Clinical Utility and DSM–V

Stephanie N. Mullins-Sweatt
National Crime Victims Research and Treatment Center,
Medical University of South Carolina, and University of Kentucky

Thomas A. Widiger
University of Kentucky

The construction of the American Psychiatric Association’s diagnostic manual has been guided primarily by concerns of construct validity rather than of clinical utility, despite claims by its authors that the highest priority has in fact been clinical utility. The purpose of this article was to further articulate the concept and importance of utility when constructing and evaluating a diagnostic construct. It is suggested that a relative emphasis on validity over utility is justifiable but that matters of clinical utility should not be neglected. Discussed in particular is ease of usage, communication, and treatment planning. Suggestions for future research are provided.

Keywords: clinical utility, diagnostic construct, DSM–V

“The Cronbach and Meehl paper, ‘Construct validity in psychological tests,’ laid the groundwork for fifty years of work on validity” (Shavelson, 2003, p. 381). There is little that would or should dissuade that conclusion. More specific to the purpose of the present article, Meehl also published another seminal article in the same year within the same journal (Psychological Bulletin). This second article was concerned with matters beyond simply construct validity. Meehl and Rosen (1955) considered “the evaluation of the efficiency of psychometric devices as well for the improvement in the interpretations made from such devices” (p. 194); more specifically, the cost–benefit ratio of an assessment decision. Their article was concerned primarily with the importance of considering base rates, but it also considered the relative cost-to-benefit ratio of any particular diagnostic cutoff score. Rorer (a graduate student of Meehl’s) and his colleagues subsequently provided a further articulation of the importance of utility when evaluating a diagnostic decision rule (Rorer, Hoffman, & Hsieh, 1966).

The central and fundamental importance of clinical utility in the construction of a diagnostic manual was expressed explicitly by the chief architects of the 4th edition of the American Psychiatric Association’s (APA) Diagnostic and Statistical Manual of Mental Disorders (DSM–IV; APA, 1994): “There is unanimous agreement, even among those engaged in research, that the primary purpose of DSM-IV is to facilitate clinical practice and communication” (Frances et al., 1991, p. 410). This emphasis is again stated quite forcefully in the first paragraph of the introduction to the DSM: “Our highest priority has been to provide a helpful guide to clinical practice” (APA, 2000, p. xxiii). However, matters of clinical utility do not in fact appear to be the highest priority in the development and consideration of the DSM. Suggested in the present article is that matters of the validity of the DSM should generally take precedence over matters of utility, but matters of utility should not be neglected.

The Concept of Clinical Utility

Clinical utility itself is considered by most to be a construct; any meaningful discussion of a construct requires its delimitation from other constructs. (“The investigation of a test’s construct validity is not essentially different from the general

Editor’s Note. Howard N. Garb served as the Action Editor for this manuscript.—MES

Stephanie M. Mullins-Sweatt, National Crime Victims Research and Treatment Center, Department of Psychiatry and Behavioral Science, Medical University of South Carolina, Department of Psychology, University of Kentucky; Thomas A. Widiger, Department of Psychology, University of Kentucky.

Thomas A. Widiger, is the lead author of the Personality Disorder Interview-IV, the manual for which is cited in this article (i.e., Widiger, Mangine, Corbitt, Ellis, & Thomas, 1994).

We express our appreciation to Geoffrey Reed for his helpful comments on an earlier version of this manuscript.

Correspondence concerning this article should be addressed to Stephanie N. Mullins-Sweatt, Medical University of South Carolina, National Crime Victims Research and Treatment Center, 165 Cannon Street, MSC 852; Charleston, SC 29425. E-mail: stephanie.sweatt@gmail.com
scientific procedures for developing and confirming theories” (Cronbach & Meehl, 1955, p. 300). Construct validation is a matter of establishing a construct’s verisimilitude (Meehl, 1978), or its truth-value: “that one’s measure of a given construct relates to measures of other constructs in theoretically predictable ways” (Smith, 2005, p. 396). Quite a few articles have been written on the optimal procedures for construct validation. Most recently, Smith (2005) suggested that there are “at least five steps in construct validity work: careful theory specification, development of informative hypothesis tests, use of sound research design, examination of the degree to which observations confirm hypotheses, and ongoing revisions of both theory and measures” (p. 406).

The authoritative statement on construct validity for psychiatric diagnoses was provided by E. Robins and Guze (1970), whose article has as much renown within psychiatry (Hyman, 2002; Kendler, 1990; Kupfer, First, & Regier, 2002; L. N. Robins & Barrett, 1989) as Cronbach and Meehl’s (1955) has within psychology. E. Robins and Guze (1970) set forth five phases for “the development of a valid classification in psychiatry” (p. 983): precise clinical description, laboratory studies, delimitation from other disorders (i.e., the disorder should be as homogeneous as possible), follow-up study (i.e., upon follow-up, one might discover that the original patients are suffering from a different disorder that could account for the original clinical picture), and family study (i.e., does the disorder in fact run in families, whether due to genetic or environmental causes).

