference from the associations found; that is, are the effects environmentally mediated? Genetically sensitive designs were needed to test for environmental mediation. Jaffee et al., (2004) used a multivariate twin analysis for this purpose and showed that the effects of physical abuse were indeed environmentally mediated. Kendler and Prescott (2006) used a discordant twins design with sexual abuse and similarly found an environmental effect. Arseneault et al., (2008) used the same type of design with bullying and, again, found environmental mediation. The second step with respect to validation refers to the somewhat different question of the biological consequences of these experiences. Ouellet-Morin et al., (2011 and 2013) showed that bullying led to increased SERT methylation and a blunted cortisol response to stress. A third step, therefore, is to consider whether the sequelae of abuse stem from the physical acts themselves or the numerous adversities (including neglect and emotional abuse) with which they tend to be associated. We are not aware of studies that have examined this matter thoroughly. It could be that single, isolated acts of physical abuse are less likely to have lasting psychological sequalae but abuse usually involves multiple acts and several varieties of abuse. There is an abundance of evidence that abuse and other adverse experiences have major biological consequences (see Rutter & Azis-Clauson, 2015, for a review). However,

the biology has proved to be much more diverse than originally envisaged, although epigenetics and HPA (neuroendocrine) effects have tended to predominate in the literature.

Whilst accepting the importance of both epigenetics and neuroendocrine effects, one of us (Rutter, 2012) has pointed out the numerous unanswered questions that remain. It is not that there is any doubt about the overall validity of the environmental effects but there are questions about whether, for example, epigenetics can differentiate between different forms of environmental adversity (say, between the consequences of acute and chronic stress). Can epigenetics explain individual differences in response (such as between sensitizing and steeling effects) and does it account for variations in psychopathological outcomes? See Krishnan et al., (2007) as a useful start in studying molecular adaptations in mice. For example, epigenetics might account for the HPA effects but the greater proximal effect on mental disorder might come from the neuroendocrine changes. In other words, there is a good deal of validation of the overall notion of the importance of environmental influences on psychopathology but much has still to be done in order to validate the specifics.

We discuss the important constraint provided by the fact that a great majority of psychiatric disorders are multifactorial and we note that the problems that this provides apply to disorders of all kinds and not just psychopathological ones. What is additionally distinctive about mental disorders is that, unlike most somatic disorders, neuropathology cannot be directly studied during life. Finally, we note that there are many examples where the failure to make much progress with validation stems from concentration on technology rather than the important clinical and developmental questions that require attention. It is sometimes supposed that the statistical advances have made all of this very much easier than used to be the case, but in considering the study of mediation, we note the limitations that still exist.

Reasons why a statistical association might not mean causation

Although everyone is aware of the fact that it is not permissible to assume causation solely on the basis of a statistical correlation or association, a high proportion of studies in the literature fail to consider whether causal inferences are justified (Rutter, 2007; Foster, 2010) Some researchers argue that they are dealing with statistical associations and not causation, although why anyone should be interested in the findings if they carry no implications for causation is a mystery. Others make explicit that the research cannot test causation but, nevertheless end the paper with suggestions on causation. Of course, with multifactorial disorders, there is no such thing as a single necessary and sufficient single cause. This is because it is usual for there to be constellation of causes acting in concert (McMahon et al., 1960; Rothman & Greenland, 2002); furthermore it is usual for several causal pathways to lead to the same end point and because a single starting point may lead to a diverse range of outcomes (Cicchetti and Cohen, 1995). The most important requirement is to consider alternative explanatory hypotheses (Cochran and Chambers, 1965; National Research Council, 2002).

The main alternatives to a causal association are: a genetic mediation of the environmental risk feature; social selection or allocation bias; reverse causation; misidentification of the risk feature; the logical problem of having to use an invariant set of associations to infer what would happen if a key association was different or changed (McKim & Turner, 1997); and finally, there is the even greater problem presented by uncertainties regarding unmeasured confounders (Academy of Medical Sciences, 2007).

Reliability and validity of measurement and analysis

Over the many years during which psychiatric classification schemes were being developed, there have been numerous concerns over the very modest chance corrected

agreement (i.e. kappa rather than percentage agreement) statistics for both inter-rater and retest reliability of psychiatric diagnoses (see Freedman et al., 2013; Regier et al., 2013). It is not just that this occurred in the past; it was found also with the DSM-5 field trials using specially trained clinicians. For example, with respect to child diagnoses, the kappa value for major depressive disorder was .28, and for conduct disorder it was .46. These low values compare unfavorably with those obtained for most research interviews. The key difference is that the research interviews provided standardization of the interview itself and not just the diagnosis.

Improving generalization through pre-specification and standardized reporting

Simmons, Nelson and Simonsohn (2011) have shown, through both simulations and actual experiments that undisclosed flexibility in data collection and analysis leads to a high likelihood of false positive findings. It had been concerns such as these (which are far from specific to psychiatry and psychology) that led to the development of standardized and structural reporting of treatment trials as exemplified in the CONSORT guidelines (www.consort.statement.org) and, later, of observational studies as exemplified in STROBE (www.strobe-statement.org). Both were widely adopted, but it seems that the mental health field (both child and adult) is not adequately adopting the guidelines (Cristea et al., 2015; Patel et al., 2015). It is notable that, in addition to partial reporting and lack of statistical power, there may be problems in publication bias and a failure to report contradictory findings. The Simmons et al (2011) paper, recognizing that the needs extended beyond trials argued for six requirements: 1) a predetermined rule (specified in the paper) on when data collection should be terminated; 2) collection of at least twenty observations per cell; 3) report of all variables collected in a study; 4) report of all experimental conditions including failures; 5) if observations are eliminated, a report on analyses of results with them included; and 6) if an analysis includes a covariate, report the statistical results without the covariate.

Issues particularly arise when examining heterogeneity (because it is clear that usually there is heterogeneity). The risk is that 'data dredging' will lead to a misleading artefactual identification of heterogeneity. This may be prevented by considering heterogeneity in advance if such comparisons are to be made. However, often the comparisons could not have been planned in advance. For example, Collishaw and colleagues (2010; Schepman et al., 2011; Sellers et al., 2015) used several of the British national cohorts to compare time trends in child and adolescent mental health across 1974, 1986 and 2008. The interest was not just whether the rates of mental health problems had increased or decreased over time, but why there had been changes. The different cohorts had been studied by different researchers and a series of methodological steps to examine possible artefacts was essential. Similar issues applied to the comparisons between the Romanian, Greek and Chinese follow-up studies of children who received early institutional care, where there were strikingly different outcome findings (Rutter & Sonuga-Barke, 2010; Vorria et al., 2015; Feast et al., 2013; Rushton et al., 2013). The standardized reporting rules could not be applied without modification, but exactly the same principles applied if bias was to be avoided.

