

## GENERAL PAPERS

# A Participant's Observations: Preparing DSM-IV

David Shaffer, MB, BS, FRCP, FRCPsych<sup>1</sup>

**Objective:** To provide an overview of the process of creating the DSM-IV from a participant's perspective.

**Method:** Narrative review.

**Results:** In its attempt to address the perceived weaknesses of both the content and development of the DSM-III and DSM-III-R, the DSM-IV development was based on commissioned reviews for sets of diagnoses, especially designed field trials and a transparent decision-making process.

**Conclusions:** Nosology is an ambitious and complicated enterprise. Classification systems must evolve as new empirical evidence is presented. Many of the obstacles faced by the DSM-IV work groups, therefore, may be moot for those formulating DSM-V. The DSM-IV development process was carefully planned, and the profession is indebted to the American Psychiatric Association (APA) for its initiative and effort in producing this system.

(Can J Psychiatry 1996;41:325–329)

**Key Words:** diagnostic classification, nosology, child

In 1989, I received a call from Allen Frances, chairman of the APA's DSM-IV Task Force, to see whether I would be interested in cochairing the Infancy, Child, and Adolescent Disorder Working Party. He said it would be 5 years of hard work, but it would be interesting. I was flattered to have been asked and agreed readily. My cochair was Magda Campbell. The working party comprised an impressive group of experts: Susan J Bradley, Dennis P Cantwell, Gabrielle A Carlson, Donald Jay Cohen, Barry Garfinkel, Rachel Klein, Benjamin Lahey, Rolf Loeber, Jeffrey Newcorn, Rhea Paul, Judith HL Rapoport, Sir Michael Rutter, Fred Volkmar, and John S Werry.

My own qualifications were questionable. I had worked on an ICD-8/ICD-9 comparison study with Mike Rutter many years ago, and I had served as a consultant to the not particularly important elimination disorder section in both DSM-III and DSM-IV. More to the point, I had gone on record in a paper written with Mike Rutter shortly after the appearance of DSM-III (1) to criticize the DSM-III's pseudo-precision,

arbitrary approach to setting criterion thresholds, incompatibility with an accepted international system, and presentation of a manual that was vastly over-authoritative. I had clearly been forgiven! I believe I deserved to be; I had come to appreciate that publishing a classification system is not a static event. It has complicated sequelae. On the one hand, if the system is—like DSM—user-friendly, as well as broadly accepted, it will bind many uncritical practitioners to it, errors and all. On the other hand, the very act of describing a diagnosis in unambiguous and replicable terms will create and give direction to a corrective process. Investigators will emerge from the woods to test validity in many different ways. Someone has to take the lead and put a definition down on paper for this process to take place, and we should be grateful that over the years the APA, with the leadership of Bob Spitzer and then Allen Frances, has been daring enough to do this.

Allen Frances's charge to our working party had been framed by the criticisms that both DSM-III and DSM-III-R had received. Timing was the first issue. The APA had been criticized for introducing DSM-III-R too soon after the introduction of DSM-III (“we were just learning how to use DSM-III,” “it made our research obsolete overnight,” “it was done for commercial reasons only, a case of planned obsolescence”), and now DSM-IV would follow after an interval of only 7 years. This was unavoidable because, by prior agreement, revisions of the DSM had to be synchronized with changes in the International Classification of Diseases (ICD),

---

Manuscript received June 1996.

<sup>1</sup>Irving Philips Professor of Child Psychiatry, Columbia University, College of Physicians and Surgeons, New York State Psychiatric Institute, New York.

Address reprint requests to: Dr David Shaffer, New York State Psychiatric Institute, 722 West 168th Street, New York, NY 10032

Can J Psychiatry, Vol 41, August 1996

and in 1994 the tenth version of the ICD was to appear. To anticipate criticism that DSM was an unstable system, Frances instructed us to keep differences between DSM-III-R and DSM-IV to a minimum. We should make only those changes that would correct some generally recognized mistake or could be justified by sound empirical evidence.