First, Pincus, et al. (2004) have provided a comparably useful definition of clinical utility, suggesting that “clinical utility is the extent to which DSM assists clinical decision makers in fulfilling the various clinical functions of a psychiatric classification” (p. 947). They further identified five components of clinical utility: conceptualization; communicating information to patients, families, health care administrators, and other relevant parties; use of the DSM in clinical practice; treatment selection; and predicting future clinical management needs.

We would respectfully suggest that though the conceptualization of the disorder and predicting future management needs may in fact be useful in clinical practice, these components are better understood as matters of construct validity rather than of clinical utility. The conceptualization of a disorder is comparable to the theoretical articulation of a construct (Cronbach & Meehl, 1955; Smith, 2005) and its delimitation from other disorders (E. Robins & Guze, 1970). Similarly, predicting a future course concerns the validation of theoretically derived hypotheses concerning the construct (Cronbach & Meehl, 1955; Smith, 2005) through follow-up studies (E. Robins & Guze, 1970). We would agree, though, that the other three primary components are the central matters of clinical utility: ease of usage, communication, and treatment planning. We discuss each in turn, with a particular emphasis on the importance of distinguishing between validity and utility and the potential impact of the latter in the construction of the DSM.

We should first note, though, that validity and utility are not entirely distinct or independent constructs. It is difficult to imagine, for instance, that there would be much (if any) clinical utility for a diagnosis that lacks any validity. The utility of a diagnosis for treatment planning and communication will be supported and informed in part by its validity. However, as we demonstrate below, the concepts are not synonymous, and it will at times be important to appreciate their distinction, particularly when making decisions regarding the construction of the DSM.

Ease of Usage

Prior to the DSM (3rd ed.; DSM–III; APA, 1980), the reliability (and validity) of psychiatric diagnoses in clinical practice and research were notoriously poor (Spitzer & Fleiss, 1974). The major innovation of the DSM–III was the provision of specific and explicit criterion sets for all of the disorders within the DSM (the only exception being schizoaffective disorder). These criterion sets have substantially increased the ability of researchers and clinicians to obtain more reliable and thereby (hopefully) more valid diagnoses (Spitzer, Williams, & Skodol, 1980).

Nevertheless, it is evident that clinicians often fail to adhere to the Diagnostic and Statistical Manual of Mental Disorders (4th ed., text revision; DSM–IV–TR; APA, 2000) criterion sets when providing clinical diagnoses (Garb, 2005). There are a number of possible reasons for this lack of adherence. These include disagreement with respect to a particular criterion set, disgruntlement with psychiatric diagnoses in general, skepticism as to the validity or utility of a particular diagnosis, and irresponsible neglect (Stengel, 1959). However, one important additional possibility is simply feasibility or ease of usage, a matter of clinical utility that was largely neglected in the construction of the DSM–III (APA, 1980) and the DSM–III–R (APA, 1987).

Ten explicit goals for the authors of the DSM were listed in the introduction to the DSM–III–R (e.g., avoidance of new terminology, theoretical neutrality, and reliability; APA, 1987). “Suitability for describing subjects in research studies” (APA, 1987, p. xx) was among the 10 listed goals, but suitability or ease of usage for describing patients in general clinical practice was not. The neglect of clinical utility is also suggested by the lack of feasibility or usage of the DSM–IV within certain clinical settings, such as primary care (Rounsaville et al., 2002). One explicit illustration of this problem is the quite lengthy criterion sets. For instance, DSM–III–R generalized anxiety disorder ballooned to 22 diagnostic criteria, and somatization disorder to 35. Criterion sets of this length would appear to be unrealistic for most practicing clinicians. Researchers have the time (and at times funding) to conduct assessments that last for hours, but clinicians do not. It was evident to the authors of the DSM–IV (APA, 1994) that the DSM was being written largely by researchers for the concerns and interests of researchers while the practical needs and concerns of clinicians were being largely neglected (Kendler, 1990). As expressed by the chief architects of the DSM–IV, “several categories in DSM–III–R are too long and complex for use in everyday practice” (Frances, Pincus, Widiger, Davis, & First, 1990, p. 1441). One of the primary mandates for the authors of the DSM–IV was to improve ease of usage (First, Frances, & Pincus, 2004; Frances, Widiger, & Pincus, 1989).

However, it is not at all clear whether the authors of the DSM–IV actually achieved the goal of constructing a DSM that was feasible for use in everyday clinical practice. The criterion sets for somatization disorder, generalized anxiety disorder, and antisocial personality disorder were shortened (First, Frances, & Pincus, 2004), and less extensive simplifications were also provided for the diagnoses of autistic disorder, conduct disorder, dementia, amnesic disorder, substance dependence, and schizophrenia (First, Pin-
Data were obtained to support the validity of the abbreviated criterion sets (Widiger et al., 1998), but no studies were conducted on the amount of time saved by the abbreviated versions, nor whether the revisions resulted in an increased adherence to the DSM (First, Pincus, et al., 2004). As suggested by Smith, McCarthy, and Anderson (2000), one of the “sins” of abbreviated measures is to “fail to show that your short form offers meaningful time or resource savings for the loss in validity” (p. 108). These types of empirical studies are critical in determining whether changes to the DSM are actually improving its clinical utility.