Meta-analyses and cross-validation

Meta-analyses have become the accepted method for establishing consensus in multiple studies addressing the same issue (Cumming, 2014). However, for these to serve their purpose, the studies included must be valid (Kraemer, 2015). Unfortunately too many include a heterogeneous mix of RCTs and observational studies, cross-sectional and longitudinal studies. For example, this would apply to the meta-analyses on institutional care (see Bakermans-Kranenburg et al., 2008). Treatment studies often need to compare across a range of different treatments. Whereas conventional meta-analyses cannot do that,

techniques are now available to bring together multiple treatments in a single meta-analysis (see Leucht et al., 2013).

In all evidence synthesis, care needs to be taken regarding the quality of the studies to be included. For example, Uher and McGuffin (2010) noted that the meta-analysis reported by Risch et al., (2009) was heavily skewed by the inclusion of very large scale studies using very weak measures. There is a real danger of the garbage in – garbage out problem. Murray (2014), commenting on the Schizophrenic Library study by Matheson et al (2014) noted that some meta-analyses are clearly biased in favor of the researchers' own views. As a consequence, meta-analyses by enthusiasts and sceptics often conflict.

With respect to improving the generalizability of single studies, a possible step forward is provided by internal cross-validation (Steyerberg, 2009) in which prediction performance is assessed on cases not included among those used to estimate prediction parameters. As the National Research Council (2002) in the US recommended with respect to educational research, there is an obligation on all researchers to try to disprove their own hypotheses before publication.

Mediation and moderation

With respect to both observational studies of risk and intervention studies it is a matter of major interest to determine which elements accounted for the risk or protection effects. The usual starting point is a strong direct effect (best measured by a path coefficient – Mackinnon and Fairchild, 2009) of the hypothesized causal variable. The mediation analyses (again using path coefficients) test the effects of the supposed causal variable on the postulated mediator, the effect of the mediator on the outcome and the overall indirect path going through the mediator. Finally, the analyses test whether the residual direct path after mediation had dropped to a non-significant level. The original

Baron and Kenny (1986) approach focused on complete mediation but modern concepts of causative processes indicate that a single total cause is highly unusual. Accordingly, partial mediation is to be expected. See Kumsta et al., 2010 for a discussion of risk, causation, mediation, and moderation in a study of institutional deprivation. There is no doubt that the aims of mediation analyses are sound but it has become clear that the statistical demands are high. Thus, even with data from a randomized trial, account needs to be taken of variables that could confound the mediator-outcome relationship (Emsley et al., 2010) and of measurement error in the mediator (Pickles et al., 2015). However, it would be a mistake to assume that any single mediation analysis, however thorough, is enough. We agree with Kazdin (2007) who argued that a range of criteria, spanning several studies are needed and here the Bradford Hill guidelines remain useful.

Moderation is a different concept in that it focuses on heterogeneity of effects, with the aim of identifying the causes of the heterogeneity. The tradition of RCTs was to rule out subgroup analyses because of a few false positive findings but it is now accepted that there is often true heterogeneity and it is crucial to determine its causation. Internal cross validation (discussed above in the previous section) should do much to reduce the risk of false positives. In addition, the notion of mediated moderation has become popular because the allocation of a mediation model to the identified moderator serves to examine the hypothesized causal process. See Manning and colleagues (2014) as an illustrative example in relation to maternal sensitivity as a possible protective factor for children exposed to interparental violence.

Testing causal inferences

Natural Experiments

'Natural experiments' constitute one important way forward (see Rutter, 2007; Thapar & Rutter, 2015b). Some commentators argue that true experiments (including RCTs) are always the best option but that is going too far (see Shadish, Cook and Campbell, 2002). Equally it is sometimes claimed that 'natural experiments' suffer from all the limitations of observational studies and hence can never satisfactorily test the causal inference. However, 'natural experiments' vary in the extent to which that is a problem and a skilled attention to design features can do much to make the causal inference more likely – particularly if several different types of 'natural experiment' are brought together. In essence, rather than relying on statistical controls, they capitalize on naturally occurring circumstances that serve to pull apart variables that ordinarily go together. They do not constitute a finite set of designs as new examples are being found all the time. Also, each 'natural experiment' has its own limitations and conclusions need to be based on circumstances in which several different 'natural experiments' that differ in their strengths and limitations all point to the same conclusion. The reviews cited above discuss the field more extensively; here we confine attention to a few examples in which the findings changed views.

First, several non-experimental studies have shown that exposure to prenatal maternal smoking predisposes to ADHD and antisocial behavior (Linnet et al., 2003). Thapar and her colleagues (Rice et al., 2009; Thapar et al., 2009) used an ingenious assisted reproductive technology (ART) design to test this inference, The design provided a good test of environmental mediation because some varieties (such as sperm donation) retain the genetic link between mother and child whereas others (such as egg donation) do not. Their findings showed that the effects of maternal smoking on ADHD and antisocial behavior were found only when the genetic link was maintained – strongly suggesting that the apparent environmental effect of maternal smoking was an artefact of genetic mediation. Obel and his colleagues (2009) used the completely different design of comparing siblings

when one was exposed to maternal smoking and the other was not. They found no environmentally mediated effect of maternal smoking on ADHD. It should be noted that both these designs did find an environmentally mediated effect on birth weight (an effect shown in all designs). In both cases, it was clear that statistical control for confounding failed to deal with the confounds. It is necessary to ask whether there was anything unusual about women who used egg donation from those who used sperm donation. Similarly, why did women smoke in one pregnancy yet not in another one? The available data provided no indication of relevant biases but the possibility remains.

A rather different example is provided by Jaffee et al., (2004) who used a prospective longitudinal design to compare the effects of corporal punishment and physical child abuse on children's propensity to engage in antisocial behavior. Before their report, most people had assumed that corporal punishment and physical child abuse were lesser and greater varieties of the same risk factor. Their findings showed that this was not so. Environmental mediation was confirmed in the case of abuse, but not for corporal punishment in which the mediation was largely genetic (implying that the effect probably mainly occurred as a result of the parental response to the child's genetically influenced behavior). However, the story did not end there. The availability of longitudinal data showed that the regular use of severe physical punishment did involve a propensity for an escalation into child abuse. The multivariate twin analyses clearly showed the need to differentiate between corporal punishment and physical child abuse, but the longitudinal data brought the two back together in terms of an escalation effect.

The third example illustrates the value of combining a natural experiment with biological data. The effects of the Dutch famine in World War II were associated with an increased risk for schizophrenia in those exposed prenatally (Stein et al., 1975). The causal inference was strengthened by a parallel effect on congenital malformations (McLellan et al

2006) but much strengthened by the epigenetic finding that an individual's prenatal exposure to famine differed from their unexposed same-sex siblings with respect to methylation of the locus for the insulin-like growth factor 2 (Heijmans et al., 2008; Tobi et al., 2012). The causal influence is also strengthened by the gestation period-specificity and the diagnostic-specificity.

Animal models

Animal models do involve true experiments. Their main limitation concerns queries over their generalization and relevance to the human situation. Thapar and Rutter (2015b) concluded that the best animal models had their origins in clinical observations or theories that were supposed to deal with the clinical features. We illustrate this with findings from three studies that either changed or modified the human findings.