Second, there had been a perception that the creation of DSM-III and DSM-III-R had taken place behind closed doors. The process for DSM-IV was to be impressively transparent. A series of critical reviews were commissioned for each set of diagnoses to determine whether there was any new empirical evidence that would warrant changing diagnostic descriptions and definitions. These reviews were to be published openly in a sourcebook (2,3) so that the basis for change would be apparent to all. Feasible proposals would be published in an options book (4) that would be widely circulated. The options book would clarify the costs and benefits of proposed changes, and it was hoped that this would stimulate comment and criticism. Finally (and most importantly) new data from specially designed field trials and from secondary analyses of existing data sets would be obtained and published.

Third, it was widely believed that many of the decisions about DSM-III and DSM-III-R had been made by a small editorial group that had selectively taken advice from some people in the field but not from others. The DSM-IV working groups would have considerable autonomy and would be drawn broadly from among experts and practitioners. The working groups were expected to promote discussion of controversial issues, to solicit suggestions actively from colleagues who were not members of the working group, and to enrol them as consultants or advisors. When this process had been completed, the working groups would present a set of proposals to the APA DSM-IV Task Force for final approval.

Fourth, DSM-III and DSM-III-R had been criticized for being *too* precise: syndrome definitions went beyond available knowledge, and the insistence on exact behaviour frequencies or number of criteria was deemed “pseudoscientific.” In DSM-IV, thresholds, frequencies, and durations would, where possible, be set on the basis of empirical data, and where data were not available, it would be just fine to use less precise terms, such as “often” and “frequent.”

Fifth, although DSM-III and DSM-III-R were in use throughout the world, there was regret at their incompatibility with the parallel, but less fully described, ICD system. As both systems were working toward major revisions, the moment seemed right for increasing compatibility. ICD-10 had adopted many of the DSM conventions and would produce a criterion-based version for research (5) with a full glossary. Liaison groups were established, and within the framework of our respective conventions, we were to do our best to bridge differences.

Although we tried to work within these guidelines, things did not always go as intended. We did a superb job in making changes transparent; the options book clearly presented

different solutions to the problems we had identified. Even though the book was widely circulated and the choices were presented for debate at public meetings of the American Academy of Child Psychiatry, the American Psychiatric Association, the Society of Professors, and the Society for Research in Child Psychopathology, among others, the working party received very few comments from outsiders. Most of the debate would take place within our own group.

We also did a good job in establishing compatibility with ICD-10. This was in large measure due to the involvement and flexibility of Mike Rutter, who was directing the revision of the ICD-10 child section. There was give and take and a good deal of patience shown on both sides, and in the end we achieved a far greater measure of compatibility between the systems than any other diagnostic section. In many instances diagnoses appear with identical criteria. One of the most conspicuous remaining differences is the retention of “combination diagnoses” in the ICD system to designate commonly comorbid conditions, for example, conduct disorder with depression. This is not compatible with the DSM principle of coding all diagnoses separately. The advantage of the ICD-10 system is that it highlights a common pattern of comorbidity that might otherwise go uncoded. The DSM requirement that all diagnoses present be listed, however, has resulted in our learning about the extent of comorbidity. This would not have been possible had we followed the original ICD-9 guidelines of assigning only one principal diagnosis or highlighting some but not all patterns of comorbidity.

It was most difficult to adhere to the instruction to be conservative and to make only those changes that were supported by sound empirical evidence. There were many instances where criteria were changed, not because they were suggested by data, but because in the absence of any empirical evidence, a new approach seemed more sensible or acceptable, or conformed better with a basic DSM principle than had the DSM-III-R version. There simply are not enough nosologically helpful studies to substantiate all parts of a classification system that has been built on haphazard clinical anecdotes and on observations made over many generations.

