

CLINICAL MODELS FOR CHILD AND ADOLESCENT BEHAVIORAL, EMOTIONAL AND SOCIAL PROBLEMS

Thomas M Achenbach & David M Ndetei



Thomas M Achenbach PhD
Professor of Psychiatry and
Psychology, University of
Vermont, Burlington, Vermont,
US.

Conflict of interest: President
of the nonprofit Research
Center for Children, Youth,
and Families, which publishes
the Achenbach System of
Empirically Based Assessment
(ASEBA).

David M Ndetei MD

Professor of Psychiatry,
University of Nairobi, Nairobi,
Kenya.

Conflict of interest: Director
of the nonprofit Africa Mental
Health Foundation, which
promotes mental health
services, training, and research
in Africa.

This publication is intended for professionals training or practising in mental health and not for the general public. The opinions expressed are those of the authors and do not necessarily represent the views of the Editor or IACAPAP. This publication seeks to describe the best treatments and practices based on the scientific evidence available at the time of writing as evaluated by the authors and may change as a result of new research. Readers need to apply this knowledge to patients in accordance with the guidelines and laws of their country of practice. Some medications may not be available in some countries and readers should consult the specific drug information since not all dosages and unwanted effects are mentioned. Organizations, publications and websites are cited or linked to illustrate issues or as a source of further information. This does not mean that authors, the Editor or IACAPAP endorse their content or recommendations, which should be critically assessed by the reader. Websites may also change or cease to exist.

©IACAPAP 2012. This is an open-access publication under the [Creative Commons Attribution Non-commercial License](#). Use, distribution and reproduction in any medium are allowed without prior permission provided the original work is properly cited and the use is non-commercial. Send comments about this book or chapter to jmreyATbigpond.net.au

Suggested citation: Achenbach TM, Ndetei DM. Clinical models for child and adolescent behavioral, emotional, and social problems. In Rey JM (ed), *IACAPAP e-Textbook of Child and Adolescent Mental Health*. Geneva: International Association for Child and Adolescent Psychiatry and Allied Professions 2012.

This chapter presents models for distinguishing among various kinds of behavioral, emotional, and social problems manifested by children. (For brevity, we use “child,” “children,” and “childhood” to include ages from birth through 18 years.) We use the term “models” to include official nosologies (i.e., classifications of diseases) such as the World Health Organization’s (1992) *International Classification of Diseases* (ICD) and the American Psychiatric Association’s (2000) *Diagnostic and Statistical Manual of Mental Disorders* (DSM). We also use “models” to include alternative nosologies, such as the *Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood, Revised* (DC:0-3R; Zero to Three, 2005), plus empirically based models derived from data on large samples of children.

The term “model” has become widely used to designate systematic representations of phenomena. It implies that particular phenomena can be represented (i.e., modeled) in multiple ways. It also implies that different models for particular phenomena can be evaluated in terms of their utility for particular purposes, rather than requiring forced choices between one model as right versus other models as wrong. Thus, for example, a DSM model may use different terms and criteria than an ICD model for attention problems, but both may be useful for different purposes and/or in different systems of care.

Some models are designed to represent relations among particular phenomena without representing causes of the phenomena. Other models are designed to represent the causes of phenomena. At their present stage of development, the clinical models presented in this chapter do not include specific causes of behavioral, emotional, or social problems. (This chapter does not address mental retardation or disorders such as Down Syndrome, Prader Willi Syndrome, Williams Syndrome, and phenylketonuria for which genetic or other physical causal factors are well documented – see Chapter C.1.) Although genetic and other physical factors are apt to affect many behavioral, emotional, and social problems, the paucity of knowledge about specific causal pathways means that clinical models must focus on phenotypic characteristics that practitioners can identify and work with. To collaborate with practitioners in advancing our understanding of etiologies, treatment effects, and outcomes, researchers also need to use similar phenotypic models to link their work to cases seen by practitioners.

ICD AND DSM

Because the ICD-11 and DSM-5 nosologies are still being developed, we cannot be specific about the formats, categories, or criteria for these nosologies. Consequently, we must base our presentation on aspects of the existing ICD and DSM models that may continue in the pending editions as well.

The ICD and DSM models are based on inputs from committees of experts who negotiate the diagnostic categories to be used and the criteria to be specified for determining whether individuals qualify for the diagnoses modeled by each category. The experts span a broad range of clinical, administrative, public health, and research experience. Drafts of the proposed categories and criteria are widely circulated for comment and are revised on the basis of the comments. Field trials may be conducted to test the criteria. For the DSM-IV-TR child diagnostic categories that are defined mainly in terms of behavioral, emotional, and social problems (e.g.,

Emil Kraepelin (1856-1926)**The putative father of psychiatric nosology**

Emil Kraepelin is credited with constructing clinical models for psychopathology that laid the foundations for psychiatric nosologies. During his medical training at the University of Leipzig in the 1870s, Kraepelin became a disciple of Wilhelm Wundt, who is regarded as the founder of experimental psychology. Using Wundt's experimental methods to study the psychological functioning of patients, he included psychological characteristics in his clinical models, and he wrote his doctoral dissertation on the place of psychology in psychiatry.

Kraepelin initially believed that careful documentation of different psychopathological phenotypes would eventually guide discovery of a particular brain disease underlying each phenotype. In 1883, Kraepelin published the very small first edition of his *Compendium der Psychiatrie*, which was followed over the next 43 years by eight more editions of progressively greater heft. Among other achievements, Kraepelin is credited with breaking what had been a single category of all psychoses into separate categories of manic-depressive disorders and dementia praecox (early dementia), later re-named schizophrenia. In the later editions of his *Compendium*, Kraepelin added psychogenic disorders, plus personality disorders that he considered to be on the border between illness and common idiosyncrasies.

As revealed in a cartoon drawn by Kraepelin, he did not take psychiatric diagnoses too seriously. Reproduced below from the *Bierzeitung* (Beer Newspaper) of 1896, the cartoon is captioned (in English translation) "Psychiatrists of Europe! Protect your most sacred diagnoses!" Kraepelin thus warned against becoming overly awed by diagnoses.



attention problems, conduct problems, oppositional defiant problems, anxiety, depression), the criteria include a list of symptoms and a diagnostic threshold for the number of symptoms from the list that must be judged present to justify a diagnosis (American Psychiatric Association, 2000). Additional criteria include the duration of symptoms, age of onset, and impairment. The criteria are similar for children of all ages and both genders. The criteria are also similar for information obtained from different sources, such as parent, teacher, and child reports, tests, and clinical observations. If there are inconsistencies or contradictions between different sources (e.g., a teacher reports attention problems but the parent and child do not), the practitioner must ultimately make a yes-versus-no judgment about whether each diagnostic criterion is met. The practitioner must then make a yes-versus-no judgment about whether the child qualifies for a diagnosis.

