Nosology and measurement in child and adolescent psychiatry

Adrian Angold and E. Jane Costello
Duke University, Durham, NC, USA

The last 50 years have witnessed enormous strides in the measurement and classification of child and adolescent psychiatric disorders. Debates about whether we should continue to depend upon a categorical nosology still continue, but we argue that, despite the absence of clear dividing lines between psychiatric disorders and normality and ubiquity of diagnostic comorbidity, the current official approach to nosology has served child and adolescent psychiatric research surprisingly well. In particular we point to the utility of non-developmental diagnostic criteria as tools for discovering developmental effects on psychopathology. We also maintain that the search for sharper boundaries between disorders is fundamentally mistaken. However, official nosologies have tended to privilege information collected in diagnostic interviews and to sideline observational and other methods that cannot easily be made to conform to the format of their criteria. We suggest that it is time to remedy this situation. The ICD-10 and DSM-IV are useless for children under the age of about two, while alternatives, such as the DC:0–3, suffer from a profound lack of empirical support. We suggest a way forward through the integration of methods from temperament and psychopathology research. Finally, we deplore the failure of standardized assessment techniques to have penetrated more deeply into everyday clinical assessment. **Keywords:** Child, adolescent, nosology, assessment, measurement.

Two papers from the first volume of the *Journal of Child Psychology and Psychiatry* stand out in terms of their citation indices, and have indeed been cited a number of times in the 21st century. Both deal with nosology. The first, by Anthony and Scott (1960), is one of the foundational papers on the still highly contentious subject of prepubertal bipolar disorder (Biederman, 2003), although it relies on the nearly extinct method of single case presentation. The second, by Lionel Hersov (1960a), is the first paper to have presented substantial empirical evidence for the now standard distinction between truants and ‘neurotic’ persistent school refusers, many of whom were clearly suffering from what we would now call ‘separation anxiety disorder’ (SAD) (Hersov, 1960b). Nearly everything we now know or believe about SAD is anticipated in these papers, but in one respect they differ from a modern paper on the topic; five categories were represented in the final diagnoses given to the children, of which only one is directly represented in the ICD-10 or DSM-IV.

The purpose of this paper is to evaluate the current position of child psychiatric nosology and its measurement and to suggest some strategies for improving nosology from a developmental perspective. But first, we need to clarify what the nosological task is.

**Carving nature at the joints**

Socrates commented to Phaedrus that there are ‘two kinds of thing that it would be gratifying to grasp in a scientific way…. First, there is perceiving together and bringing into one form items which are scattered in many places, in order that one may define each thing’ (Plato, Phaedrus, 265d5). When asked what the second kind of thing might be, he responds ‘being able to cut up whatever it is again, kind by kind, according to its natural joints, and not to try to break any part into pieces, like an inexpert butcher…’ (ibid. 266a). Our aim here is to attempt to call down Socrates’ blessing by looking ‘to one and to many,’ and considering both the metaphorical ground (points of similarity) and tension (points of dissimilarity) (Richards, 1936) generated by Socrates’ comments.

**Are there any joints?**

The ground of Socrates’ simile is what usually gets attention. Life would be so much easier if we could have psychiatric disorders neatly carved up like so many Sunday joints. The big problem is that we have failed to find any cleanly separable joints. It has been argued that the fact that we now recognize diagnostic comorbidity as the norm (Angold, Costello, & Erkanli, 1999) indicates that we have not identified meaningful categories. However, collapsing categories is little help, because even if we leave ourselves with just two (behavioral/externalizing and emotional/internalizing), there is still a great deal of comorbidity between them (Angold et al., 1999). On the other hand, it is not hard to identify substantial differences between the most molecular categories. For instance, lumping together the highly comorbid childhood separation anxiety disorder and generalized anxiety disorder would obscure marked differences in their sex ratios and age-prevalence.
to nourish it and control its activity. Similarly, if we want to understand psychopathology we need to get to grips with how and why things like anxiety and depression often go together. For understanding the nature of the processes underlying depression and anxiety we suggest that attempting to better separate them is exactly the wrong way to go, because we have overwhelming evidence that they are linked. But linkage is not the same as identity. Krueger (2006, p. 156) puts it well (and Socrates would surely approve); ‘... both ‘lumping’ and ‘splitting’ perspectives on the organization of psychopathology are partially correct, and they can be reconciled by adopting a dimensional-hierarchical model of etiologic contributions within this domain that recognizes etiologic factors at continuously varying levels of specificity versus breadth.’