First, there were the Rosenzweig findings (Rosenzweig et al., 1962) on deprivation and stimulation in rats. The early results were done at a time when it was assumed the effects would only be found at a time of maximal brain growth. The later findings compared juveniles and adults, with the striking finding that effects were broadly similar in adults as in juveniles although the effects were greater in the young (Rosenzweig and Bennett, 1996). This led on to human studies that similarly showed that experiences in adult life had effects, not only on behavior, but also on neural structure and formation.

The second example concerns the already discussed (see page 14) testing of age-moderated effects of cannabis.

The third example concerns the work of Levine (1957) in showing that intermittent foot shocks in rodents actually lead to a diminished susceptibility to later stress and a greater size of the neuroendocrine system (Levine, 1962). This, in turn, led on to a later experimental study by Lyons et al., (2010) showing that protective effects of brief

intermittent mother-infant separations in squirrel monkeys. In conjunction with other data this led to the conclusion that exposure to challenge and manageable stress in humans can enhance resilience (Rutter, 2013).

Until recently the potential validity of an animal model depended on the face-validity of the animal behavioral response as functionally analogous to the human response of interest. The increasing scope to examine whether the biological response pathway is similar across both animal and human provides substantial additional reassurance of the relevance of the animal work. The work by Battaglia et al., (2014) showing that carbon dioxide had a similar physiological effect in humans and mice is an example.

Of course, like human studies, animal models involve a high risk of false positive findings if small samples are used. The best studies not only have adequate-sized samples but also include a range of other methodological steps such as including random allocation to groups (see Conti et al., 2012 for a primate example).

One of the poorly highlighted features of much animal work has been that some studies report only the results for male animals. This is because the response of female animals might be affected by reproduction cycles, so complicating interpretation. The likely consequence is that potentially important sex differences may have been overlooked. However, as Hostinar et al., (2014) discussed, the lack of a proper study of sex differences and their meaning applies to both humans and animal models.

Ways forward

There are two critical longstanding limitations that are proving particularly challenging. First, it is not ethically feasible to examine brain tissues during life. This makes the situation in psychiatry very different from that in, say, cancer where the biological findings have had a huge importance in relation to both prognosis and treatment. This

limitation is all the more restricting because the brain is substantially more complex (with respect to cell variety, regional specificity and interconnectedness) than say the lung or liver. Second, the empirical findings are clear in showing that the overlap among diagnoses continues to be much greater than previously assumed. That means that we must continue to avoid reification of current diagnostic systems and will need creative new thought on how to deal with uncertainty in the diagnostic system. One possible way forward lies in a focus on neural networks based on gene functions rather than physical gene proximity (see Chen et al., 2015) and this will clearly require skilled bioinformatics collaboration (see Mostafavi and Morris, 2012).

As we have discussed, biological findings are likely to be helpful in this connection, but a step-by-step approach putting together clinical and biological findings will be needed. Concerns have been expressed that too many research findings fail to be replicated. In part this is a function of lack of adequate statistical power in much research in psychiatry and psychology (both child and adult) but it also stems from a failure to deal with the various sources of possible bias. The CONSORT and STROBE guidelines were introduced to begin to deal with these and we strongly recommend their continuing use together with full transparency on what analyses were undertaken and why. However, we are most concerned that it seems that these guidelines are sometimes being ignored; that will have to be corrected. It is good that there is an increasing recognition that an unregulated flexibility in the handling of data analysis carries with it a major risk of false positives (Simmons et al., 2011). Nevertheless, we worry that pressures to publish mean that the aim becomes publication in a prestigious journal and the higher rates of citation for some claimed result, rather than securing some result that is valid and honest and will be confirmed by others (see Nurse, 2015).

Standardization, such as represented by preplanned requirements to specify the approach to data analysis, have been very valuable but we emphasize the need also to adopt the principles to the many situations of secondary analysis such as those undertaken by researchers who had not designed the study. We gave the examples of the use of multiple cohorts to examine possible changes over time in rates of disorder, and the contrasting of the various studies of the long-term outcome following early institutional care to elucidate why the findings involved some important differences.

Less progress has been made in the synthesis of evidence that comes from more diverse sources. We need to be triangulating and testing our understanding using all the available methodologies; clinical, epidemiological, RCTs and natural experiments; biological, cognitive and behavioral studies; human, animal and where possible post-mortem and in-vitro studies; and imaging, genomics and epigenetics. Such synthesis remains largely informal and one very demanding of interdisciplinary expertise. These concerns also relate to the design of single studies and research programs. These too should where possible test causal inference in multiple ways and use complementary methodologies. There is no one solution or combination that suits all circumstances and the need is to consider a range of approaches (see also Lewis et al., 2013).

The press for larger studies should not be allowed to prejudice the quality of measures of the environment. We commented on this point when discussing G x E but the issue is a broader one. Probably, too, there is a need to develop appropriate biological measures for younger children. Diary methods or ecological momentary assessment have proved helpful but, probably, much greater use could be made of the record of social communication, activity and environmental features offered by mobile phones (now used by most adolescents and many children). Google Street View has been found to provide a good

measure of neighborhood quality (Odgers et al., 2012). More progress, however, is needed on the measurement of individual differences in environmental experiences.

In the body of the paper we have discussed the strengths and limitations of genetics, imaging, epigenetics and other technologies. Although the potential of these has been somewhat oversold, it will be essential to try to make appropriate use of them – especially in combinations. Two areas that warrant mention are prevention studies based on a biology-based high risk (as in the BASIS study of parent-mediated prevention of autism – Green et al., 2015), and studies using risk indices based on genetic, demographic and clinical features to predict individual differences in outcomes following CBT for anxiety disorders (Hudson et al., 2013; Roberts et al., 2014).

We have noted the threat that is inherent in an excessive reliance on hypothesis-free technology-driven research (such as with GWAS), and urged the need for the biological study of neglected areas of clinical and developmental research where there are important leads that should be followed. The same applies to the neglected topic of the meaning of sex differences in both behavioral and brain responses to environmental stimuli, individual differences in environmental vulnerability and the phenomenon of resilience, and the biological embedding of experiences.

Finally, we consider that a change in clinical culture will be needed to fund large scale intervention studies that have become accepted in other areas of medicine, to persuade practitioners to use manualized methods of treatment, and to fund the training of clinical and service administrative staff in the methods of research.

In the introduction to this paper we drew attention to the many excellent examples in child psychiatry and psychology of high quality research, particularly in the fields of epidemiology and longitudinal studies that have often (but not always) used adequate sized

samples and which pay appropriate attention to the avoidance of bias and to testing of the causal inference. However, if the challenges are to be met it will be necessary to pay attention to the advice given by Einstein and Infeld (1938): To raise new questions, new possibilities, to regard old questions from a new angle, requires creative imagination and marks real advance in science.