DSM dimensional diagnostic criteria

In preparation for DSM-5, the American Psychiatric Association appointed a task force to consider possibilities for “dimensional” (i.e., quantitative) diagnostic criteria. The task force has published a great deal of support for dimensionalizing criteria for many kinds of child and adult diagnoses (Helzer et al, 2008). However, it is not clear at this writing whether or how DSM-5 criteria might be dimensionalized. One possibility appears to be that criteria for clinical diagnoses will retain the yes-versus-no format of DSM-IV-TR but that separate research diagnostic criteria may include some kind of dimensionalization.

ICD-10 research diagnostic criteria

The ICD-10 *Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research* (World Health Organization, 1993) added research criteria to the criteria that were previously published for clinical diagnoses (World Health Organization, 1992). The ICD-10 research diagnostic criteria resemble DSM-IV-TR criteria in that they specify symptoms and other features that must be judged yes-versus-no. After users make yes-versus-no judgments of each criterion, they must make a yes-versus-no judgment of whether a child qualifies for a particular diagnosis.

Differences between ICD-10 and DSM clinical diagnostic criteria

The ICD-10 clinical diagnostic criteria (World Health Organization, 1992) differ from DSM-IV-TR criteria in that they lack explicit statements of specific symptoms, other features, and decision rules for making yes-versus-no judgments to decide whether a child qualifies for a particular diagnosis. The ICD-10 and DSM-IV-TR also differ considerably in their descriptions of diagnostic categories and in some of the categories themselves. For example, the ICD-10 has diagnostic categories for sibling rivalry disorder, hyperkinetic conduct disorder, and disinhibited attachment disorder that lack counterparts in the DSM-IV-TR.

The ICD-10 and DSM-IV-TR additionally differ in their subdivisions of certain diagnostic categories. For example, DSM-IV-TR divides its category for attention deficit hyperactivity disorder (ADHD) into predominantly inattentive type, predominantly hyperactive-impulsive type, and combined type. By contrast, the ICD-10 category most similar to ADHD is designated as hyperkinetic disorder, which is not subdivided into categories like those of the DSM-IV-TR. On the other hand, the ICD-10 subdivides conduct disorder into conduct disorder confined to the family context, unsocialized conduct disorder, and socialized conduct disorder.

ICD-10 differs from DSM-IV in that ICD-10 lacks explicit statements of specific symptoms, other features, and decision rules for making yes-versus-no judgments to decide whether a child qualifies for a particular diagnosis.

Although the DSM-IV-TR Appendix H lists ICD-10 code numbers for many DSM diagnostic categories, practitioners should not expect much agreement between particular DSM diagnoses and ICD diagnoses made for the same children, even when the DSM and ICD diagnoses have the same code numbers.

DIAGNOSTIC CLASSIFICATION FOR INFANCY AND EARLY CHILDHOOD

The *Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood* (DC:0-3) was first published in 1994, followed by a revised edition (DC:0-3R) in 2005 (Zero to Three, 1994, 2005). A key purpose was to respond “to the failure of the DSM system to include (1) sufficient coverage of syndromes of early childhood that needed clinical attention or (2) sufficient consideration of developmental features of early disorders” (Zero to Three, 2005, p 4). Like the DSM and ICD, the DC:0-3 was developed by having experts negotiate diagnostic categories and criteria.

DC:0-3R Axes

The DC:0-3R includes the following five axes:

- I. Clinical disorders
- II. Relationship classification
- III. Medical and developmental disorders and conditions
- IV. Psychosocial stressors
- V. Emotional and social functioning.

Some of the Axis I clinical disorders, such as *posttraumatic stress disorder*, *separation anxiety disorder*, and *generalized anxiety disorder*, have counterparts in DSM-IV-TR. However, as they are designed for ages 0-3, the DC:0-3R criteria are quite different from DSM-IV-TR criteria and are illustrated with clinical examples for ages 0-3. Other DC:0-3R disorders do not have clear counterparts in DSM-IV-TR. Examples include *deprivation/maltreatment disorder*, which “occurs in the context of deprivation or maltreatment” (p 17); *prolonged bereavement/grief reaction*; *mixed disorder of emotional expressiveness*, “characterized by a child’s difficulty in expressing a developmentally appropriate range and intensity of emotions over at least a 2-week period” (p 27); and *hypersensitive*, referring to hypersensitivity to sensory stimuli. A diagnostic category designated as *multisystem developmental disorder* is designed for children less than 2 years old who manifest problems like those specified for pervasive developmental disorder in the DSM-IV-TR.

The DC:0-3R Axis II *Relationship Classification* has no counterpart in DSM-IV-TR, but is considered to be especially important for young children. As a basis for classifying relationships, the DC:0-3R provides the 100-point Parent-Infant Relationship Global Assessment Scale (PIR-GAS), on which clinicians rate parent-infant relationships. PIR-GAS ratings of 81-100 indicate *adapted relationships*, ratings of 41-80 indicate *features of a disordered relationship*, and ratings of 0-40 indicate a *disordered relationship*. The DC:0-3R also provides the Relationship Problems Checklist for clinicians to rate the following qualities of parent-infant relationships: overinvolved, underinvolved, anxious/tense, angry/hostile, verbally abusive, physically abusive, and sexually abusive. For each of these qualities, clinicians are to indicate whether there is no evidence, some evidence needing



Click on the picture to access the DSM-5 website that gives details of proposed changes to the classification and their rationale.

DC:0-3 was developed to respond to the DSM system not including sufficient coverage of syndromes of early childhood and of their developmental features.

further investigation, or substantial evidence, based on their judgments of the behavioral quality of interaction, affective tone, and psychological involvement.

The DC:0-3R Axis V *Emotional and Social Functioning* might be considered somewhat analogous to the DSM-IV-TR Axis V Global Assessment of Functioning. However, the DSM-IV-TR Axis V is assessed with a single 100-point scale for rating psychological, social, and occupational functioning. By contrast, the DC:0-3R Axis V comprises clinicians' 6-point ratings of the following capacities for emotional and social functioning (p. 62):

- Attention and regulation
- Forming relationships/mutual engagement
- Intentional two-way communication
- Complex gestures and problem solving
- Use of symbols to express thoughts/feelings
- Connecting symbols logically/abstract thinking

The DC:0-3R Axis III *Medical and Developmental Disorders and Conditions* and Axis IV Psychosocial Stressors are roughly analogous to the DSM-IV-TR Axis III General Medical Conditions and Axis IV Psychosocial and Environmental Problems, although the content and clinical examples for the DC:0-3R axes are specific to ages 0-3.