The absence of pertinent information was apparent from the quality of the commissioned literature reviews. Some reviews were forced, through the paucity of material, to include both good and bad studies; as a result, many were difficult to interpret or provided very little information that addressed diagnostic or criterion issues. Overall, the reviews did very little to influence the process. By contrast, the field trials and secondary data analyses were enormously influential. The multisite child and adolescent field trials supported by the National Institute of Mental Health (NIMH) focused on the disruptive behaviour disorders (6,7) and autism (8), while the John D and Catherine T McArthur Foundation supported secondary analyses of existing data sets (9). These collaborations give a measure of the resources and great energy that were being poured into this project.

The field trials examined patients attending a variety of inpatient and outpatient units chosen as likely to include subjects with the diagnosis that was being studied (the disruptive disorder trial also included 26 adolescents residing in a detention facility). In these trials, patients were assessed with forms that addressed past, current, and proposed criteria. The patients were also rated on a variety of measures of impairment and were assigned diagnoses by clinicians who had access to systematic information about their symptoms. Algorithms were drawn up and new and old diagnostic formats were compared with each other and with the clinical diagnoses.

This design was adequate for assessing the overlap between what the clinicians believed to be the diagnosis and both the old and new DSM formulations. It indicated the extent to which the new system would reflect actual and current practice. The design also gave information about how the criteria might be adjusted, for example, by changing the symptom mix or the number of criteria required for a diagnosis, to make the best and most economical match with the clinical standard (10). Nevertheless, the use of established clinical standards as a criterion is of questionable value for a process that is in the business of changing those standards. Further, the reliability of clinical diagnoses is relatively poor, and this also mitigates their value as a criterion (11). The field trials were, by design, unable to indicate what proportion of nontroubled, nonreferred children would also meet diagnostic criteria. This is not a trivial issue because in community-based studies of unreferred subjects, Shaffer and others (12) found that there are many children and teenagers who meet DSM-III-R criteria for diagnoses but who do not appear to have significant social or academic impairment and who have never been referred for any clinical service.

The field trials' cross-sectional design, as well as their use of referred convenience (and therefore potentially unrepresentative) samples, also limited their ability to inform on discriminant validity, that is, on whether related diagnoses differ from one another not only in terms of their defining criteria but also in terms of antecedent risk factors or natural history. For example, it is not clear whether oppositional defiant disorder and conduct disorder or Asperger's disorder and autism are distinct entities or are instead simply mild and severe forms of the same condition. Some of these issues were dealt with by the secondary analyses of community samples of high-risk populations (9,13), but others will have to await findings from epidemiologically based research, such as the current NIMH Use, Need, Outcomes, and Costs in Child and Adolescent Populations (UNOCAP) study (14).

The delegation of responsibility to the working groups functioned well. Each meeting was attended by either Allen Frances, Michael First, the editor of DSM-IV, or Harold Pincus, vice chair of the task force, or most often by all 3. They also participated in each conference call. They advised the working party of conventions, rules, and principles, but made no attempt to influence issues of content. As a result, the working party developed a strong identification with the

revision. When, at the last moment, under pressure from the printers demanding their copy, Allen Frances and Michael First made a few stylistic changes without consulting the work group, the indignation was intense.

The working party met face to face at each of the APA and the American Academy of Child Psychiatry meetings and sometimes at other points during the year. At these meetings, a forum to present the latest proposals or findings from the secondary data analyses or field trials was also convened. The forums were always well attended, with between 80 and 200 people, but there were relatively few comments or suggestions. We did, however, receive well-formulated requests from child psychiatrists interested in the dissociative disorders who wanted to see a new diagnosis of dissociative disorder of childhood. In the end, this was not approved because the proposed criteria would overlap with those for other disorders and because a number of the proposed criteria would require a level of inference that went beyond that usually accepted in DSM. There were also representations to propose the development of a new diagnosis of adjustment disorder with suicidal features. It was pointed out that although many teenagers who make a suicide attempt do not meet any criteria for any Axis I diagnosis and that many make their suicide attempt after experiencing an external stress, they usually receive treatment of a type and intensity that is reserved for those with a psychiatric diagnosis. The proposal was eventually rejected because it was felt that it would discourage clinicians from ascertaining other psychiatric conditions that can lead to suicidal behaviour. The DSM-IV Task Force was generally very resistant to approving new diagnoses unless they were well supported by research findings.