The development of the DSM–IV was informed by 37 data reanalyses funded in part by the MacArthur Foundation and 12 field trials funded in part by the National Institute of Mental Health (NIMH; First, Frances, & Pincus, 2004; Widiger et al., 1998). These empirical studies provided very useful and constructive information for the development of the DSM–IV, but, with only a few exceptions, they were clearly focused on matters of validity rather than clinical utility (see Widiger et al., 1998). In contrast, the development of the DSM–III included an extensive field trial within general clinical practice (Spitzer, Forman, & Nee, 1979). Naturalistic studies of practicing clinicians to consider their typical usage of the DSM or the effect of a proposed revision on clinical practice have not generally been conducted by the authors of the diagnostic manual since the DSM–III. Instead, the focus has been on the reliability and validity of the disorders when diagnosed within research settings using semistructured interviews. There have been quite a few studies of clinicians’ application of the diagnostic manual by other researchers, and these studies have typically suggested that clinicians are continuing to fail to adhere closely to the diagnostic criterion sets (Garb, 2005; Kutchins & Kirk, 1997). Prototypal Matching

Another approach to improving ease of usage beyond the abbreviation of criterion sets is to abandon criterion sets altogether in favor of a paragraph description of a prototypic case of a respective disorder (consisting of approximately 20 sentences), to which clinicians would match their impression of the patient on a 5-point Likert scale (First & Westen, 2007). There are data to suggest that clinicians prefer this prototypal matching approach over the DSM–IV–TR diagnostic criterion sets (Spitzer, First, Shedler, Westen, & Skodol, 2008), and it is evident that the prototypal matching approach is much easier to use and would save a considerable amount of time. For example, it presently takes approximately 2 hr to assess all of the DSM–IV–TR personality disorder diagnostic criteria (Widiger & Samuel, 2005). With the prototypal matching approach, proponents suggest that “clinicians could make a complete Axis II diagnosis in 1 or 2 minutes” (Westen, Shedler, & Bradley, 2006, p. 855).

Specific and explicit diagnostic criterion sets are perhaps used relatively infrequently in general clinical practice outside of the United States (as well as inside). The World Health Organization’s (WHO’s; 1992) International Classification of Diseases (ICD-10) includes two separate editions for psychiatry: one for researchers (with specific and explicit criterion sets) and one for clinicians (narrative paragraph descriptions), in the belief that clinicians will be unable to adhere to the specific and explicit criterion sets. This approach was considered for the DSM–IV but was rejected because it was difficult to endorse a policy for clinical diagnoses that was not acceptable for research diagnoses (First et al., 2004).

Prototypal matching would address ease of usage, but it also would essentially be a return to the idiosyncratic, unreliable, and invalid diagnostic practices of the DSM, 2nd edition (DSM–II; APA, 1968). “To make a diagnosis, diagnosticians rate the overall similarity or ‘match’ between a patient and the prototype using a 5-point rating scale, considering the prototype as a whole rather than counting individual symptoms” (Westen et al., 2006, p. 847). No explicit guidelines are provided as to which of the features included in the narrative description should be emphasized, or even necessarily used (Spitzer et al., 1980). Clinicians are likely to vary widely in how they interpret and apply the narrative descriptions, as they did for the DSM–II (Spitzer & Fleiss, 1974; Stengel, 1959).

“There is clearly a trade-off between assessment time and validity” (Smith et al., 2000, p. 108), and the validity of a diagnosis should not be sacrificed for the sake of ease of usage. No interrater reliability data have been published on the prototypal matching procedure. Agreement between different clinicians interviewing the same patient at different times was quite poor for the DSM–III–II prototypal matching diagnoses (Spitzer & Fleiss, 1974), and there is perhaps little reason to expect that it would be much better even when the narrative paragraphs consist of sentences that were selected on the basis of empirical data and are rated on a 5-point scale.

Westen et al. (2006) reported support for the validity of prototypal matching, but there have been two fundamental difficulties with the validity of these findings. First, the external validity data (e.g., DSM–IV diagnoses) were provided by the same persons who provided the prototypal ratings. In addition, these clinicians already knew the patients very well (average length of treatment was 16 months), and it is quite likely that they already had a clear understanding of the clinical diagnoses. What is needed are studies on the validity (and reliability) of assessments of patients being seen for the first time. This is the typical context for which the DSM has been constructed and is being used, in both clinical and research settings.