For practitioners who work with infants and young children, the DC:0-3R certainly touches on many more relevant aspects of functioning, problems, and issues than the ICD or DSM-IV-TR. However, proper use of the DC:0-3R's five axes requires extensive training of practitioners, as well as extensive interactions with each child and its caregivers under multiple conditions. Further research is needed to determine how to train practitioners to use the DC:0-3R with adequate reliability, validity, and utility.

GAPS BETWEEN NOSOLOGICAL MODELS AND CLINICAL ASSESSMENT

The ICD, DSM, and DC:0-3R nosologies are intended to embody concepts of disorders distilled from inputs by many experts. Except for use of standardized cognitive tests to determine whether children meet diagnostic criteria for mental retardation or learning disorders, the nosologies do not specify clinical assessment procedures on which to base judgments of each criterion for making diagnoses of particular disorders. As an example, the DSM-IV-TR criteria for ADHD list nine symptoms of inattention (e.g., "is often forgetful in daily activities") and nine symptoms of hyperactivity-impulsivity (e.g., "often talks excessively"). For a child to meet the overall diagnostic criterion for ADHD, the practitioner must decide that at least 6 of the 9 symptoms from at least one list "have persisted for at least 6 months to a degree that is maladaptive and inconsistent with developmental level." The practitioner must also decide that some symptoms caused impairment before the child was 7 years old, that the symptoms currently cause impairment in at least two settings (e.g., home and school), and that there is clear evidence "of clinically significant impairment in social, academic, or occupational functioning" (American Psychiatric Association, 2000, pp92-93).

To diagnose children, practitioners must decide what information to obtain for making yes-versus-no judgments of each criterion for all disorders that the



Uniformed letter carrier with child in mailbag (c 1900)

This postman posed for a humorous photograph with a young boy in his mailbag. After parcel post service was introduced in the US in 1913, at least two children were sent by the service. With stamps attached to their clothing, the children rode with railway and city carriers to their destination. The Postmaster General quickly issued a regulation forbidding the sending of children in the mail after hearing of those examples (National Postal Museum, US)

children might have, how to obtain the information, and from whom. In most cases, practitioners would want to interview the child and at least one parent figure. However, interview time is precious and is seldom sufficient to assess all potentially relevant symptom criteria. For each ADHD symptom that they report to the practitioner, children and parents may not be able to accurately say whether the symptom caused impairment before age 7 and whether it currently causes impairment in at least two settings.

If a child attends school, the practitioner would typically want to obtain information from the child's teacher(s) to provide a picture of the child's functioning in school. In fact, to judge whether the child meets criteria for ADHD, the practitioner typically needs to know whether at least 6 of the 9 symptoms from one or both lists cause impairment in the school setting. Because teachers often instigate referrals for help and because many child problems interfere with functioning in school, teachers' input can be valuable in multiple ways, in addition to being needed to judge whether particular symptoms cause impairment in school. Because practitioners are seldom able to interview teachers, other methods are needed to obtain information from them.

Whatever the assessment methods, reports by parents, teachers, and children often differ with respect to criteria required for diagnoses (De Los Reyes, 2011). Consequently, practitioners may find it hard to make yes-versus-no diagnostic judgments on the basis of conflicting information from different sources. Practitioners may also be handicapped by the failure of nosologies to

specify procedures for assessing behavioral, emotional, and social problems in different settings and by gaps between what parents, teachers, and children are able to report versus the diagnostic requirements for yes-versus-no judgments. Practitioners therefore need practical methods for assessing children's problems in multiple settings and for dealing with discrepancies between different sources of information.

Standardized diagnostic interviews

Various standardized diagnostic interviews (SDIs) have been developed to obtain assessment data from informants in order to determine whether criteria for diagnoses are met. Parallel versions of several SDIs are available for administration to children and their parents, although agreement between diagnoses made from child and parent interviews is typically low.

Among the SDIs for diagnosing children, the *Diagnostic Interview Schedule for Children* (DISC; Shaffer et al, 2000) is perhaps the most widely used. The DISC is highly structured. This means that it comprises narrowly focused questions for obtaining yes-versus-no answers as to whether each DSM diagnostic criterion is met. Because the respondent's yes-versus-no answers are the basis for determining whether each criterion is met, highly structured SDIs like the DISC are called "respondent-based" interviews. Neither clinical skills nor clinical judgments are needed to administer the DISC. However, intensive training is needed to learn how to ask the questions properly, to record the answers, and to respond flexibly when answers cannot be clearly coded as yes-versus-no. To cover all the criteria for symptoms, impairment, occurrence in multiple settings, duration, age of onset, etc. for all the DSM diagnoses relevant to children, the DISC contains thousands of questions. There are "skip-out" options for skipping questions for some diagnoses if it appears that a child will not meet criteria for those diagnoses. However, administration of the DISC usually takes at least an hour for each informant (e.g., each parent and child) and may take several hours for interviews regarding clinically referred children who have many problems.

In addition to respondent-based SDIs like the DISC, a second type of SDI is known as "interviewer-based." These SDIs must be administered by clinically trained interviewers who can phrase the SDI's semistructured questions in ways that are appropriate for each interviewee. The interviewers must also be trained to judge the interviewee's responses as indicating whether the child does or does not meet each criterion for each diagnosis. The most widely used interviewer-based SDI is the *Kiddie Schedule for Affective Disorders and Schizophrenia* (K-SADS; Kaufman et al, 1997).

A third type of SDI combines aspects of respondent-based and interviewer-based SDIs. An example is the *Development and Well-being Assessment* (DAWBA; Goodman et al, 2000). The DAWBA includes structured interviews for administration to parents and 11- to 16-year-olds. The DAWBA's interviews include some open-ended questions for following up on the interviewees' responses to the structured questions. A brief questionnaire is used to obtain data from teachers. The data from parents, teachers, and youths are brought together by a computer program that generates likely diagnoses. Experienced clinical raters then view the computer output and decide whether to accept or overturn the computer-generated diagnoses.

Relations between standardized diagnostic interviews and clinical diagnoses

SDIs are now widely used in research on both child and adult disorders. Consequently, many published findings about associations between diagnoses and other characteristics are based on diagnoses made with SDIs. However, owing to the cost and time required to train interviewers and to administer SDIs, they are seldom used in clinical practice. We must therefore ask, would children receiving particular diagnoses from SDIs receive the same diagnoses from clinical evaluations? To answer this question, a meta-analysis was performed on data from 38 studies that reported agreement between diagnoses made from SDIs and from clinical evaluations of the same patients (Rettew et al, 2009). Agreement was expressed in terms of the percentage of cases that received the same diagnoses from SDIs and from clinical evaluations. The *kappa* statistic (Cohen, 1960) was used to correct for chance agreement. The agreement between children's SDI diagnoses and their diagnoses made from clinical evaluations averaged 39%, after correction for chance. In other words, SDIs and clinical evaluations yielded the same diagnoses for 39% of children but different diagnoses for 61% of children. The agreement percentages varied considerably among diagnoses, with a low of 19% for generalized anxiety disorder and a high of 86% for anorexia nervosa (to obtain large enough numbers, data for children and adults were combined when computing these percentages). For some disorders, such as anorexia nervosa, the high agreement indicates that findings with SDI diagnoses can be applied to clinical diagnoses. However, the low agreement for other diagnoses, such as generalized anxiety disorder, indicates that findings with SDIs cannot be applied to clinical diagnoses.