Figure 1 Changing prevalences of separation anxiety disorder (SAD) and generalized anxiety disorder (GAD) by sex and age in the Great Smoky Mountains Study curves (see Figure 1 which shows data from the Great Smoky Mountains Study; Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). The situation is no better at the most severe end of the psychopathological spectrum. Autism is as well-validated a category as we have, but it shades into a variety of autism spectrum disorders.

But why should ‘good disorders’ have sharply demarcated boundaries? There are plenty of examples of physical diseases that do not. For instance, to continue the arthrological analogy, rheumatoid arthritis and systemic lupus erythematosus (SLE) may be difficult to distinguish, and The American College of Rheumatology criteria for SLE were first issued in 1971, updated in 1982 and 1997, and employ a ‘Chinese menu’ format in which four of 11 diagnostic criteria must be met (http://www.rheumatology.org/publications/classification/index.asp?aud=mem). It is also well recognized that some individuals with SLE may never meet these criteria. Such diseases result from multiple intercorrelated factors in highly integrated complex systems. The brain is the most complex organ we have, and so it seems in hindsight that we really should not have been surprised or upset that psychiatric disorders have fuzzy boundaries.

Does that mean that there are no joints to find? Not at all. The problem with the analogy is in the focus on carving, not in the existence of joints. The leg of lamb we put in the oven is good for eating but useless for walking because it has been skinned, disarticulated, hung, and perhaps frozen. If we want to understand the functional anatomy of the leg we have to take account of structures that hold the two sides of the joint together, the muscles and ligaments that cross it, and the blood vessels and nerves that come from far away

Are there any categorical disorders?

Sadly, the diagnostic categories vs. dimensions debate seems to have been sparked again in the adult literature by the approach of DSM-V. For instance, Widiger and Clark (2000, p. 954) have suggested that ‘... as one builds towards DSM-V, what may emerge is a structured set not of categorical diagnoses but of component dimensions, a set of symptom cluster building blocks from which the panoply of diagnoses could be constructed.’ They go on to say, with evident approbation, ‘... the fundamental structure of future DSMs may not be composed of individual diagnoses as it is now. Rather, it may consist of an ordered matrix of symptom-cluster dimensions....’ The paper provides a compelling critique of many aspects of the DSM-IV (and its predecessors), but we believe that its proposed solution to these problems is profoundly mistaken.

We are persuaded that ADHD, ODD, CD, depression and the various anxiety disorders represent extremes of more or less continuously distributed phenomena (Pickles & Angold, 2003). But that does not mean that we eschew the use of diagnostic categories in these areas. The symptom groupings prescribed in the DSM-IV have stood up quite well to tests using confirmatory factor analysis based on scalar data (Lahey et al., 2008). The categorical and scalar approaches have happily converged on the same solutions. Several years ago Pickles and Angold commented that ‘the central question is not “is psychopathology scalar or categorical”, but “under what circumstances does it make sense to regard psychopathology as being scalar and under what circumstances does it make sense to treat psychopathology as being categorical. An essential part of our argument is that it is necessary to shift the debate away from trying to determine whether there are categorical states or dimensional levels of psychopathology, to considering the forms of relationships with other processes, whether
epidemiological or clinical” (Pickles & Angold, 2003, p. 529; see also Rutter & Sroufe, 2000).