Structured Assessment Instruments

A trend in a direction largely opposite to increasing ease of usage is to require the administration of formal psychological assessment instruments (particularly structured or semistructured interviews) when obtaining mental disorder diagnoses (Widiger & Clark, 2000). As expressed by the vice chair of the DSM–V and colleagues, “Although diagnostic criteria are the framework for any clinical or epidemiological assessment, no assessment of clinical status is independent of the reliability and validity of the methods used to determine the presence of a diagnosis” (Regier et al., 1998, p. 114). The DSM–III (APA, 1980) innovation of providing relatively specific and explicit diagnostic criteria is not realized if clinicians do not in fact adhere to the criterion sets and assess them in a comprehensive, systematic, and consistent fashion (Rogers, 2003). Reliable assessments do appear to occur when clinicians and researchers assess criterion sets with structured and/or semistructured interviews, the preferred method within clinical research (Rogers, 2003; Zimmerman, 2003).
Researchers would be hard-pressed to get their findings published if they failed to document that their diagnoses were based on a systematic, replicable, and objective method of assessment, yet no such requirements are generally provided for routine clinical assessments. Clinicians prefer to rely on their own experience, expertise, and subjective impressions obtained through unstructured interviews (Westen, 1997), but it is precisely this reliance on subjective and idiosyncratic clinical interviewing that often undermines the reliability (and ultimately the validity) of clinical diagnoses (Garb, 2005; Rogers, 2003; Wood, Garb, Lilienfeld, & Nezworski, 2002).

A proposal for the DSM–V is to incorporate specified clinical assessments within diagnostic criterion sets to provide a formal requirement that a systematic, replicable method of assessment be conducted (Widiger & Clark, 2000). At first blush, this might appear to be a radical suggestion, but formal measures are already required components of the diagnostic criterion sets for learning disorders and mental retardation (APA, 2000, pp. 49, 53). “This exception points the way for research that could lead to incorporation of psychological tests results as diagnostic criteria for other mental disorders” (Rounsaville et al., 2002, p. 24). Few clinicians would attempt to diagnose mental retardation solely on the basis of a brief, unstructured clinical interview. Perhaps as much consideration should be provided for other mental disorders.

Incorporating formal measures within diagnostic criteria would decrease rather than increase ease of usage, and the authors of the DSM are reluctant to further burden the practicing clinician. For example, most sleep disorder specialists use the International Classification of Sleep Disorders (ICSD) developed by the American Sleep Disorders Association (2005). The ICSD requires the use of polysomnographic information for the diagnosis of a number of sleep disorders (e.g., time of onset of rapid-eye movement sleep). The 12 DSM–IV–TR sleep disorder diagnoses are coordinated with the ICSD but differ significantly in failing to include polysomnographic diagnostic criteria. Detailed references are made to polysomnographic findings within the text of the DSM–IV (and the DSM–IV–TR), and it was acknowledged by its authors that “for sleep disorders other than insomnia, such as narcolepsy and sleep apnea, the utility of sleep laboratory testing is widely accepted” (Buysse, Reynolds, & Kupfer, 1998, pp. 1104–1105). Nevertheless, polysomnography findings were not required in the DSM–IV because of the extensive cost of the technology and their lack of availability within many clinical settings (Buysse et al., 1998; First, Pincus, et al., 2004).

An additional issue facing any decision to include formal assessment measures within a diagnostic criterion set is what specific assessment instrument to require. One could leave this decision open to the clinician, but there is considerable variability in the reliability and validity of alternative measures (Garb, 2003). However, it might be rather difficult and premature to require (for instance) one particular semistructured interview in preference to any other specific interview schedule, given the lack of data to demonstrate conclusively the predominance of any one particular instrument and the presence of data to indicate variability in findings across alternative measures (Widiger & Lowe, in press). For the diagnosis of mental retardation, the DSM–IV–TR simply requires “an individually administered IQ test” (APA, 2000, p. 49), and for a reading disorder one is required to administer an “individually administered standardized test” (APA, 2000, p. 53). We would suggest that the requirement for other mental disorders specify the administration of a semistructured interview or a standardized self-report inventory.

It is evident that the requirement of a measure that objectively and systematically assesses a diagnostic criterion set would improve significantly the reliability and validity of clinical diagnoses (Rogers, 2003; Zimmerman, 2003; Zimmerman, Chelminski, McGlinchey, & Young, 2006). Rather than simplify the DSM to conform to how clinicians diagnose mental disorders in general clinical practice (First & Westen, 2007), it might be preferable to alter the practice of clinicians to conform to how diagnoses are being made reliably within systematic research (Widiger & Clark, 2000). In fact, it may not be the case that clinicians are adverse to conducting formal assessments with semistructured interviews simply because it is easier to provide diagnoses without them (Weiner, 2007). The neglect of systematic assessments can reflect a number of possible reasons, such as not believing that these test instruments will in fact be helpful for a diagnosis or a differential diagnosis. However, their neglect also may reflect other trends and forces, such as difficulty obtaining reimbursement for the time spent in psychological testing (Butcher, 2006). If objective, systematic, and comprehensive assessments became a required component of a diagnosis, then perhaps clinicians would be able to obtain adequate reimbursement for their time and effort, as in the case of mental retardation and other childhood disorders, and would then include, if not embrace, their implementation in general clinical practice (Weiner, 2007).