Because various SDIs have been used to obtain the diagnoses reported in published studies, we also need to ask whether each child would receive the same diagnoses from different SDIs. Surprisingly little research has been published on this important question. In what may be the only published study of agreement between child diagnoses made from the most widely used SDIs, only 3% of the diagnoses made from DISC interviews of children and their parents agreed with diagnoses made from K-SADS interviews of the same children and their parents (Cohen et al, 1987). Thus, for many children, we cannot expect different SDIs to yield the same diagnoses nor can we expect SDIs to yield the same diagnoses as clinical evaluations.

Instruments for rating children's problems

To help practitioners obtain information from parents, teachers, and children, various standardized rating instruments have been developed. Many of these instruments are designed to be self-administered by people who can read, but they can also be administered by lay interviewers who read the items aloud and enter the respondent's answers on a rating form or computer keyboard. If a practitioner looks at the results of rating instruments completed by one or more parent figures, the practitioner can then tailor precious interview time to focus on problem areas revealed by the results and on issues not tapped by rating instruments. By using instruments that provide parallel forms completed by parent figures, teachers, and the child, the practitioner can compare results to identify agreements and discrepancies. The practitioner can then use interviews and other means to investigate further. The use of parallel rating instruments completed by parents, teachers, and children enables practitioners to obtain and compare a great

deal of information from multiple informants at little cost in practitioner time. Rating instruments are also available for completion by people who provide day care for young children, observers who rate children's behavior in group settings such as classrooms, and practitioners who administer clinical interviews and/or standardized ability tests (Achenbach, 2009).

The items of some rating instruments are intended to approximate the symptom criteria embodied in official nosologies. An example is the *ADHD Rating Scale* (DuPaul et al, 1998), which has 18 items approximating the nine inattention and nine hyperactivity-impulsivity symptom criteria for DSM-IV-TR ADHD. Unlike the DSM-IV-TR criteria, however, each ADHD Rating Scale item is rated 0 = *Never or rarely*, 1 = *Sometimes*, 2 = *Often*, or 3 = *Very often*. The ratings are summed to yield separate scores for scales comprising the inattention and hyperactivity-impulsivity items. Clinical cutpoints have been established on distributions of scale scores obtained from parents' and teachers' ratings of samples of children in the US. Thus, although the items approximate the DSM-IV-TR symptom criteria, the raters are not required to make yes-versus-no responses and practitioners can see how scale scores compare with normative distributions of scale scores and with cutpoints on those distributions of scale scores.

Other rating instruments also focus mainly on a particular diagnostic category such as ADHD but include additional problems as well, e.g., the *Conners Rating Scales-Revised* (CRS-R; Conners, 2001). Still other rating instruments include more diverse items that are scored on scales constructed by a combination of statistical analyses and the authors' decisions about which items to assign to particular scales. The *Strengths and Difficulties Questionnaire* (SDQ; Goodman, 1997), for example, has 25 items, which are scored on five scales constructed by the author. The *Behavior Assessment System for Children-2* (BASC-2; Reynolds & Kamphaus, 2004) is a much longer instrument that is scored on numerous scales based on a combination of statistical analyses and the authors' decisions about which items to assign to particular scales. Instruments such as the ADHD Rating Scale, CRS-R, SDQ, and BASC-2 are useful for obtaining informants' ratings of various kinds of problems, but they do not close the gaps between clinical assessment and nosological models any more than SDIs do.

CLINICAL MODELS DERIVED FROM RATINGS OF CHILDREN'S PROBLEMS

The nosological models have been constructed largely via "top-down" procedures. In other words, construction of the nosological models has started "at the top" with diagnostic categories proposed by experts. For each category, the experts have then proposed criteria for determining whether a child's problems fit the category. Although input from other experts and from field trials has been used to adjust some criteria, the diagnostic categories and criteria have not been derived from data on actual samples of children. Furthermore, the diagnostic criteria are the same for boys and girls, all ages, all sources of information, and all societies.

Just as multiple nosological models may co-exist, so too, alternative kinds of models may co-exist with nosological models. One alternative has involved using "bottom-up" procedures to derive clinical models from ratings of many children. This has been done by having parents, teachers, and children rate diverse problems

as 0 = *Not true*, 1 = *Somewhat or sometimes true*, and 2 = *Very true or often true* for thousands of children referred for mental health services and thousands of children sampled from general populations (Achenbach & Rescorla, 2000; 2001). The items were iteratively selected and refined on the basis of input from practitioners in fields related to mental health, from research on problems reported for children referred for mental health services, and from analyses of ratings and suggestions provided by parents, teachers, and children on successive pilot editions of the rating instruments. Rather than being selected on the basis of experts' diagnostic concepts, the final sets of items have been selected on the basis of their ability to discriminate significantly between children who were referred for mental health or special education services versus demographically similar children who were not referred for services. Important criteria for retention of problem items were that (a) they could be understood and rated by parents, teachers, and/or children, and (b) they were rated significantly higher for referred children than for demographically similar non-referred children.

Although many of the items have counterparts among the nosological criteria, other items were also found to discriminate between referred and non-referred children even though they were not among the nosological criteria. Equally important, the items were worded so that parents, teachers, and children could rate them, rather than being worded in terms of nosological criteria that parents, teachers, and children may not understand. The final versions of the rating instruments can be completed in about 15 minutes and include assessments of competencies as well as problems. The instruments include the *Child Behavior Checklist* for ages 1½-5 and 6-18, the *Caregiver-Teacher Report Form* for ages 1½-5, the *Teacher's Report Form* for ages 6-18, and the *Youth Self-Report* for ages 11-18 (Achenbach & Rescorla, 2000; 2001).

Syndrome models

After the rating instruments were developed, ratings of thousands of children were statistically analyzed to identify actual patterns (i.e., syndromes) of problems that tend to occur together. The syndromes were given names that summarize their constituent problems. For ages 6-18, the syndromes are designated as *Anxious/Depressed*, *Withdrawn/Depressed*, *Somatic Complaints* (without known medical cause), *Social Problems*, *Thought Problems*, *Attention Problems*, *Rule-Breaking Behavior*, and *Aggressive Behavior*. Counterparts of several of these syndromes were found for ages 1½-5, but syndromes designated as *Emotionally Reactive* and *Sleep Problems* were also found. Children's scores on each syndrome are computed by summing the 0-1-2 ratings of the syndrome's items. To make it easy for practitioners to compare a child to non-referred peers, the syndrome scores are displayed on profiles in relation to norms for the child's age and gender, the type of informant (parent, teacher, self), and the relevant society (explained later).