The tasks of the working group changed over time. At the beginning, different groups made tentative proposals for changes in the criteria. After a while, the literature reviews that had been commissioned early in the process began to arrive. As our meetings continued, the issues sharpened, and certain topics took up increasing amounts of time. The issues that seemed to occupy the group most were 1) the structural relationship between oppositional and conduct disorder and 2) the different variants of attention deficit hyperactivity disorder (ADHD). These were diagnoses in which we all had some competence, and they were also the issues most driven by data from the reanalyses of 2 important studies (Loeber [9,13] and Lahey [15]) and the disruptive disorders field trials.

We eventually approved a fairly large number of changes: 2 groups of conditions (the gender identity and eating disorders) were moved out of the child section when the section name was altered from "Disorders that Are Usually First Evident in Infancy, Childhood, or Adolescence" to "Disorders Usually First Diagnosed in Infancy, Childhood, or Adolescence;" 3 diagnoses were deleted, all from the anxiety disorder group (avoidant disorder, overanxious disorder of childhood, and identity disorder); 2 diagnoses were added, both to the pervasive developmental disorder group (Rett's disorder and Asperger's disorder); new subtypes were devel-

oped for the disruptive disorders (ADHD and conduct disorder) and for reactive attachment disorder; one deleted DSM-III diagnosis was restored (attention deficit disorder without hyperactivity); the diagnostic criteria for autistic disorder were substantially revised; a number of small changes were made to individual criteria; and a small number of criteria were added or deleted from several diagnoses.

The reason for deleting avoidant disorder and overanxious disorder of childhood was that their criteria were very similar to those for social phobia and generalized anxiety disorder. Even though it was not clear that these disorders were the same in children as in adults, the general principle of not using the age of a patient to designate a diagnosis was important. The diagnosis of identity disorder was dropped because of the absence of any empirical evidence to show that it existed independently of a mood or anxiety disorder. The working party examined case records of adolescents who had been assigned this diagnosis by members of the Society for Adolescent Psychiatry and found that every one of these cases also met criteria for a mood or anxiety disorder. Furthermore, a literature search revealed no publications on the diagnosis for the previous 8 years. These generally negative findings, coupled with its reliance on inferential criteria, led us to recommend deletion. In the 2 years since the DSM-IV has been released, we have not received a single complaint about this decision.

There were some misgivings about each of the 2 new diagnoses. Rett's disorder appears to be a genetically determined neurological condition that is associated with mental retardation and language delay (16). Should it be classified as a mental or a neurological disorder? Although there was a good case for classifying it as a neurological disorder, ICD includes it as a mental disorder, and we had no authority to create a new neurological diagnosis. The other new diagnosis, Asperger's disorder, first described over 50 years ago, had been well described in a number of publications that had appeared since DSM-III-R. It was not clear from the literature, however, whether Asperger's was a distinct disorder or a mild variant of autism.

The diagnosis that had been brought back from DSM-III, attention deficit disorder predominantly inattentive type (called attention deficit disorder without hyperactivity in DSM-III), was a good example of the heuristic role of DSM. The diagnosis was first created for DSM-III because of a general—but not well-documented—belief among clinicians that, while many hyperactive children become less hyperactive as they grow older, they remain inattentive. In addition, there are other children who have great difficulty paying attention to specified tasks but who had never been hyperactive. The publication of criteria for this diagnosis then led to systematic studies (17), which confirmed that a number of children do meet criteria for attention deficit disorder without hyperactivity and that they are different in a number of ways from those that meet criteria for ADHD. When the diagnosis was deleted in DSM-III-R—with no reason for the deletion

having been published—there was an immediate call for its restoration.