**Computer-Administered Assessment**

The limited amount of time that is generally available to many clinicians could effectively rule out the administration of semistructured interviews within many clinical settings, particularly if no reimbursements are provided for their administration. A substantial amount of time can be saved, however, by first using a self-report inventory as a screening device, followed by a semistructured interview that is confined to just one or two diagnoses (Widiger & Samuel, 2005). Self-report inventories can also be useful in alerting clinicians to areas of dysfunction that the clinician might otherwise have missed.

Self-report inventories, however, can also be expensive in time or money to score. An additional alternative is the use of onsite computerized interviews or test administration (Garb, 2007). A computerized interview or test can be completed while a patient is waiting for an appointment, the results can be more revealing than a face-to-face interview (to the extent that respondents are more likely to disclose sensitive information to computers than to human interviewers), and the results can be directly entered into a scoring system, the results of which will be immediately available to the clinician. Such an assessment approach could also be useful for family physicians who may rely on their memory and limited understanding of the diagnostic criteria when assessing for mental disorders for prescribing psychotropic medications.

**Diagnostic Efficiency**

Computerized-administered assessments, however, may also be prohibitively expensive within some clinical settings. The WHO (1992) provides narrative paragraph descriptions rather than spe-
specific and explicit criterion sets within the ICD-10 in part because of the limited resources within many countries and clinical settings to systematically assess diagnostic criterion sets (First & Westen, 2007). Computerized assessments could be comparably infeasible within such settings.

If clinicians are focusing on just a subset of features or symptoms when deciding on a clinical diagnosis (rather than systematically assessing all of the diagnostic criteria or matching the patient to an entire global prototype), then it might be useful to provide them a rank order of the diagnostic efficiency of a respective criterion set. Most of the criterion sets for the DSM–IV personality disorders are presented in a descending order of diagnostic value (Widiger, Mangine, Corbitt, Ellis, & Thomas, 1994). This fact was not noted within the manual in part because there were a few notable exceptions (e.g., new diagnostic criteria were placed at the end of the list due to the absence of sufficient data for their ranking). It is evident that one can readily identify more efficient usage of existing diagnostic criterion sets (Chorpita & Nakamura, 2008; Frick et al., 1994; Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984). Not all of the diagnostic criteria need to be assessed, and some diagnostic criteria are considerably more informative than others. If clinicians do not in fact have the time or means to consider all of the diagnostic criteria, then the authors of the criterion sets should at least provide clinicians with empirically validated information as to which diagnostic criteria are most informative.

Facilitating Communication

The primary purpose of an official diagnostic nomenclature is to provide a common language of communication (Kendell, 1975). The impetus for the development of the DSM–I was the crippling confusion generated by the absence of an authoritative, common language. Medical centers, clinics, and clinicians were not using the same diagnoses, thereby hindering substantially meaningful communication and consistency in care (Widiger, 2008). “For a long time confusion reigned. Every self-respecting alienist, and certainly every professor, had his own classification” (Kendell, 1975, p. 87).

A difficulty for the authors of the DSM–V is the diversity of interested parties who communicate in terms of the DSM. The DSM–IV–TR is the common language for communication among a variety of agencies, organizations, companies, and professions. As expressed in the introduction to the diagnostic manual, “an official nomenclature must be applicable in a wide diversity of contexts” (APA, 2000, p. xxiii). Beyond simply treatment decisions, the diagnostic categories of the DSM–IV–TR impact decisions concerning disability, incarceration, public health care policy, and insurance coverage. The chief architects of the DSM–IV asked, rhetorically, “How does one balance the needs of the many users of DSM–IV – for example, clinicians, researchers, trainees and students, administrators, record room librarians and ... insurance companies, lawyers, disability boards?” (Frances et al., 1991, p. 410). It is perhaps unlikely that a uniform diagnostic system could in fact be optimal, or even compatible, for all of these different applications. The APA, however, is surprisingly sanguine about the DSM’s utility across different clinical and social contexts. With one notable exception, it is stated that, “fortunately, all these many uses are compatible with one another” (APA, 2000, p. xxiii).

The one exception noted within the DSM–IV–TR is the utility of the DSM for forensic decisions. Here, it is explicitly acknowledged that there is an “imperfect fit between the questions of ultimate concern to the law and the information contained in a clinical diagnosis” (APA, 2000, p. xxxiii). We would suggest, though, that a similar concern applies to many other social and clinical decisions.

One of the recurring arguments against shifting the APA DSM to a dimensional classification is the ease of communication for diagnostic categories (First, 2005). As expressed by the chair of the DSM–IV, “there is an economy of communication and vividness of description in a categorical name that may be lost in a dimensional profile” (Frances, 1993, p. 110). However, it does not appear to be the case, as many clinicians well know, that a DSM–IV–TR mental disorder diagnosis provides sufficient information for many social and clinical decisions. The simplicity of being able to use the same diagnostic threshold for all social and clinical decisions might be offset by the inconsistency in the needs and concerns of these different decisions.