Additional statistical analyses have identified associations among subsets of the syndromes that are designated as *Internalizing* (anxiety, depression, social withdrawal, somatic complaints) and *Externalizing* (rule-breaking and aggressive behavior). Scores for Internalizing and Externalizing are computed by summing the 0-1-2 ratings of the items comprising these broad groupings. A *Total Problems* score is also computed by summing the 0-1-2 ratings of all the problem items on a form. Scores for Internalizing, Externalizing, and Total Problems are displayed on

profiles in relation to norms for the child's age and gender, the type of informant, and the relevant society.

DSM-oriented scales

The items of the rating instruments were selected through bottom-up procedures that started with ratings of children rather than top-down procedures that started with diagnostic categories. Nevertheless, some of the problem items have counterparts among diagnostic criteria and some of the statistically identified syndromes comprise problems like those of certain diagnostic categories. For example, the Attention Problems syndrome includes many problem items analogous to the symptom criteria for ADHD. Furthermore, significant associations have been found between certain diagnoses and high scores on certain syndromes (Achenbach & Rescorla, 2001).

In order to help practitioners see relations between diagnostic categories and the data obtained with the rating instruments, expert psychiatrists and psychologists from many cultures were asked to identify rating instrument items that they considered to be very consistent with DSM-IV-TR diagnostic categories. The items identified by large majorities of experts for particular diagnostic categories were then grouped into DSM-oriented scales corresponding to the diagnostic categories. For ages 6-18, the DSM-oriented scales are designated as *Affective Problems*, *Anxiety Problems*, *Somatic Problems*, *Attention Deficit Hyperactivity Problems*, *Oppositional Defiant Problems*, and *Conduct Problems*. For ages 1½-5, the experts' judgments yielded several DSM-oriented scales analogous to those for ages 6-18, but also a scale designated as *Pervasive Developmental Problems* (Achenbach & Rescorla, 2000). This scale has been found to identify children diagnosed as having autism spectrum disorders (Muratori et al, 2011; Sikora et al, 2008). A child's score on each DSM-oriented scale is computed by summing the 0-1-2 ratings of the problem items comprising the scale. To enable practitioners to compare the child's scores with scores for non-referred peers, the scale scores are displayed on profiles in relation to scores for the child's age and gender, the type of informant, and the relevant society, as illustrated in Figure A.3.1 for self-ratings by 16-year-old Richard (not his real name). The DSM-oriented scales, the syndrome scales, and their respective profiles thus provide alternative models for understanding and using informants' ratings of children's behavioral, emotional, and social problems on the same rating instruments.

Cross-informant comparisons

Because parents, teachers, and children are aware of different aspects of children's functioning, discrepancies are often found between the problems reported by different informants. When discrepancies are found, it is tempting to conclude that one informant is right while another is wrong. However, different informants may provide useful information about differences in how a child functions in different contexts and with different people. Discrepancies between informants' reports may also reflect differences between the informants' perceptions and evaluations of children's functioning. These differences need to be considered when deciding whether a child needs mental health services, what kind of services, and how parents, teachers, and the child may participate in the services.

Figure A.3.1
Richard's
profile
of DSM-
oriented YSR
scales (from
Achenbach
& Rescorla,
2007).