The ICD system (on which the DSMs are glosses) was intended to serve as a tool capable of covering the entire range of diseases in a uniform format for the purposes of statistical analysis. That is still its primary function, and it is still obvious that there is a need for a categorical international nosology for statistical and epidemiological purposes. It is also obvious that there are plenty of situations (e.g., whether to give treatment or not) in which categorical decisions have to be made, even if the underlying pathological phenomena are continuously distributed. More generally, case-control designs can be a very efficient means of identifying risk factors for and clinical correlates of psychopathology. On the other hand, the dimensional properties of many disorders mean that even when we collect risk factor and diagnostic data in population surveys we might often be better off from a power point of view analyzing diagnostic interview-derived symptom scales rather than diagnoses. So the answer to the question posed in the heading is ‘yes, when it’s useful for them to be.’ Helzer and colleagues (Helzer, Kraemer, & Krueger, 2006) have suggested that it would be a good idea to develop continuous measures to lie side by side with the categorical diagnoses that emerge in the DSM-V. It now seems unlikely that this will happen, but it is worth noting that many structured diagnostic interviews collect a great deal of information about severity, duration and the like at the symptom level, so the tools are already in place for implementing such a plan.

Do we need to redefine the joints?

Consider the case of major depression and generalized anxiety disorder (GAD). There is strong evidence that they ‘belong together.’ They are very highly comorbid both cross-sectionally and longitudinally at all ages (Watson, 2005), and large twin studies have found that their genetic correlation is unity in women and over .7 in men (Kendler, Gardner, Gatz, & Pedersen, 2007). Early suggestions that GAD might be a developmental prodrome of depression have now been supplanted by evidence that depression is as likely to precede GAD as vice versa (Moffitt et al., 2007). Both also respond to selective serotonin reuptake inhibitors (and other ‘antidepressants’ for that matter; Hughes et al., 2007; Seidel & Walkup, 2006). Surely, here is a case where we have carved at a non-existent joint. But consider also that only just under two-thirds of cases of GAD have lifetime comorbidity with depression and many people with major depression never have GAD. Second, there has long been evidence (including evidence from twin studies) that the environmental risk factors for depression and anxiety disorders are different (Kendler, 1996). There are also differences in the brain circuitry thought to underlie depression and anxiety (see, e.g., Anand et al., 2005; McClure et al., 2007). So there is a case for treating unipolar depression and GAD as aspects of a single disorder, but there is also a case for separating them. When a situation like this occurs, it is usual to say that the question needs further study, but we do not think that further study will help here. Neither the similarities nor the differences will go away, however much data we collect. The important issue is not whether they are ‘really’ both aspects of one disorder, or ‘really’ separate disorders, but how the similarities and differences that we observe come about. No simple pattern of cuts can be made that will properly partition the developmental processes. Any point at which we make a cut will sacrifice important developmental ‘tendons.’ The nosological status of these two disorders is no longer really a matter of science at all; we already know enough to see that it is a matter of the pragmatics of the nosology. We have to decide whether it will prove more useful to join unipolar depression and GAD or to keep them separate. For the purposes for which the nosology is designed, does it make more sense to have them articulated or disarticulated? We should be asking things like: will joining them produce more helpful international statistics, or more appropriate clinical evaluations or treatments, or clearer communication among clinicians and patients? Researchers will continue to combine or separate them for the particular purposes of their studies regardless of their nosological status.

The utility of non-developmental criteria for developmental research

The ‘crude’ application of ‘adult-style’ diagnostic criteria to children and adolescents has been extraordinarily successful in demonstrating that many adult disorders definitely occur in childhood (e.g., major depression and social phobia), starting much earlier than used to be thought even possible, while indicating that some others really are very uncommon before later adolescence (e.g., panic disorder, classical mania and anorexia nervosa; Costello, Foley, & Angold, 2006). What is particularly striking is the degree to which, in the former case, subsequent research has also indicated that there are substantial differences between the childhood cases and presentations in adulthood. Depression is, perhaps, the best example here, where there is solid evidence for differences in associations with gender (Angold, Costello, & Worthman, 1998), sexual abuse (Hill, Pickles, Rollinson, Davies, & Byatt, 2004), sleep disturbances (Birmaher et al., 1996b), glucocorticoid function (Birmaher et al., 1996b), and treatment response (Birmaher, Neal, Williamson, Brent, & Kaufman, 1996a) between child and adolescent and adult depressions. We do not know why and how these differences arise, but it is undeniable that our non-developmental criterion sets have been the source of results that require a great deal of
developmental explanation. Fixing the phenomenological goalposts has proved to be a good strategy for identifying developmental differences in play.’