Kraemer, Noda, and O’Hara (2004) argued that “a categorical diagnosis is necessary” (p. 21) for psychiatry. “Clinicians who must decide whether to treat or not treat a patient, to hospitalize or not, to treat a patient with a drug or with psychotherapy, or what type, must inevitably use a categorical approach to diagnosis” (Kraemer et al., 2004, p. 12). Although seemingly compelling, this is not an accurate characterization of actual clinical practice. In many of these common clinical situations, the decision is not in fact clearly categorical in nature. Typically, there is a decision concerning a degree of medication dosage, frequency of therapy sessions, or even degree of hospitalization (e.g., day hospital, partial hospitalization, residential program, and traditional hospitalization). Proponents of the “hybrid” approach of diagnosis (wherein the categories are retained but scaled in terms of severity or number of criteria met) suggest that these decisions can best be made by creating dimensional profiles of the existing categories (Oldham & Skodol, 2000). However, this approach still falls short in its recognition of subthreshold cases and may be as time-consuming and complex as the present categorical system (Verheul, 2006).

Even more important, it is evident that these many different clinical decisions will not be well informed by a uniform diagnostic threshold. The present diagnostic thresholds are not set at a point that is optimal for any particular social or clinical decision, and yet are used to inform most to all of them (Regier & Narrow, 2002). A diagnostic system with the flexibility to set different thresholds for different social and clinical decisions could be considerably more useful and credible than the present system. A flexible (dimensional) classification could be preferable to governmental, social, and professional agencies because it would provide more reliable, valid, and explicitly defined bases for making important social and clinical decisions. It is in part for this reason that the authors of the DSM–V are considering the provision within the DSM of a variety of supplementary dimensional scales of functioning to facilitate clinical decisions (e.g., Shear, Bjelland, Beesdo, Gloster, & Wittchen, 2008).

Public Health Care

One area of particular pragmatic concern is the need for treatment and public health care policy. Regier (vice chair of the DSM–V) and colleagues (1998) lamented the high prevalence rates
obtained for many mental disorders within NIMH epidemiology studies. “In the current US climate of determining the medical necessity for care in managed health care plans, it is doubtful that 28% or 29% of the population would be judged to need mental health treatment” (Regier et al., 1998, p. 114). Regier et al. noted that “major policy questions were raised about the need for mental health services that were implied by these high rates . . . such . . . as indicating a bottomless pit of possible demand for mental health services” (Regier et al., 1998, p. 110). Regier et al. suggested that the diagnostic thresholds for many disorders in the DSM–IV–TR should be raised to be more consistent with public health care policy and funding. “Hence, additional impairment and other criteria should be developed for future epidemiological surveys” (Regier et al., 1998, p. 114).

The authors of the DSM–IV added a requirement for a clinically significant level of impairment to many of the diagnostic criterion sets, which helped address this concern (Regier & Narrow, 2002). “It would appear that the lowest or most conservative [prevalence] rates are obtained by DSM-IV, followed by DSM-III, with DSM-III-R having the highest rates” (Regier & Narrow, 2002, p. 26). However, as Regier and Narrow argued, the concern still remains. “Despite the reassurance that clinically significant disorders may be closer approximations of need for care, diagnosis by itself is not a sufficient indicator of treatment need” (p. 27), and available evidence indicates that many people who meet diagnostic criteria for disorders have neither functional impairment nor a perceived need for treatment. Hence, some adjustments in the criteria—particularly for those that would operationalize clinical significance seem appropriate. (Regier & Narrow, 2002, p. 28)

Adjusting the threshold for diagnosis on the basis of level of impairment, however, has received considerable criticism (e.g., Spitzer & Wakefield, 1999; Wakefield & First, 2002). Wakefield and Spitzer (2002) indicated that service utilization (or treatment need) is a very problematic indicator for the presence of a disorder. Wakefield and colleagues provide numerous examples of criterion sets from the DSM–IV–TR that may fail to distinguish between problems in living and true psychopathology due to the reliance on indicators of distress, duration, or impairment rather than an underlying pathology (Wakefield, 1997, 2007; Wakefield & First, 2002; Wakefield, Schmitz, First, & Horwitz, 2007).

Regarding the presence of an underlying pathology may also be highly problematic as a basis for establishing the presence of disorder (Widiger, 2008). However, more to the point, Regier and colleagues appear to be confusing a matter of clinical utility with a matter of validity. They purport that raising diagnostic thresholds on the basis of level of impairment would help provide “accurate reflections of the level of true psychopathology in the population” (Regier & Narrow, 2002, p. 28), but their concerns are really more with respect to “policy implications” (p. 28). Public health care policy is of considerable importance to the mental health field, but it need not be consonant with the actual rate of psychopathology within the population.

One solution might be to explicitly separate utility and validity concerns within the DSM. The optimal diagnostic threshold for the existence of a disorder and need for treatment (i.e., if the latter informs public health care policy) are unlikely to be compatible. As suggested by Wakefield and Spitzer (2002), “disability should be considered an independent dimension” (p. 39). Not all physical disorders have the same level of impairment, nor do they all warrant professional intervention. Simply because a person has a physical disorder does not necessarily imply a need for professional intervention, nor will all mental disorders have the same implications for need for treatment.