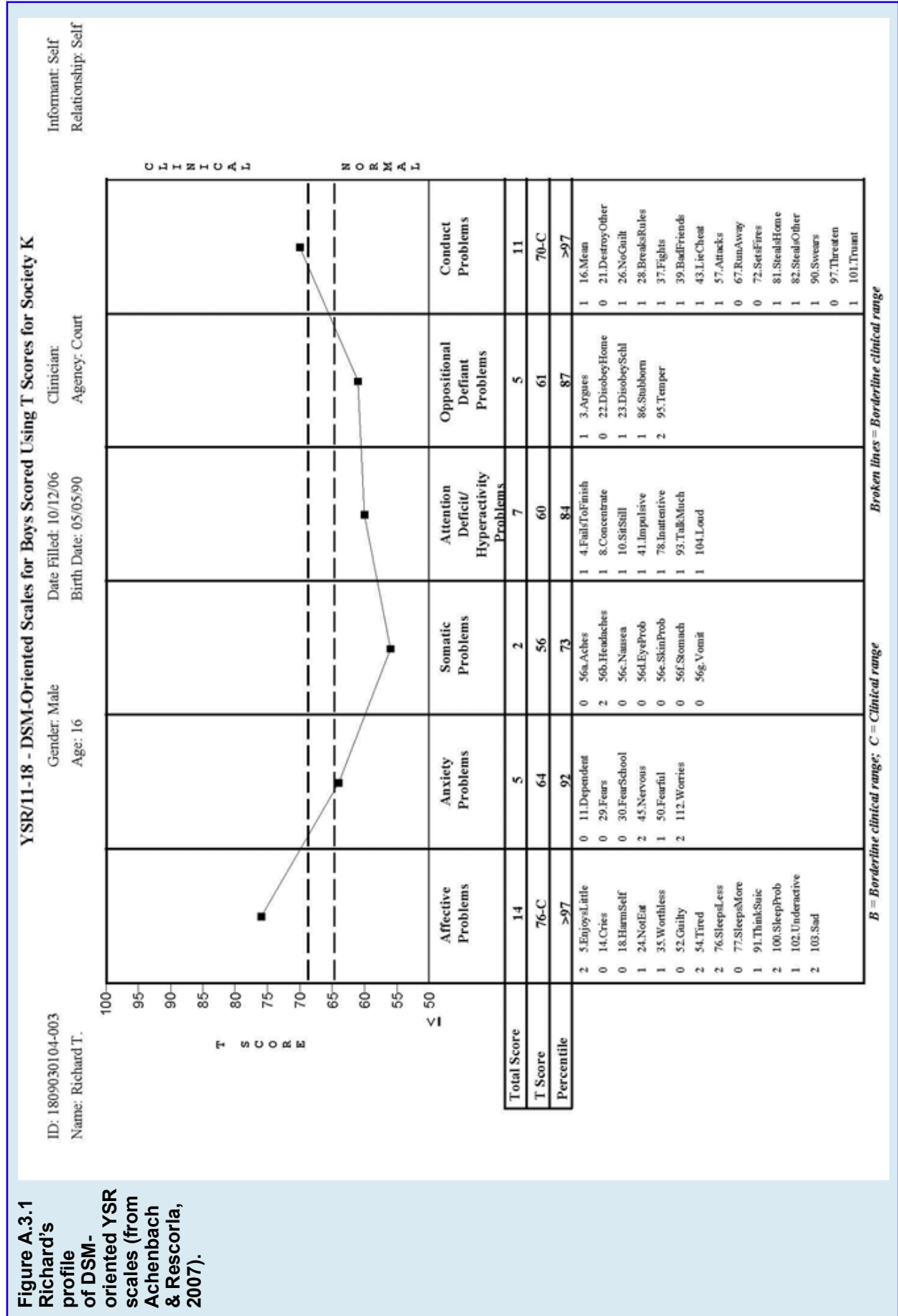
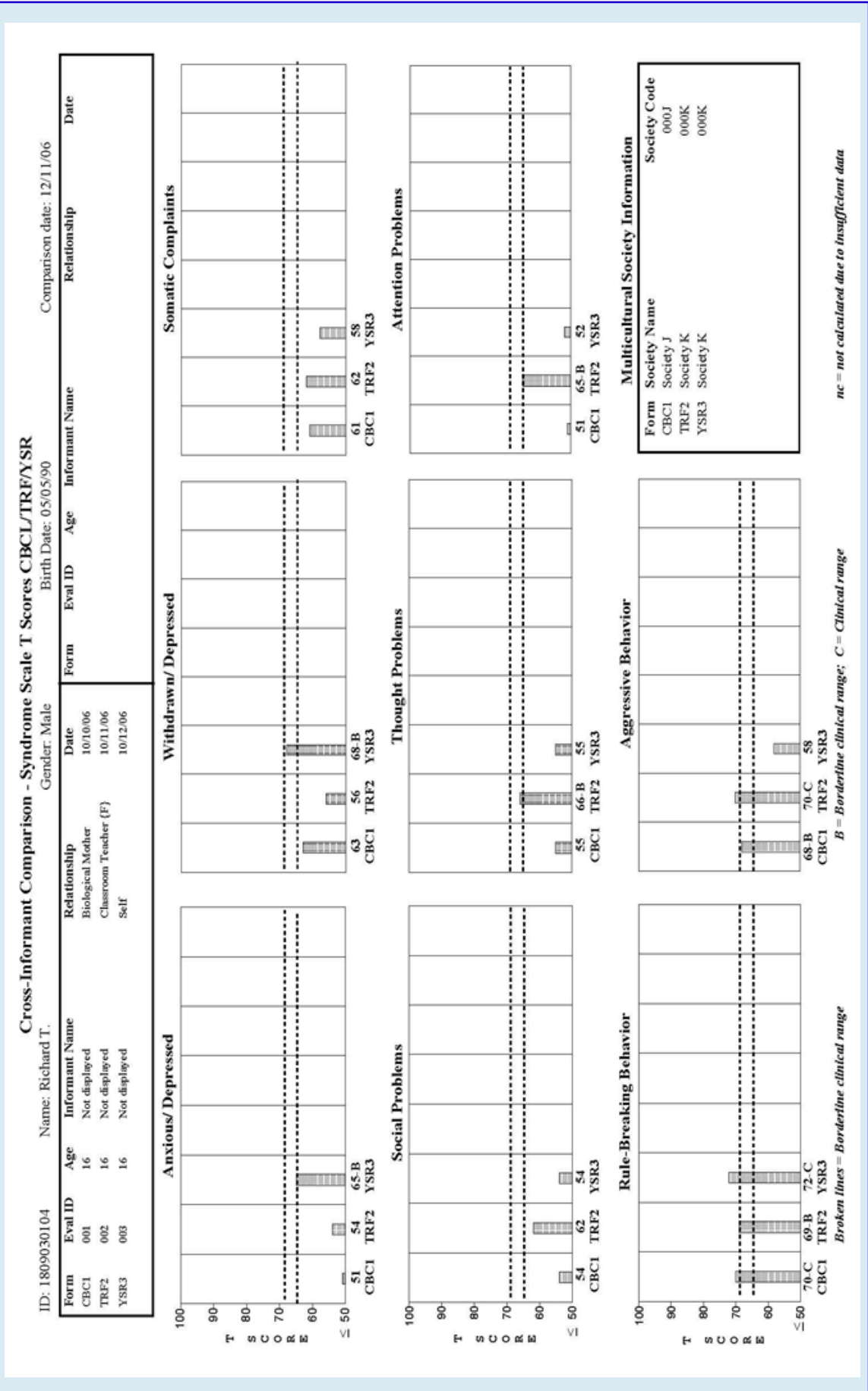


Figure A.3.2 Cross-informant comparisons of Richard's scores on syndrome scales in relation to Society J norms for the CBCL and Society K norms for the TRF and YSR (from Achenbach & Rescorla, 2007).



To make it easy for practitioners to detect similarities and discrepancies between problems reported by parents, teachers, and children, the computer software for using the rating instruments produces side-by-side displays of the 0-1-2 ratings of each problem item by any combination of up to eight informants. The side-by-side displays of item ratings enable the practitioner to quickly identify problems that are endorsed by all informants, problems that are not endorsed by any informants, and problems that are endorsed by some informants. As an example, if certain problems are endorsed only by teachers, this would suggest that these problems are specific to the school context. On the other hand, if certain problems are endorsed only by parents, this would suggest that these problems are specific to the family context.

In addition to the side-by-side displays of problem ratings, the computer software also displays bar graphs of syndromes and DSM-oriented scales scored from each informant's ratings. As an example, Figure A.3.2 shows the bar graphs of DSM-oriented scales scored from ratings of 16-year-old Richard by his mother, teacher, and Richard himself. By looking at the bar graphs, the practitioner can quickly see whether any scores are in the borderline clinical range (between the two broken lines) or in the clinical range (above the top broken line). The practitioner can also see whether there are important differences between scores obtained from different informants.

Multicultural applications

The ICD, DSM, and DC:0-3R models have not been extensively tested outside a handful of societies. Consequently, much remains to be learned about how well their symptom criteria, diagnostic thresholds (i.e., the number of symptoms required for diagnoses), and other features apply to children from the many societies where the nosological models have not been tested. As significant changes are likely in ICD-11 and DSM-5, it is to be hoped that drafts of the new criteria will be subjected to multicultural tests before the criteria are finalized.

Meanwhile, clinical models derived from data on problems rated for ages 1½-18 have been tested in 47 societies, and translations of the rating instruments are available in the 86 languages listed in Table A.3.1. The syndrome models were tested by statistically analyzing ratings of tens of thousands of children by their parents, teachers, daycare providers, and—for ages 11-18—the children themselves (Achenbach & Rescorla, 2012). Nearly all of the analyses supported the syndrome models that were originally derived from ratings of Anglophone children from Australia, Jamaica, the UK, and the US. The findings mean that the patterns of co-occurring problems embodied in the syndromes scored from the parent, teacher, and self-rating instruments were also found in the other societies.