Key Points

1. Substantial achievements from hypothesis-driven analysis of longitudinal data from adequately powered studies with high quality measurement and from the development and testing of psychological therapies.
2. Threats to validity derive from:

   i). the inability to study brain tissues in life

   ii). classification systems that involve extensive overlap among diagnoses

   iii). data dredging

   iv). pressures to publish driven by motives other than production of solid findings

   v). the neglect of key clinical and developmental issues that is coincident with an excessive emphasis on technology-driven research.

3. Suggested remedies include:

   a). the consistent use of CONSORT and STROBE guidelines regarding preplanned analysis

   b). a willingness to fund large scale studies of quality with adequate power

   c). appropriate testing of causal inferences using randomized trials, 'natural experiments' and animal models

d). the use of biological measures, along with epidemiological and other strategies to examine validity

Acknowledgements

This review article was invited by the journal, for which a small honorarium was offered to aid expenses. AP was part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The article has undergone full external peer review.

Academy of Medical Sciences. (2007). *Identifying the environmental causes of disease: How should we decide what to believe and when to take action?* London: Academy of Medical Sciences.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

Arango, C., Giraldez, M., Merchan-Naranjo, J., Baeza, I., Castro-Fornieles, J., Alda, J-A ... Parellada, M. (2014). Second-generation antipsychotic use in children and adolescents: A six-month prospective cohort study in drug-naive patients. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53, 1179-1190.

Arseneault, L., Cannon, M., Poulton, R., Murray, R., Caspi, A., & Moffitt, T.E. (2002). Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. *British Journal of Psychiatry*, 325, 1212-1213.

Arseneault, L., Milne, BJ., Taylor, A., Adams, F., Delgado, K., Caspi, A., & Moffitt, TE. (2008). Being bullied as an environmentally mediated contributing factor to children's internalizing problems: A study of twins discordant for victimization. *Archives of Pediatrics & Adolescent Medicine,* 162, 145-150.

Bakermans-Kranenburg, M.J., van IJzendoorn, M.H., & Juffer, F. (2008). Earlier is better: A meta-analysis of 70 years of intervention improving cognitive development in institutionalized children. *Monographs of the Society for Research in Child Development*, 73, 279-293.

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic , and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Baron-Cohen, S., Leslie, A.M., Frith, U. (1985) Does the autistic child have a "theory of mind?" *Cognition*, 21, 37-46.

Battaglia, M., Ogliari A., D'Amato, F., & Kinkead, R. (2014). Early-life risk factors for panic and separation anxiety disorder: Insights and outstanding questions arising from human and animal studies of CO2 sensitivity. *Neuroscience and Biobehavioral Reviews*, 46, 455-464.

Beck, A.T., & Dozois, D.J.A. (2014). Cognitive theory and therapy: past, present and future. In S. Bloch, S.A. Green, & J. Holmes (Eds.), *Psychiatry: Past, present, and prospect* (pp. 366-382). Oxford: Oxford University Press.

Berens, A.E., & Nelson, C.A. (2015). The science of early adversity: is there a role for large institutions in the care of vulnerable children? *Lancet*, published online January 29, 2015 http://dx.doi.org/10.1016/S0140-6736(14)61131-4.

Bloch, S., Green, S.A., Holmes, H. (2014) *Psychiatry: Past, present, and prospect*. Oxford: Oxford University Press.

Bolton, P., Carcani-Rathwell, I., Hutton, J., Goode, S., Howlin, P., & Rutter, M. (2011). Epilepsy in autism: Features and correlates. *British Journal of Psychiatry*, 198, 289-294.

Boyce, W.T. & Kobor, M.S. (2015). Development and the epigenome: the 'synapse' of gene-environment interplay. *Developmental Science*, 18, 1-23.

Bradford Hill, S. (2015). The environment and disease: association or causation? *Journal of the Royal Society of Medicine*, 108, 32-37.

Bradley, C. (1937-1938). The behavior of children receiving Benzedrine. *American Journal of Orthopsychiatry*, 94, 577-585.

Byrd, A.L., & Manuck, S.B. (2014). MAOA, childhood maltreatment, and antisocial behavior: Meta-analysis of a gene-environment interaction. *Biological Psychiatry*, 75, 9-17.

Byrd, A.L., Wright, A.G.C., Sen, S., Shedden, K., Manuck, S.B. (2015) Reply to: The Liptak-Stouffer test for meta-analyses. *Biological Psychiatry*, 77, e3-e4.

Calhoun, V.D. & Arhabshirani, MR. (2014). Neuroimaging-based automatic classification of schizophrenia. In I. Singh, WP. Sinnoff-Armstrong & J. Sarnlesan (Eds.), *Bioprediction, biomarkers and bad behavior: Scientific, legal and ethical challenges* (pp. 206-229). New York: Oxford University Press.

Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16, 260-266.

Caspi, A., Hariri, A. R., Holmes, A., Uher, R., & Moffitt, T. E. (2010). Genetic sensitivity to the environment: The case of the serotonin transporter gene and its implications for studying complex diseases and traits. *American Journal of Psychiatry*, 167, 509-527.

Caspi, A., Langley, K., Milne, B., Moffitt, T.E., O'Donnovan, M., Owen, M.J., Polo Thomas, M., Poulton, R., Rutter, M., Taylor, A., Williams, B., & Thapar, A. (2008). A replicated molecular genetic basis for subtyping antisocial behavior in children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 65, 203-210.

Caspi, A., Moffitt, T.E., Cannon, M., McClay, J., Murray, R., Harrington, H.L., . . . Braithwaite, A. (2005). Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene X environment interaction. *Biological Psychiatry*, 57, 1117-1127.

Chen, L., Pan, H., Tuan, T.A., The, A.L., MacIsaac, J.L., Mah, S.M. …. And the GUSTO Study Group. (2015). Brain-derived neurotrophic factor (BDNF) Val66Met polymorphism influences the association of the methylome with maternal anxiety and neonatal brain volumes. *Development and Psychopathology*, 27, 137-150.

Cicchetti, D. & Cohen, D. (1995). *Developmental Psychopathology*. New York: John Wiley & Sons, Inc.

Cochran, W.G. & Chambers, S.P. (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society*. Series A (General), 128, 234-266.

Collins, F. (2010). *The language of life: DNA and the revolution in personalized medicine*. London: Profile Books.

Collishaw, S., Maughan, B., Natarajan, L., & Pickles, A. (2010). Trends in adolescent emotional problems in England: A comparison of two national cohorts twenty years apart. *Journal of Child Psychology and Psychiatry*, 51, 885-894.

Conti, G., Hansman, C., Heckman, J.J., Novak, M.F.X., Ruggiero, A. & Suomi, S.J. (2012). Primate evidence on the late health effects of early-life adversity. *Proceedings of the National Academy of Sciences*, 109, 8866-8871.

Cristea, I.A., Kok, R.N., & Cuijpers, P. (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: meta-analysis. *The British Journal of Psychiatry*, 206, 7-16.