Stigmatization

The APA DSM is, of course, also used to communicate information to the general public concerning psychopathology, and a recurring issue for the DSM throughout its history are matters of stigmatization. As expressed by Hinshaw and Stier (2008), “Despite clear gains in public knowledge related to mental illness over the past half-century, levels of stigmatization as appraised by attitude surveys appear to have increased rather than decreased in the United States” (p. 368). Stigma contributes to lower rates of research funding, lower employment, poor housing, family burden, and personal shame. “Controlled research indicates that the negative impacts of stigmatization [can] outweigh the impairments related to various forms of mental disorders themselves” (Hinshaw & Stier, 2008, p. 369).

This concern has also been raised in regard to the diagnostic thresholds, more specifically, that with each edition of the DSM, there is an increasing pathologizing of normal problems in living (Cosgrove & Caplan, 2004; Horwitz, 2002; Houts, 2002; Kutchins & Kirk, 1997). As expressed by Hinshaw and Stier (2008), “Many in the general population are bound to be skeptical of the idea that 25% or more of the current population suffers from a mental disorder or that the lifetime risk is nearly 50%” (p. 376).

However, these concerns, much like those of Regier and Narrow (2002), are driven as much by matters of clinical utility as they are by validity (Widiger & Miller, 2009). Perhaps it should be more surprising to find that scientific research and increased knowledge have failed to lead to the recognition of more instances of psychopathology rather than being concerned that the DSM is increasing with each subsequent revision (Wakefield, 1998, 2001). In fact, there is quite a bit of data to suggest that the present manual is actually inadequate in its coverage (Judd, Schettler, & Akiskal, 2002; Magruder & Calderone, 2000). One of the more common diagnosis in general clinical practice is not otherwise specified (NOS; Clark, Watson, & Reynolds, 1999). The NOS diagnosis is provided when a clinician has determined that psychopathology is present but the symptomatology fails to meet criteria for any one of the existing disorders. There are a number of reasons that clinicians use NOS, including the possibility that clinicians are not adhering to the diagnostic criteria. However, another explanation for clinicians providing the diagnosis of NOS for anxiety, mood, developmental, personality, eating, and other disorders may be a testament to the inadequate coverage of the present system (Bowlard, Zimmerman, & Gaus, 2007; Phelps, Angst, Katzow, & Sadler, 2008; Schmidt et al., 2008; Verheul & Widiger, 2004).

Rather than argue against expanding the DSM out of concerns regarding stigmatization (Cosgrove & Caplan, 2004; Hinshaw & Stier, 2008; Kutchins & Kirk, 1997), one could in fact argue for the expansion of the DSM out of concerns regarding stigmatization. “The fact that some psychiatric disorders have a surprisingly high prevalence rate in the general population does not necessarily mean that they have been inappropriately defined” (Kendell, 2002, p. 6). It is perhaps the assumption that only a small minority of the
population suffers from psychopathology that is contributing to a negative outgroup perception. As Kendell (2002) further suggested that

Some people do regard it as absurd that at any given time 10%–15% of the population is suffering from a mental disorder and that nearly 30% may do so in the course of a year; however, I believe that this reaction is related to the stigma associated with mental illness. (p. 6)

Optimal psychological functioning, as in the case of optimal physical functioning, might represent an ideal that is achieved by only a very small minority of the population (Vaillant, 2003). Few persons would suggest that it is unlikely that they will ever become physically ill, yet it seems that few persons would also suggest that it is likely that they will ever become mentally ill. It is unclear why the prevalence rate of mental disorders should be so much lower than the prevalence rate of physical disorders (Kendell, 2002), as if most persons have been fortunate to have obtained no problematic genetic dispositions or vulnerabilities and have never experienced significant stress, pressure, or conflict that strained or injured their psychological functioning (Widiger & Miller, 2009).

Treatment Planning

“The ‘holy grail’ of clinical utility is the positive effect of a change in the diagnostic system on [treatment] outcome” (First, Pincus, et al., 2004, p. 951). As noted earlier, the central and fundamental importance of treatment planning for the DSM is noted quite explicitly in the first paragraph of the introduction to the DSM–IV–TR: “Our highest priority has been to provide a helpful guide to clinical practice” (APA, 2000, p. xxiii).

Nevertheless, it might be with respect to treatment planning that the DSM–IV–TR is most problematic. As suggested by the chair and vice chair of the DSM–V in their evaluation of the present success of the DSM, “With regard to treatment, lack of treatment specificity is the rule rather than the exception” (Kupfer et al., 2002, p. xviii). This strikingly negative statement might be a bit harsh and overstated (Kendell & Jablensky, 2003). Any naturalistic study of treatment planning and implementation would readily identify quite a bit of differences in the treatment approach (for instance) of persons receiving a diagnosis of mental retardation versus sexual dysfunction versus substance dependence versus schizophrenia.