To provide norms appropriate for children in the 47 societies, distributions of Total Problems scores were compared for the samples of children from all the societies. Based on these comparisons, societies with relatively low, medium, or high problem scores were identified. To enable practitioners to compare a child's scores with scores obtained by peers in the appropriate society, norms have been constructed for low-scoring, medium-scoring, and high-scoring societies. The three sets of multicultural norms are available in computer-scoring software that enables practitioners to display a child's syndrome, DSM-oriented, Internalizing,



Click on the picture
to access the ASEBA
website.

Table A.3.1 Translations of the Achenbach System of Empirically Based Assessment (ASEBA) forms^a

Afaan Oromo (Ethiopia)	Georgian	Portuguese (Brazil)
Afrikaans	German	Portuguese (Portugal)
Albanian (Albania, Kosovo)	Greek	Portuguese Creole
American Sign Language	Gujarati (India)	Punjabi (India)
Amharic (Ethiopia)	Haitian Creole	Romanian
Arabic	Hebrew	Russian
Armenian	Hindi	Samoan
Auslan (Australian Sign Language)	Hungarian	Sepedi (Northern Sotho, South Africa)
Bahasa (Indonesia)	Icelandic	Serbian
Bahasa (Malaysia)	Italian	Sesotho (Southern Sotho, South Africa)
Bangla (Bangladesh)	Japanese	Sinhala (Sri Lanka)
Bengali	Kannada (India)	Slovak
Bosnian	Khmer (Cambodia)	Slovenian
Braille	Kiambu (Kenya)	Somali
British Sign Language	Kiswahili (Kenya)	Spanish (Castilian)
Bulgarian	Korean	Spanish (Latino)
Catalan (Spain)	Laotian	Swedish
Chinese	Latvian	Tagalog (Philippines)
Croatian	Lithuanian	Tamil (India)
Czech	Macedonian	Telugu (India)
Danish	Malay	Thai
Dutch (Netherlands, Flanders)	Malayalam (India)	Tibetan
Estonian	Maltese	Turkish
Farsi (Persian, Iran)	Marathi (India)	Ukrainian
Finnish	Nepalese	Urdu (India, Pakistan)
French (Belgian)	Norwegian	Vietnamese
French (Parisian)	Papiamentu (Curacao)	Visayan (Philippines)
French (Quebecois)	Pashto (Afghanistan, Pakistan)	Xhosa (South Africa)
Ga (Ghana)	Polish	Zulu

^aLanguages into which at least one ASEBA form has been translated (from Achenbach & Rescorla, 2012).

Externalizing, and Total Problems scores in relation to norms for the appropriate society, as well as for the child's age and gender and the type of informant.

OVERVIEW OF CLINICAL MODELS

We have outlined several models for distinguishing among various kinds of behavioral, emotional, and social problems. At this stage of mental health science, no single model can be regarded as totally correct, with the others being regarded as wrong. Instead, the different models may have different strengths and weaknesses for different purposes. Consequently, for the foreseeable future, it may be helpful to think in terms of multiple models for guiding the collection and use of clinical data for evidence-based practice.

The ICD and DSM models for childhood behavioral, emotional, and social problems are components of nosologies that cover many kinds of problems across the lifespan. Like the models for other age groups, the models for children's problems comprise diagnostic categories proposed by experts who then specify criteria for determining which categories a child's problems match. The criteria require the same yes-versus-no judgments for children of both genders, different ages, and different cultural backgrounds, as well as for data obtained from different sources. In evaluating DSM-IV and ICD-10, the world's most famous child psychiatrist, Sir Michael Rutter (2011), concluded that "there are far too many diagnoses, and a ridiculously high rate of supposed comorbidity" (p. 647), and that "the use of dimensions could reduce the extent of misleading supposed comorbidity" (p. 655).

Like the ICD and DSM, the DC:0-3R is a nosology whose categories were proposed by experts who then specified criteria for determining which categories a child's problems match. Unlike the ICD and DSM, however, the DC:0-3R was designed to provide better coverage of early childhood syndromes and developmental features. Also unlike the ICD and the DSM, the DC:0-3R provides rating scales for parent-infant relationships, which are regarded as especially important for understanding the problems of very young children.

The failure of the ICD and DSM to specify procedures for assessing behavioral, emotional, and social problems has spurred the development of various SDIs. In recognition of the need for data from parents as well as from the child, SDIs designed for diagnosing children include interviews for parents. Some SDIs also provide questionnaires for teachers. However, differences between parent, teacher, and child reports make it hard to draw yes-versus-no conclusions about each diagnostic criterion.

As SDIs are widely used to make diagnoses for research purposes, many publications report findings for diagnoses made with SDIs. However, the extensive training and large blocks of time needed to administer SDIs limit their use in clinical practice. Consequently, it must be asked whether research findings for diagnoses made from SDIs can be applied to diagnoses made from clinical evaluations. Meta-analyses indicate good agreement between certain diagnoses—such as anorexia nervosa—made from SDIs and clinical evaluations, but poor agreement for some

"There are far too many diagnoses, and a ridiculously high rate of supposed comorbidity."

"The use of dimensions could reduce the extent of misleading supposed comorbidity."

Sir Michael Rutter (2011, p 655)

common diagnoses, such as generalized anxiety disorder. Across all diagnoses in the meta-analyses, SDIs and clinical evaluations yielded the same diagnoses for 39% of the child patients but different diagnoses for 61%, after correction for chance.

In addition to SDIs that yield yes-versus-no judgments about diagnostic criteria, various instruments have been developed for rating children's problems. Some rating instruments comprise items approximating the symptom criteria for a particular diagnostic category, such as ADHD. Other instruments include more diverse items that may approximate certain diagnostic criteria but do not include one-to-one counterparts of diagnostic categories.

Another approach to constructing clinical models is by working from the bottom-up by starting with informants' ratings of diverse problem items for thousands of children. The ratings are then analyzed statistically to identify syndromes of co-occurring problems. For use in clinical evaluations, children's syndrome scores are computed by summing the ratings of the items comprising each syndrome. To enable practitioners to determine whether scores are in the normal range, borderline clinical range, or clinical range, the syndrome scales are displayed on profiles in relation to norms for the child's age and gender, as well as for the type of informant, such as parents, teachers, and children themselves. To take account of differences between problems reported for children in different societies, the practitioner can select norms based on multicultural data from many societies. And to help the practitioner make cross-informant comparisons, item and scale scores from multiple informants can be displayed side-by-side. The scale scores from each informant are standardized for the child's age and gender, the type of informant, and the multicultural norm group selected by the practitioner.