Croning, M.D.R., Marshall, M.C., McLaren, P., Armstrong, J.D., & Grant, S.G.N. (2009). G2Cdb: the genes to cognition database. *Nucleic Acids Research*, 37. doi: 10.1093/na/gkn700.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25, 7-29.

Dodge, K.A., & Rutter, M. (2011). *Gene-Environment Interactions in Developmental Psychopathology*. New York: The Guilford Press.

Einstein, A. & Infeld, L. (1938). *The Evolution of Physics*. Cambridge: Cambridge University Press.

Eisenberg, L., Lachman, R., Molling, P., Lockner, A., Mizelle, J., & Conners, K. (1963) A psychopharmacologic study in a training school for delinquent boys. *American Journal of Orthopsychiatry,* 33, 431-447.

Ellis, B.J., & Boyce, W.T. (2011). Differential susceptibility to the environment: Toward an understanding of sensitivity to developmental experiences and context. *Developmental Psychopathology,* 23, 1-5.

Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., Pickles, A., Baron-Cohen, S., Bolton, P., Johnson, MH., BASIS team (2012) Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current Biology*, 22, 338-342.

Emsley, R., Dunn, G., White, IR. (2010). Mediation and moderation of treatment effects in randomized controlled trials of complex interventions. *Statistical Methods in Medical Research*, 19, 237-270.

Fedorowicz, V.J., & Fombonne, E. (2005). Metabolic side effects of atypical antipsychotics in children: A literature review. *Journal of Psychopharmacology*, 19, 533-550.

Feast, J., Grant, M., Rushton, A., Simmonds, J., & Sampeys, C. (2013) *Adversity, adoption and afterwards: A mid-life follow-up study of women adopted from Hong Kong*. London: British Association for Adoption & Fostering.

Fergusson, DM., Lynskey, MT., & Horwood, J. (1996). Factors associated with continuity and changes in disruptive behavior patterns between childhood and adolescence. *Journal of Abnormal Child Psychology*, 24, 533-554.

Fombonne, E., Wostear, G., Cooper, V., Harrington, R., & Rutter, M. (2001). The Maudsley long-term follow-up of child and adolescent depression. 1. Psychiatric outcomes in adulthood. *British Journal of Psychiatry,* 179, 210-217.

Fonagy, P., Cottrell, D., Phillips, J., Bevington, D., Glaser, D., & Allison, E. (2015). *What works for whom?: A critical review of treatments for children and adolescents*. New York: The Guilford Press.

Foster, E.M. (2010). Causal inference and developmental psychology. *Developmental Psychology,* 46, 1454-1480.

Freedman, R., Lewis, D.A., Michels, R., Pine, D.S., Schultz, S.K., Tamminga, C.A., Gabbard, G.O., Gau, S.S., Javitt, D.C., Oquendo, M.A., Shrout, P.E., Vieta, E., & Yager, J. (2013). The initial field trials of DSM-5: new blooms and old thorns. *American Journal of Psychiatry*, 170, 1-5.

Frick, P.J., Ray, J.V., Thornton, L.C., & Kahn, R.E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents?: A comprehensive review. *Psychological Bulletin*, 140, 1-57.

Frith, C. (2003). What do imaging studies tell us about the neural basis of autism? In G. Bock, & J. Goode (Eds.), *Autism: neural basis and treatment possibilities* (pp. 149-176). Chichester: John Wiley and Sons.

Ginzburg, D.M., Bohn, C., Hofling, V., Weck, F., Clark, D.M., & Stangier. U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy*, 50, 747-752.

Goodman, R., & Yude, C. (1997). Do unilateral lesions of the developing brain have side-specific psychiatric consequences in childhood? *Laterality*, 2, 103-115.

Green, J., Charman, T., Pickles, A., Wan, M.W., Elsabbagh, M., Slonims, V., Taylor, C., McNally, J., Booth, R., Gliga, T., Jones, E.J.H., Harrop, C., Bedford, R., Johnson, M.H., & the BASIS team. (2015). Parent-mediated intervention versus no intervention for infants at high risk of autism: a parallel, single-blind, randomised trial. *Lancet*, 2, 133-140.

Harrington, R., Fudge, H., Rutter, M., Pickles, A., & Hill, J. (1990). Adult outcomes of childhood and adolescent depression. I. Psychiatric status. *Archives of General Psychiatry*, 47, 465-473.

Hazell, P., O'Connell, D., Heathcote, D., Robertson, J., & Henry, D. (1995). Efficacy of tricyclic drugs in treating child and adolescent depression: a meta-analysis. *British Medical Journal*, 310, 897-901.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E. & Lumey, L. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences*, 105, 17046-17049.

Henquet, C., Murray, R., Linszen, D., & van Os, J. (2005). The environment and schizophrenia: The role of cannabis use. *Schizophrenia Bulletin*, 31, 608-612.

Henquet, C., Di Forti, M., Morrison, P., Kuepper, R., & Murray, R.M. (2008). Gene-environment interplay between cannabis and psychosis. *Schizophrenia Bulletin*, 34, 1111-1121.

Hermelin, B., & O'Connor, N. (1970). *Psychological experiments with autistic children*. London: Pergamon Press.

Hodgson, K., Uher, R., Crawford, A.A., Lewis, G., O'Donovan, M.C., Keers, R., ... McGuffin, P. (2014). Genetic predictors of antidepressant side effects: A grouped candidate gene approach in the Genome-Based Therapeutic Drugs for Depression (GENDEP) study. *Journal of Psychopharmacology*, 28, 142-150.

Hostinar, C.E., Sullivan, R.M., & Gunnar, M.R. (2014). Psychobiological mechanisms underlying the social buffering of the hypothalamic-pituitary-adrenocortical axis: A review of animal models and human studies across development. *Psychological Bulletin*, 140, 256-282.

Howe, G.W., Reiss, D., & Yuh, J. (2002). Can prevention trials test theories of etiology? *Development and Psychopathology*, 14, 673-694.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2009). Savant skills in autism: Psychometric approaches and parental reports. *Philos Trans R Soc Lond B Biol Sci*, 364, 1359-1367.

Hudson, J. L., & Lester, K.J., Lewis, C.M., Tropeano, M., Creswell, C., Collier, D.A., Cooper, P., Lyneham, H.J., Morris, T., Rapee, R.M., Roberts, S., Donald, J.A., & Eley, T.C. (2013). Predicting outcomes following cognitive behaviour therapy in child anxiety disorders: the influence of genetic, demographic and clinical information. *Journal of Child Psychology and Psychiatry*, 54, 1086-1094.

Hyman, S.E. (2014). Psychiatry and neuroscience. In S. Bloch, S.A. Green & J. Holmes (Eds.), *Psychiatry: Past, present, and prospect* (pp. 1-21). Oxford: Oxford University Press.

International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455, 237-241.

Iyegbe, C., Campbell, D., Butler, A., Ajnakina, O., & Sham, P. (2014). The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for GxE research. *Soc Psychiatry Psychiatr Epidemiol*, 49, 169-182.