The diagnosis of autism was also returned to an earlier format. The DSM-III-R diagnosis of autistic disorder was substantially different from the infantile autism diagnosis in DSM-III. The DSM-III-R format allowed one to choose the necessary criteria from a large selection without any requirement that there be symptoms from such cardinal areas of abnormality as language abnormality, impairment of social relationships, need for sameness, and bizarre behaviours, as was the case in DSM-III. As a result, a very heterogeneous group of children appeared to qualify for the diagnosis. Practical experience with the DSM-III-R version led to discontent about this among clinicians, and the basis for this discontent was confirmed by the autism field trial (8). Data from the trial played an important role in restoring its form to one that was closer to DSM-III.

Although avoidant and overanxious disorders have now been integrated with social phobia and generalized anxiety disorders, a number of the adult criteria have been changed. Children may not recognize that a source of anxiety is excessive or irrational (although this may also be true for adults); that criterion, therefore, is not required of children in any of the anxiety disorders, including obsessive-compulsive disorder. Another problem with the anxiety disorders is the lack of specificity of symptoms. Many anxiety symptoms are compatible with more than one disorder; for example, panic attacks may occur in the context of separation anxiety. This cross-compatibility could lead to a specious comorbidity among anxiety disorders. To minimize this problem, we reworked the diagnostic rules to emphasize that double counting should be avoided.

There were also important issues that were not fully discussed, including how best to deal with the problem that symptoms may be present without any features of impairment. The solution provided in DSM-IV is that all diagnoses must be associated with significant social or other impairment. This is a potentially problematic decision, without parallel in the rest of medicine, where asymptomatic conditions are widely accepted.

Another issue that was not fully worked through is what to make of the very high rate of diagnostic comorbidity that is found among children and adolescents. Is this a consequence of shared genetic or environmental exposures, or does it arise because one disorder leads sequentially to another? The challenge to nosology is to make sure that comorbidity is not spurious, that it does not result because we use similar criteria for several diagnoses (as in the case of irritability, a criterion shared by mania and depression) or because criteria that seem on paper to be distinct are mistaken for one another by a parent replying to a structured interview or a clinician with new or moderate skills (for example, the similarity between the concept of *irritability* in mania and such criteria for ADHD as *blurting out answers before the question has been completed* or *interrupting or intruding on others* or

*difficulty waiting turn or often avoids dislikes or is reluctant to engage in tasks that require sustained mental effort*). We also need to be sure that we are not drawing the boundaries of disorders in the wrong places. If a given diagnosis, as it occurs in nature, nearly always coexists with another, maybe the true disorder incorporates features of both. Boundary problems of this kind almost certainly exist for the anxiety and pervasive developmental disorders, but we lacked the data to address and profitably resolve those problems.

Paradoxically, we paid most attention to the subjects we already knew most about. I believe this was a function of the availability or nonavailability of data. When data were present, issues were depersonalized, and there was open and candid discussion. The lively and critical discussions around the disruptive disorders were a contrast to the often overly respectful and undercritical mood that took over when a working party member had to defend a position for which there were few data.

It is likely that the group that will work on DSM-V will have a far easier time. The growth of biological research may lead to more robust markers of a diagnosis than symptoms have proved to be. The development of criterion-based structured interviews, such as the Diagnostic Interview Schedule for Children (DISC), used in large, community-based longitudinal studies, will generate a wealth of data that will be used to examine key issues of criteria and construct validity. It should be a breeze!

#### Acknowledgements

This work was made possible by NIMH Research Training Grants MH 38198-11A2 and MH 16434 as well as NIMH Center Grant MH 4 3878 AI.