The demonstration that current diagnostic categories work as well in children as they do in adults has also stimulated the idea that ‘childhood’ disorders might also be found in adults. ADHD is now well established as an adult disorder (Kessler et al., 2006), and adult separation anxiety disorder is moving in the same direction (Shear, Jin, Ruscio, Walters, & Kessler, 2006).

The adult-style approach to diagnosis has now been shown to work pretty well the way down to the age of two (Egger et al., 2006). Findings such as this contradict the idea that disorders become progressively differentiated across childhood, because they already appear to be differentiated by the age of two. They also speak to an important founding hypothesis of the field of developmental psychopathology. In 1984, Sroufe and Rutter (1984, p. 24) suggested that ‘... the strongest predictors of later pathology are not likely to be early replicas of the behavioral indicators of adult pathology. The strongest predictors likely will be adaptational failures, defined in age-appropriate terms.’ However, it now seems that many of the most common disorders can already be identified in the preschool years (Egger et al., 2006), and these ‘early replicas’ might well be the strongest predictors of many later pathologies.

But we should not conclude from this that, from a conceptual standpoint, developmental psychopathology has generally been in error. Quite the reverse. Our non-developmental nosology has proven surprisingly robust in childhood, but it has also shown that we need to understand how similarities and differences in what are nosologically the ‘same’ disorders arise. Here the methods of developmental psychopathology will be needed more than ever (Rutter & Sroufe, 2000). Developmental psychopathology is also committed to dealing with the interrelationships between ‘normality’ and ‘abnormality,’ and so comfortably contains the conceptual apparatus for dealing with continuously distributed attributes of disorder. It also provides us with guidelines on how to proceed in situations in which there is almost wholesale failure of the standard nosology.

Where the nosology fails and what to do about it

If children of two already have diagnosable disorders, we must look for their causes still earlier in life. But the limited behavioral and linguistic repertoire of children under two means that we cannot identify a very large proportion of the symptoms required for ICD-10 or DSM-IV diagnoses – the classification breaks down. Attempts have been made to classify disorders in younger children, of which the best-known is the DC:0-3R (Zero to Three, 2005). The problem here is that its diagnostic categories have been subjected to very little empirical research, but now require immediate attention.

Fortunately, we have been given a head start by infant temperament research. Temperamental extremes have often been regarded as being risk factors for later disorders, but the time has come to determine to what extent they already are the disorders of later life. The clearest example here is provided by research on behavioral inhibition. It has been shown that behaviorally inhibited children are at risk for later anxiety disorders, that behavioral inhibition is familially linked to other anxiety disorders (Rosenbaum et al., 1991; Shamir-Essakow, Ungerer, & Raape, 2005), and that behaviorally inhibited children show patterns of neurophysiological arousal otherwise associated with anxiety disorders (Kagan, Reznick, & Snidman, 1987; Kagan & Snidman, 1999).

So, perhaps it is reasonable to suppose that a 15-month-old identified as being behaviorally inhibited already has what we would later call social phobia. The fact that many inhibited children do not go on to meet the criteria for social phobia (Kagan & Snidman, 1999) or later manifest other disorders (such as depression, Gladstone & Parker, 2006), or recover, does not eliminate this possibility, because diversity of outcomes is characteristic of psychiatric disorders at every age. Furthermore, the behavioral inhibition literature provides a model for the exploration of the earliest roots of psychopathology in that it has identified patterns of negative reactivity in early infancy that are associated with the manifestation of later behavioral inhibition (Kagan & Snidman, 1999).

Despite their origins in different research traditions (Frick, 2004), examination of the types of behavior considered relevant in the measurement of early temperament and psychopathology reveals a good deal of overlap. Lahey has pointed out that there is no inherent natural distinction between temperament and psychopathology (Lahey, 2004), rather they are overlapping constructs derived from two different research traditions. At the ages at which both temperament and psychopathology measures are available they are very highly correlated (Copeland, Landry, Stanger, & Hudziak, 2004). That does not mean that temperament and psychopathology are the same thing – there are obvious conceptual differences in their formulations (Frick, 2004), but the integration of early temperament approaches with those of psychopathologists (something that should come naturally to developmental psychopathologists; Rutter & Sroufe, 2000) seems to offer an obvious way forward for the task of establishing an empirically sustainable psychiatric nosology for the first two years of life.