Jaffee, S.R., Caspi, A., Moffitt, T.E., Polo-Tomas, M., Price, T.S. & Taylor, A. (2004). The limits of child effects: Evidence for genetically mediated child effects on corporal punishment but not only physical maltreatment. *Developmental Psychology*, 40, 1047-1058

Kanner, L., & Eisenberg, L. (1956). Early infantile autism 1943-55. *American Journal of Orthopsychiatry*, 26, 556-566.

Kapur, S., Phillips, A.G., & Insel, T.R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, 17, 1174-1179.

Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: Evidence of genetic moderation. *Archives of General Psychiatry*, 68, 444-454.

Kazdin, A.E. (2007). Mediators and mechanisms of change in psychotherapy research. Annual Review of Clinical Psychology, 3, 1-27.

Kendler, K.S. (2014). DSM issues: Incorporation of biological tests, avoidance of reification, and approach to the "Box Canyon Problem". *American Journal of Psychiatry*, 171, 1248-1250.

Kendler, K.S., Jaffee, S.R., & Romer, D. (2011). *The dynamic genome and mental health: The role of genes and environment in youth development*. Oxford: Oxford University Press.

Kendler, K.S., & Prescott, C.A. (2006). *Genes, environment, and psychopathology: Understanding the causes of psychiatric and substance us disorders*. New York: Guilford Press.

Kraemer, H.C. (2015). Evaluating interventions. In A. Thapar, D. Pine, J.F. Leckman, S. Scott., M. Snowling., & E. Taylor (Eds.), *Rutter's Child and Adolescent Psychiatry* (6th ed.). Oxford: Wiley Press.

Krishnan, V., Han, M-H., Graham, D.L., Berton, O., Renthal, W., Russo, S.J. .. Nestler, E.J. (2007). Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell*, 131, 391-404.

Kumsta, R., Rutter, M., Stevens, S., & Sonuga-Barke, E.J. (2010). IX. Risk, causation, mediation, and moderation. *Monographs of the Society for Research in Child Development*, 75, 187-211.

Layard, R., & Clark, D.M. (2014). *Thrive: The power of evidence-based psychological therapies*. London: Penguin.

Leucht, S., Cipriani, A., Spineli, L., Mavridis, D., Orey, D., RIchter, F., Samara, M., Barbui, C., Engel, R.R., Geddes, J.R., Kissling, W., Stapf, M.P., Lassig, B., Salanti, G., Davis, J.M. (2013). Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet*, 382, 951-962.

Levine, S. (1957). Infantile experience and resistance to physiological stress. *Science*, 126, 405-406.

Levine, S. (1962). The effects of infantile experience on adult behavior. In A.J. Bachrach (Ed). *Experimental foundations of clinical psychology* (pp. 139-169). New York: Basic Books.

Lewis, S.J., Relton, C., Zammit, S., & Smith, G.D. (2013). Approaches for strengthening causal inference regarding prenatal risk factors for childhood behavioural and psychiatric disorders. *Journal of Child Psychology and Psychiatry*, 54, 1095-1108.

Lichenstein, P., Carlström, E., Råstam, M., GIllberg, C., & Anckarsäter, H. (2010). The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *American Journal of Psychiatry,* 167, 1357-1367.

Linnet, K.M., Dalsgaard, S., Obel, C., Wisborg, K., Henriksen, T.B., Rodriguez, A., & al., et. (2003). Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: Review of the current evidence. *American Journal of Psychiatry,* 160, 1028-1040.

Lipman, R.S. (1974). NIMH-PRB-support of research in minimal brain dysfunction in children. In C.K. Conners (Ed.), *Clinical use of stimulant drugs in children* (pp. 202-213). Amsterdam: Excerpta Medica.

Logothetis, N.K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453, 869-878.

Lyons, D.M., Parker, K.J. & Schatzberg, A.F. (2010). Animal models of early life stress: implications for understanding resilience. *Developmental Psychobiology*, 52, 616-624.

Mackinnon, D.P. & Fairchild, A.J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science*, 18, 16.

Malhotra, D., & Sebat, J. (2012). CNVs: Harbinger of a rare variant revolution in psychiatric genetics. *Cell*, 148, 1223-1241.

Manning, L.G., Davies, P.T., & Cicchetti, D. (2014). Interparental violence and childhood adjustment: How and why maternal sensitivity is a protective factor. *Child Development*, 85, 2263-2278.

Matheson, S.L., Shepherd, A.M., & Carr, V.J. (2014). How much do we know about schizophrenia and how well do we know it? Evidence from the Schizophrenia Library. *Psychological Medicine*, 44, 3387-3405

McCracken, J. (2011). Pharmacotherapy for autism spectrum disorders. In D. G. Amaral, G.Dawson & D. H. Geschwind (Eds.), *Autism Spectrum Disorders* (pp. 1309-1322). New York: Oxford University Press.

McGowan, P.O., Sasaki, A., Huang, T.C., Unterberger, A., Suderman, M., Ernst, C., Meaney, M.J., Turecki, G., Szyf, M. (2008). Promoter-wide hypermethylation of the ribosomal RNA gene promoter in the suicide brain. *PlosOne* 3:e2085.

McKim, V.R. & Turner, S.P. (1997). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*. Notre Dame, IN: University of Notre Dame Press.

McLellan, J.M., Susser, E., & King, M-C. (2006). Maternal famine, de novo mutations, and schizophrenia. *Journal of the American Medical Association*, 296, 582-584.

McMahon, B., Pugh, T.F., & Ibsen, J. (1960). *Epidemiological Methods*. Boston: Little Brown.

Moffitt, T.E., Caspi, A., Rutter, M., & Silva, P.A. (2001*). Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. Cambridge, UK: Cambridge University Press.

Mohiuddin, S., & Ghaziuddin, M. (2013). Psychopharmacology of autism spectrum disorders: A selective review. *Autism*, 17, 645-654.

Molling, P.A., Lockner, A.W. Jr., Sauls, R.J., & Eisenberg, L. (1962). Committed delinquent boys: The impact of perphenazine and of placebo. *Archives of General Psychiatry*, 7, 70-76.

Mostafavi, S. & Morris, Q. (2012) Combining many interaction networks to predict gene function and analyze gene lists. *Proteomics*, 12, 1687-1696.

Murray, R. (2014). On collecting meta-analyses of schizophrenia and postage stamps. *Psychological Medicine*, 44, 3407-3408.

National Research Council. (2002*). Scientific research in education*. Washington: National Academy Press.

Nurse, P. (2015). Trust in science. *Royal Society Notes and Records*, 69, 217-222.

Obel, C., Linnet, K.M., Henriksen, T.B., Rodriguez, A., Järvelin, M.R., Kotimaa, A., Moilanen, I., Ebeling, H., Bilenberg, N., Taanila, A., Ye, G., Olsen, J. (2009). Smoking during pregnancy and hyperactivity-inattention in the offspring – comparing results from three Nordic cohorts. *International Journal of Epidemiology*, 38, 698-705.