#### References

- Rutter M, Shaffer D. DSM-III: a step forward or back in terms of the classification of child psychiatric disorders? *J Am Acad Child Psychiatry* 1980;19:371-94.
- American Psychiatric Association. DSM-IV sourcebook. Volume 1. Washington (DC): American Psychiatric Association; 1994.
- American Psychiatric Association. DSM-IV sourcebook. Volume 2. Washington (DC): American Psychiatric Association; 1996.
- American Psychiatric Association. DSM-IV options book. Washington (DC): American Psychiatric Association; 1991.
- World Health Organization. The ICD-10 classification of mental and behavioral disorders: diagnostic criteria for research. Geneva: World Health Organization; 1993.
- Lahey BB, Applegate B, Barkley RA, Garfinkel B, McBurnett K, Kerdyk L, and others. DSM-IV field trials for oppositional defiant disorder and conduct disorder in children and adolescents. *Am J Psychiatry* 1994;151:1163-71.
- Lahey BB, Applegate B, McBurnett K, Biederman K, Greenhill L, Hynd GW, and others. DSM-IV field trials for attention deficit/hyperactivity disorder in children and adolescents. *Am J Psychiatry* 1994;151:1673-85.
- Volkmar FR, Klin A, Siegel B, Szatmari P, Lord C, Campbell M, and others. Field trial for autistic disorder in DSM-IV. *Am J Psychiatry* 1994;151:1361-7.
- Loeber R, Green SM, Keenan K, Lahey BB. Which boys will fare worse? early predictors of the onset of conduct disorder in a six-year longitudinal study. *J Am Acad Child Adolesc Psychiatry* 1995;34:499-509.
- Frick PJ, Lahey BB, Applegate B, Kerdyck L, Ollendick T, Hynd GW, and others. DSM-IV field trials for the disruptive behavior disorders: symptom utility estimates. *J Am Acad Child Adolesc Psychiatry* 1994;33:529-39.
- Piacentini J, Shaffer D, Fisher PF, Schwab-Stone M, Davies M, Gioia P. The diagnostic interview schedule for children: revised version III criterion validity. *J Am Acad Child Adolesc Psychiatry* 1993;32:658-65.
- Shaffer D, Fisher P, Dulcan M, Davies M, Piacentini J, Schwab-Stone M, and others. The NIMH diagnostic interview schedule for children (DISC-2): description, acceptability, prevalences, and performance in the MECA study. *J Am Acad Child Adolesc Psychiatry* 1996;35:865-77.
- Loeber R, Keenan K, Lahey BB, Green SM, Thomas C. Evidence for developmentally based diagnoses of oppositional defiant disorder and conduct disorder. *J Abnorm Child Psychol* 1993;21:377-410.
- National Institute of Mental Health. Request for applications MH-94-009: cooperative agreement for a multi-site study of mental health service use, need, outcomes, and costs in child and adolescent populations. 1994.
- Lahey BB, Loeber R, Hart EL, Frick PJ, Applegate P, Zhang Q, and others. Four-year longitudinal study of conduct disorder in boys: patterns and predictors of persistence. *J Abnorm Psychol* 1995;104:83-93.
- Armstrong DD. The neuropathology of Rett syndrome: overview. *Neuropediatrics* 1995;26:100-4.
- Lahey BB, Schaughency EA, Hynd GW, Carlson CL, Nieves N. Attention deficit disorder with and without hyperactivity: comparison of behavioral characteristics of clinic-referred children. *J Am Acad Child Adolesc Psychiatry* 1987;26:718-23.

#### Résumé

**Objectif :** Donner un aperçu du processus de la production du DSM-IV du point de vue d'un participant.

**Méthode :** Présentation d'une revue narrative.

**Résultats :** Dans sa tentative à faire face aux faiblesses perçues du DSM-III et du DSM-III-R, tant au niveau de leur contenu que dans leur préparation, la production du DSM-IV fut basée sur des évaluations recrutées de séries de diagnostics, particulièrement des essais conçus sur place et selon un processus apparent de prise de décision.

**Conclusions :** La nosologie étant une étude ambitieuse et complexe, les systèmes de classification doivent évoluer dans la mesure où l'évidence empirique est présentée. Plusieurs des difficultés auxquelles l'équipe responsable du développement du DSM-IV a dû s'attaquer peuvent donc être sans rapport pour ceux qui préparent le DSM-V. Le processus de la préparation du DSM-IV a été soigneusement planifié et la profession est redevable à l'American Psychiatric Association (APA) pour son initiative et son effort dans la réalisation de ce système.