Measurement

The last 50 years have witnessed the development of a vast range of measures of psychopathology, and
more are coming along all the time. So plenty of scalar and diagnostic tools are available that cover the period from the preschool years to adulthood. There can be little doubt that the existence of standard nosologies has been helpful to interview developers because they have bounded the set of topics that have to be covered. Neither has this bounding been a limitation, because modules can always be (and often have been) developed to cover material not specified by the nosology. However, we want to consider two notable failures to move forward in the assessment area.

Getting beyond the interview

It is now standard procedure to base most research diagnoses upon standardized interviews, preferably with more than one reporter. That is fine, as far as it goes, but in clinical practice we expect to observe the child, often in multiple settings. Many observational schemes have been developed for research, but these are typically not included in the research diagnostic process; rather they are treated as adjunctive information. The reason is not far to seek. The style of the ICD-10 and DSM-IV criteria, with their many frequency, duration and timing rules, lend themselves to the interview format (even though we know that much of this information is of doubtful validity; Angold, Erkanli, Costello, & Rutter, 1996; Breton et al., 1995), but observational measures do not fit this mold well. As a result, observation and testing, and interviews with younger children (Valla, Bergeron, & Smolla, 2000) have been largely removed from the domain of making research diagnoses, except where such testing is specified in the criteria for the disorders (as is the case with learning disorders). The notable exception to this general rule lies in the combined use of the Autism Diagnostic Interview (ADI-R; Rutter, Le Couteur, & Lord, 2003) and the Autism Diagnostic Observation Schedule (Lord, Rutter, DiLavore, & Risi, 2003) in the assessment of autism. The latter provides ‘presses’ for behaviors diagnostic of autism. This approach is one that deserves wider consideration across the field as a whole. However, its implementation would necessitate much debate about when and how an observed phenomenon could be allowed to stand in for a symptom like ‘often argues with adults.’ Work has begun in this area for preschoolers with the recent development of the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS; Wakschlag et al., 2002), which is consciously modeled after the ADOS. In fact the integration of such material into the nosology is likely to become part of a broader need to modify the overall format of ICD and DSM criteria as ‘tests’ for psychiatric conditions emerge. With genetic, neurochemical, and neuro-imaging methods advancing at a great rate, it seems likely they will, at some point, begin to impact the nosology. Our guess is that we will eventually find ourselves in a position similar to that of the rheumatologists, whose diagnoses are based on range of clinical historical, observational and laboratory findings, with no single type of information taking precedence.

Why don’t clinicians use standardized measures?

Now that we have all these standardized measures, and it is nearly impossible to get a research paper published without using them, why are they not de rigueur in clinical practice? Standardized psychiatric interviews were originally developed because of the weaknesses of clinical interviewers. Many of these weaknesses spring directly from biases in human information processing, which standardized interviews are designed to overcome. The general failure to adopt standardized assessments in clinical practice seems to us to be akin to a situation in which rheumatologists knew about X-rays, but decided only to use them for research purposes and not in their day-to-day practices. It continues to amaze us that psychiatrists, clinical psychologists and social workers can complete their entire trainings without being taught to use a standardized diagnostic interview. We realize that making standardized measures part of routine clinical practice will be expensive and time consuming, but we think we owe it to our patients.

Key points

- Non-developmental official psychiatric nosologies have served child and adolescent psychiatry surprisingly well.
- Fixing the ‘nosological goalposts’ has proved to be an effective way of detecting developmental effects on psychopathology.
- The format of official nosologies has privileged interview-based diagnostic measures and sidelined observational and other assessment methods.
- There is an urgent need to provide an empirically-based psychiatric nosology for children under the age of two. The temperament literature provides a valuable starting point for this endeavor.
- Unstandardized clinical assessment is no longer acceptable as a research tool. It is time that standardized assessments were incorporated into non-research settings.
Correspondence to
Adrian Angold, Duke University Medical Center, Durham, NC 27710, USA; Email: adrian.angold@duke.edu

References


modulation of fear circuit function in pediatric generalized anxiety disorder. *Archives of General Psychiatry*, 64, 97–106.


Manuscript accepted 9 June 2008