Odgers, C.L., Caspi, A., Bates, C.J., Sampson, R.J. & Moffitt, T.E. (2012). Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *Journal of Child Psychology and Psychiatry*, 53, 1009-1017.

Ouellet-Morin, I., Danese, A., Bowes, L., Shakoor, S., Ambler, A., Pariante, C., Papadopoulos, A., Caspi, A., Moffitt, TE., & Arseneault, L. (2011). A discordant MZ twin design shows blunted cortisol reactivity among bullied children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 50, 574-582.

Ouellet-Morin, I., Wong, CCY., Danese, A., Pariante, CM., Papadopoulos, AS., Mill, J., & Arseneault, L. (2013). Increased SERT methylation is associated with bullying victimization and blunted cortisol response to stress in childhood: A longitudinal study of discordant MZ twins. *Psychological Medicine*, 43, 1813-1823.

Patel, M.X., Collins, S., Hellier, J., Bhatia, G., & Murray, R.M. (2015). The quality of reporting of phase II and III trials for new antipsychotics: a systematic review. *Psychological Medicine*, 45, 467-479.

Pelphrey, K., Vander Wyk, B., Crowley, M. (2015). Neuroimaging in child psychiatry. In A. Thapar, D. Pine, J.F. Leckman, S. Scott., M. Snowling., & E. Taylor (Eds.), *Rutter's Child and Adolescent Psychiatry* (6th ed.). Oxford: Wiley Press.

Pfeifer, J.H. & Peake, S.J. (2012) Self-development: Integrating cognitive, socioemotional, and neuroimaging perspectives. *Developmental Cognitive Neuroscience*, 2, 55-69.

Pickles, A., Harris, V., Green, G., Aldred, C., McConachie, H., Slonims, V., LeCouteur, A., Hudry, K., Charman, T. and the PACT Consortium. (2015). Treatment mechanism in the MRC preschool autism communication trial: implications for study design and parent-focussed therapy for children. *Journal of Child Psychology and Psychiatry*, 56, 162-170.

Pickles, A., Simonoff, E., Conti-Ramsden, G., Falcaro, M., Simkin, Z., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2009). Loss of language in early development of autism and specific language impairment. *Journal of Child Psychology and Psychiatry*, 50, 843-852.

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., . . . Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 466(7304), 368-372. doi: 10.1038/nature09146

Pourcain, B., Mandy, W., Heron, J., Golding, J., Davey Smith, G., & Skuse, D. (2011). Links between co-occurring social-communication and hyperactive-inattentive trait trajectories. *Journal of the American Academy of Child & Adolescent Psychiatry*, 50, 892-902.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L. & Petersen SE. (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59, 2142-2154.

Rapoport, J.L., Buchsbaum, M.S., Weingartner, H., Zahn, T.P., Ludlow, C., & Mikkelsen, E.J. (1980). Dextroamphetamine: Its cognitive and behavioral effects in normal and hyperactive boys and normal men. *Archives of General Psychiatry*, 37, 933-943.

Rapoport, J.L., Buchsbaum, M.S., Zahn, T.P., Weingartner, H., Ludlow, C., & Mikkelsen, E.J.. (1978). Dextroamphetamine: cognitive and behavioral effects in normal prepubertal boys. *Science*, 199, 560-563.

Rawlins, M. (2008). De Testimonio: On the evidence for decisions about the use of therapeutic interventions. *Clinical Medicine*, 8, 579-588.

Regier, D.A., Narrow, W.E., Clarke, D.E., Kraemer, H.C., Kuramoto, S.J., Kuhl, E.A., & Kupfer, D.J. (2013). DSM-5 field trials in the United States and Canada, part ii: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, 170, 59-70.

Rice, F., Harold, G.T., Boivin, J., Hay, D.F., van den Bree, M., & Thapar, A. (2009). Disentangling prenatal and inherited influences in humans with an experimental design. *Proceedings of the National Academy of Sciences*, 106, 2464-2467.

Rietveld, C.A., Conley, D., Eriksson, N., Esko, T., Medland, S.E., Vinkhuyxen, A.A.E. ... Social Science Genetics Association Consortium. (2014). Replicability and robustness of Genome-Wide-Association studies for behavioral traits. *Psychological Science*, 1-12. doi: 10.1177/0956797614545132

Risch, N., Herrell, R., Lehner, T., Liang, K.Y., Eaves, L., Hoh, J., Griem, A., Kovacs, M., Ott, J., & Merikangas, K.R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression. *Journal of the American Medical Association*, 301, 2462-2471.

Roberts, S., Lester, K.J., Hudson, H.L., Rapee, R.M., Creswell, C., Cooper, P.J., Thirlwall, K.J., Coleman, J.R.I., Breen, G., Wong, C.C.Y. and Eley, T.C. (2014). Serotonin transporter methylation and response to cognitive behaviour therapy in children with anxiety disorders. *Translational Psychiatry*, 4, 1-5.

Rosenzweig, M.R. & Bennett, E.L. (1996). Psychobiology of plasticity: Effects of training and experience on brain and behavior. *Behavioral Brain Research*, 78, 57-65.

Rosenzweig, M.R., Krech, D., Bennett, E.L., & Diamond, M.C. (1962). Effects of environmental complexity and training on brain chemistry and anatomy: A replication and extension. *Journal of Comparative and Physiological Psychology*, 55, 429-437.

Rothman, K.J. & Greenland, S. (2002). Causation and causal inference. In R. Detels, J. McEwan, R. Beaglehole & H. Tanaka (Eds). *Oxford Textbook of Public Health*, 4th Edn. (pp. 641-653). Oxford: oxford University Press.

Rushton, A., Grant, M., Feast, J., & Simmonds, J. (2013). The British Chinese Adoption Study: Orphanage care, adoption and mid-life outcomes. *Journal of Child Psychology and Psychiatry*, 54, 1215-1222.

Rutter, M. (2007). Proceeding from observed correlation to causal inference: The use of natural experiments. *Perspectives on Psychological Science*, 2, 377-395.

Rutter, M. (2008). Introduction: Whither gene-environment interactions*? Novartis Found Symp*, 293, 1-12; discussion 68-70.

Rutter, M. (2012). Achievements and challenges in the biology of environmental effects. *Proceedings of the National Academy of Sciences*, 109, 17149-17153.

Rutter, M. (2013). Resilience: Clinical implications. *Journal of Child Psychology and Psychiatry*, 54, 474-487.

Rutter, M. (2014) Biomarkers: Potential and Challenges. In I. Singh, W.P. Sinnoff-Armstrong & J. Sarnlesan (Eds.), *Biopredictions, biomarkers and bad behavior: Scientific, legal and ethical challenges* (pp. 188-205). New York: Oxford University Press.