Pediatric clinic in Kenya



An 8-year-old Kenyan boy was brought to a local medical clinic by his mother. About 2 years earlier, she had noticed that he seemed very restless and that other children refused to play with him. He was very distractible and had to be supervised carefully. Home helpers employed by the mother often quit because they could not put up with him. He sustained a variety of injuries, including bruises, fractures, and burns. His teachers complained that he was very careless, had poor handwriting, and did not complete his work. He blurted out answers when it was not his turn and was very disruptive in class. Although he appeared bright, he was easily distracted and was quarrelsome during playtime, wanting to be in the center of all the children's activities. He lost important school items, prompting his mother to visit the school often in order to look for lost property and to attend disciplinary meetings.

To provide model-based assessment data, the mother was asked to complete the Conners Parent Rating Scale (CPRS). Based on elevated scores on multiple CPRS problem scales, the mother's interview reports, and a brief session with the boy, the clinic pediatrician recommended that the boy be seen by a psychiatrist for a more extensive evaluation than the local medical clinic could provide.

When the psychiatrist evaluated him, he noted that the boy was well-groomed but that he had multiple bruises at different stages of healing. The bruises appeared to be the results of falls and bumping into objects rather than having been inflicted by other people, however, he said that he was often punished at home because of his behavior and asked the doctor to talk to his parents about these punishments. A mental status examination revealed he was

well-oriented in time and place, his memory was good, and there was no evidence of perceptual disturbances. However, he was easily distracted and made many inappropriate noises. He could not sit still, walked around the room, and asked the doctor questions about pictures on the office walls. His language skills were poorly developed, and he did not express ideas clearly.

In reviewing the CPRS results forwarded by the boy's pediatrician, the psychiatrist saw that scores were elevated on multiple scales, with especially high scores on the ADHD Index. Although the boy clearly needed help in multiple areas, the psychiatrist concluded that the high scores for ADHD argued for a trial of medication to see whether it would reduce problems of inattention and hyperactivity sufficiently to make the boy amenable to help with other problems.

When discussing his conclusions with the boy's mother, the psychiatrist found that she had already learned about ADHD from a friend whose child had similar problems. Consequently, the mother was not surprised by the psychiatrist's conclusions and was willing to have a trial of medication. With the mother's permission, the psychiatrist also communicated his conclusions and plan to the boy's teacher, who was also concerned about his behavior. The psychiatrist prescribed 10 mg of methylphenidate twice daily and asked mother and teacher to complete the CPRS at monthly intervals to evaluate response to treatment and as a basis for deciding on possible further interventions. Mother was also given information about ADHD and about more appropriate disciplinary measures.

The statistically derived syndromes provide bottom-up models for understanding children's problems. However, top-down models are also provided for scoring the same rating instruments in terms of DSM-oriented scales based on clinical judgments by experts from many cultures.

THE NEED FOR MODEL-BASED ASSESSMENT IN DEVELOPING COUNTRIES

The needs for mental health services in developing or low income countries are apt to be at least as great as in developed countries, but low income countries have far fewer skilled mental health workers. The photo shows one of the authors in a clinic in a low income country where children are waiting to be seen by a nurse – who is overwhelmed by the sheer number of patients. There is thus an urgent need to use assessment instruments that may assist overworked health care providers. These instruments should be able to obtain reliable information from parents and – whenever possible – teachers and the children themselves. Such instruments can serve, at least, the following purposes: (a) to increase awareness of mental disorders in children; (b) to integrate data about these problems into health information systems; and (c) to assist in identifying cases for referral to specialists, if needed. The case vignette above illustrates a model-based assessment in a developing country.

REFERENCES

- Achenbach TM (2009). *The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- Achenbach TM & Rescorla LA (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- Achenbach TM & Rescorla LA (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- Achenbach TM & Rescorla LA (2012). *Multicultural Guide for the ASEBA Forms & Profiles for Ages 1½-18 (2nd ed)*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, text revision. Washington, DC: Author.
- Cohen J (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37-46. doi:10.1177/001316446002000104
- Cohen P, O'Connor P, Lewis S et al (1987). Comparison of DISC and K-SADS-P interviews of an epidemiological sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26:662-667. doi:10.1097/00004583-198709000-00009
- Conners CK (2001). *Conners' Rating Scales-Revised Manual*. North Tonawanda, NY: Multi-Health Systems.
- De Los Reyes A (2011). More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 40:1-9. doi:10.1080/15374416.2011.533405
- DuPaul GJ, Power TJ, Anastopoulos AD et al (1998). *ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation*. New York: Guilford.
- Goodman R (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38:581-586. doi:10.1111/j.1469-7610.1997.tb01545.x
- Goodman R, Ford T, Richards H et al (2000). The Development and Well-Being Assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 41:645-655. doi:10.1111/j.1469-7610.2000.tb02345.x
- Helzer JE, Kraemer HC, Krueger RF et al (Eds) (2008). *Dimensional Approaches in Diagnostic Classification: Refining the Research Agenda for DSM-V*. Washington, DC: American Psychiatric Association.
- Kaufman J, Birmaher B, Brent D et al (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36:980-988. doi:10.1097/00004583-199707000-00021
- Muratori F, Narzisi A, Tancredi R et al (2011). The CBCL 1.5-5 and the identification of preschoolers with autism in Italy. *Epidemiology and Psychiatric Sciences*, 20:329-338. doi:10.1017/S204579601100045X
- Rettew DC, Doyle A, Achenbach TM et al. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research*, 18:169-184. doi:10.1002/mpr.289
- Reynolds CR & Kamphaus RW (2004). *Behavior Assessment System for Children (2nd ed)*. Circle Pines, MN: American Guidance Service.
- Rutter M (2011). Child psychiatric diagnosis and classification: concepts, findings, challenges and potential. *Journal of Child Psychology and Psychiatry*, 52:647-660. doi:10.1111/j.1469-7610.2011.02367.x
- Shaffer D, Fisher P, Lucas CP et al (2000). NIMH Diagnostic Interview Schedule for Children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39:28-38. doi:10.1097/00004583-200001000-00014
- Sikora DM, Hall TA, Hartley SL et al (2008). Does parent report of behavior differ across ADOS-G classifications: Analysis of scores from the CBCL and GARS. *Journal of Autism and Developmental Disorders*, 38:440-448. doi:10.1007/s10803-007-0407-z
- World Health Organization (1992). *Mental Disorders: Glossary and Guide to their Classification in Accordance with the Tenth Revision of the International Classification of Diseases (10th ed)*. Geneva: World Health Organization.
- World Health Organization (1993). *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. Geneva, Switzerland: Author.
- Zero to Three: National Center for Infants, Toddlers, and Families (1994; 2005). *Diagnostic Classification 0-3. Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood (1st edition, revised edition)*. Washington, DC: Author.