Rutter, M., & Azis-Clauson, C. (2015). Biology of environmental effects. In A. Thapar, D. Pine, J.F. Leckman, S. Scott., M. Snowling., & E. Taylor (Eds.), *Rutter's Child and Adolescent Psychiatry* (6th ed.). Oxford: Wiley Press.

Rutter, M., & Pine, D. (2015). Diagnosis, diagnostic formulations, and classification. In A. Thapar, D. Pine, J.F. Leckman, S. Scott., M. Snowling., & E. Taylor (Eds.), *Rutter's Child and Adolescent Psychiatry* (6th ed.). Oxford: Wiley Press.

Rutter, M., & Sonuga-Barke, E.J. (2010). X. Conclusions: Overview of findings from the E.R.A. study, inferences, and research implications. Deprivation-specific Psychological Patterns: Effects of Institutional Deprivation (pp. 212-229). Oxford*: Monographs of the Society for Research in Child Development* 75.

Rutter, M., & Thapar, A. (2014). Genetics of autism spectrum disorders. In R. Volkmar, S.J. Rogers, R. Paul & K.A. Pelphrey (Eds.), *Handbook of Autism and Pervasive Developmental Disorders, Assessment, Interventions, Policy, and the Future* (4th ed.) (pp.411-423) Hoboken NJ: Wiley Press.

Rutter, M., & Uher, R. (2012). Classification issues and challenges in child and adolescent psychopathology. *International Review of Psychiatry*, 24, 514-529.

Rutter, M, Tizard, J., & Whitmore, K. (1970). *Education, health and behaviour*. London: Longman.

Sackett, D.L., Haynes, R.B., & Tugwell, P. (1985). *Clinical epidemiology: A basic science for clinical medicine*. Boston: Little, Brown.

Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., Richardson, W.S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, 312, 71-72.

Sakurai, T., Cai, G., Grice, D.E., & Buxbaum, J.D. (2011). In E. Hollander, A. Kolevzon, & J.T. Coyle (Eds.), *Textbook of Autism Spectrum Disorders* (pp. 281-298). Arlington, VA: American Psychiatric Publishing.

Schepman, K., Collishaw, S., Gardner, F., Maughan, B., Scott, J., Pickles, A. (2011). Do changes in parent mental health explain trends in youth emotional problems? *Social Science & Medicine*, 73, 293-300.

Schizophrenia Working Group of the Psychiatric Genomics. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511, 421-427.

Schneider, M., & Koch, M. (2003). Chronic pubertal, but not adult chronic cannabinoid treatment impairs sensorimotor gating, recognition memory, and the performance in a progressive ratio task in adult rats. *Neuropsychopharmacology*, 28, 1760-1769.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., . . . Kendall, J. (2007). Strong association of de novo copy number mutations with autism. *Science*, 316, 445-449.

Sellers, R., Maughan, B., Pickles, A., Thapar, A., & Collishaw. S. (2015) Trends in parent- and tech-rated emotional, conduct and ADHD problems and their impact in prepubertal children in Great Britain: 1999-2008. *Journal of Child Psychology and Psychiatry*, 56, 49-57.

Shadish, WR., Cook, TD., & Campbell, DT. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Simmons, J.P., Nelson, L.D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47, 921-929.

Snyder, A.Z., & Raichle, M.E. (2012) A brief history of the resting state: The Washington University perspective. *Neuroimage*, 62, 902-910

Stein, Z.A., Susser, M., Saenger, G., & Marolla, F. (1975). *Famine and human development: The Dutch hunger winter of 1944-1945.* New York: Oxford University Press.

Steyerberg, E. (2009). *Clinical prediction models: A practical approach to development, validation and updating*. New York: Springer.

Stigler, K., Potenza, M., Posey, D., & McDougle, C. (2004). Weight gain associated with atypical antipsychotic use in children and adolescents: Prevalence, clinical relevance, and management. *Pediatric Drugs*, 6, 33-44.

Thapar, A., & Cooper, M. (2013). Copy number variation: What is it and what has it told us about child psychiatric disorders? *Journal of the American Academy of Child & Adolescent Psychiatry*, 52, 772-774.

Thapar, A., & Rutter, M. (2015a). Neurodevelopmental disorders. In A. Thapar, D. Pine, J.F. Leckman, S. Scott., M. Snowling., & E. Taylor (Eds.), *Rutter's Child and Adolescent Psychiatry* (6th ed.). Oxford: Wiley Press.

Thapar, A., & Rutter, M. (2015b). Using natural experiments and animal models to study causal hypotheses in relation to child mental health problems. In A. Thapar, D. Pine, J.F. Leckman, S. Scott., M. Snowling., & E. Taylor (Eds.), *Rutter's Child and Adolescent Psychiatry* (6th ed.). Oxford: Wiley Press.

Thapar, A., Rice, F., Hay, D., Boivin, J., Langley, K., van den Bree, M., Rutter, M., & Harold, G. (2009). Prenatal smoking might not cause attention-deficit/hyperactivity disorder: Evidence from a novel design. *Biological Psychiatry*, 66, 722-727.

Tobi, E.W., Slagboom, P.E., van Dongen, J., Kremer, D., Stein, A.D., Putter, H., Heijmans, B.T. & Lumey, L. (2012). Prenatal famine and genetic variation are independently and additively associated with DNA methylation at regulatory loci with IGF2/H19. *PLoS One*, 7, e37933.

Trezza, V., Cuomo, V., Vanderschuren, L.J.M.J. (2008). Cannabis and the developing brain: Insights from behavior. *European Journal of Pharmacology*, 585, 441-452.

Uher, R., & McGuffin, P. (2010). The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update. *Molecular Psychiatry*, 15, 18-22.

Van Dijk, K.R., Sabuncu, M.R. & Buckner, R.L. (2012) The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, 59, 431-438.

Vargha-Khadem, F., Isaace, E., Van der Werf, S., Robb, S. & Wilson, J. (1992). Development of intelligence and memory in children with hemiplegic cerebral palsy. *Brain*, 115, 315-329.

Viding, E., & McCrory, E.J. (2015) Developmental risk for psychopathy. In A. Thapar, D. Pine, J.F. Leckman, S. Scott., M. Snowling., & E. Taylor (Eds.), *Rutter's Child and Adolescent Psychiatry* (6th ed.). Oxford: Wiley Press.

Vorria, P., Ntouma, M., Rutter, M. (2015). Vulnerability and resilience after early institutional care: The Greek Metera study. *Development and Psychopathology* doi: 10.1017/S0954579415000243

Wadsworth, M.E.J. (1991). *The imprint of time*. Oxford: Oxford University Press.

Weisz, J.R. & Kazdin, A.E. (Eds.) (2010). *Evidence-based psychotherapies for children and adolescents* (2nd edn). New York: Guilford Press.

Williams, N., Zaharieva, I., Martin, A., Langley, K., Mantripragada, K., Fossdal, R., . . . Thapar, A. (2010). Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: A genome-wide analysis. *Lancet*, 376, 1401-1408. doi: 10.1016/s0140-6736(10)61109-9