However, it is evident that the APA DSM is not fulfilling its anticipated potential in leading to specific treatment implications for many of the individual disorders commonly seen in general practice. As suggested by Charney et al. (2002) in their white paper for a DSM–V Research Planning conference,

it can be concluded that the field of psychiatry has thus far failed to identify a single neurobiological phenotypic marker or gene that is useful in making a diagnosis of a major psychiatric disorder or for predicting response to psychopharmacologic treatment. (p. 33)

This conclusion has led many to suggest that the DSM should be governed more strongly and specifically by treatment implications or, for similar reasons, by common etiology (Charney et al., 2002; Helzer & Hudziak, 2002; Phillips, Price, Greenberg, & Rasmussen, 2003; Tsuang, Stone, Tarbox, & Faraone, 2003).

One point of caution with respect to this emerging shift in the foundation for the diagnostic manual is that the emphasis, to date, has been largely on genetic etiology, endophenotypes, pathophysiology, and pharmacotherapy. There is a distinct and evident lean toward the neurophysiological perspective on etiology, pathology, and treatment in this literature and the suggested impact on the construction of the DSM (e.g., Charney et al., 2002; Helzer & Hudziak, 2002; Hyman, 2007; Phillips et al., 2003). Changes were in fact made to the DSM–IV on the basis of “implications for treatment selection” (APA, 1994, p. 781). These included the addition of the specifiers for atypical features, postpartum onset, and rapid cycling, as well as the inclusion of the new diagnosis of bipolar II disorder (First, Pincus, et al., 2004). In all of these instances, the revisions were based on expectations concerning pharmacological interventions. It is not clear whether any recent revisions have been guided by research or by expectations concerning any other form of clinical intervention.

One of the intentions of the authors of the DSM–III was to have the DSM be atheoretical, or at least be reasonably neutral with respect to alternative models of psychopathology (Spitzer et al., 1980). The DSM is used by clinicians and researchers from a wide variety of theoretical perspectives (Frances et al., 1989). Nevertheless, each theoretical perspective also seems to find the manual to be at least somewhat cumbersome and problematic for clinical care, including the interpersonal and systems theoretical perspectives (Beach, Wamboldt, Kaslow, Heyman, & Reiss, 2006), the psychodynamic (PDM Task Force, 2006), the behavioral (Folette & Houts, 1996), and even the neurobiological (Charney et al., 2002).

No DSM can be truly atheoretical or entirely neutral, particularly if its construction is to be guided by empirical research (Frances et al., 1989). If the empirical research favors one particular model relative to another (and it would be inevitable for this to occur, unless all theoretical models were equally valid), then perhaps the DSM–V will represent more heavily the theoretical model with the strongest empirical support. Nevertheless, it should probably continue to hold theoretical neutrality as a prominent and fundamental aspiration, particularly in the absence of conclusive research (Spitzer, 1985, 2001). The DSM–IV–TR is the authoritative language for professional and scientific communication. Impacting this language in any theoretical direction not only would provide an explicit authority for this particular perspective but also would have a more subtle and cumulative effect on subsequent scientific discourse and clinical practice (Wakefield, 1998). A language that favors one particular perspective does not provide an equal playing field.

The nature and extent to which the DSM should be guided or governed by any treatment perspective is also unclear. Not only would this further hinder the clinical application of the manual for treatment modalities left out of the construction of the manual, it is not really clear that the validity of a diagnosis depends on the existence of an effective treatment. This is like trying to identify what is wrong with a mechanism solely on the basis of the existing tools with which one has to fix it. Any such DSM is also likely to be confusingly unstable. As each new tool or technique for effecting change in mood, beliefs, attitudes, or behaviors becomes available, requisite changes in what is considered to be a disorder would need to occur. If the manual was based on response to pharmacological therapy, for instance, one would probably have to lower diagnostic thresholds substantially, as there is insufficient research to indicate that normal psychological functioning is not
responsive to pharmacologic intervention (Harmer, Hill, Taylor, Cowen, & Goodwin, 2003; Harmer, Shelley, Cowen, & Goodwin, 2004; Knutson et al., 1998).

Nevertheless, it is also the case that a DSM without clear or specific treatment implications is fundamentally flawed with respect to its purported highest priority, contributing once again to consideration being given to shifting the manual to a dimensional rather than to a categorical classification (Kupfer et al., 2002). The heterogeneity of diagnostic membership, the lack of precision in description, the excessive diagnostic co-occurrence, the failure to lead to a specific diagnosis, the reliance on the NOS wastebasket diagnosis, and the unstable and arbitrary diagnostic boundaries of the DSM–IV–TR diagnostic categories all contribute to difficulties in identifying specific treatment implications (Smith & Combs, in press; Widiger & Mullins-Swheat, 2009).

Research

Changes in DSM-IV were made with the explicit goal of improving clinical utility . . . . [However] no formal effort was made to empirically examine whether these changes actually improved clinical utility. Instead, the field trials and data reanalyses primarily evaluated proposed criteria sets in terms of reliability, validity (using clinical diagnoses as the standard), and the extent to which the proposed criteria set identified different individuals as having the disorder. Purported improvements in clinical utility were simply assumed to be the case. (First, Pincus, et al., 2004, p. 947)

There has been quite a bit of research on the diagnostic activity and judgments of clinicians. However, much of this research has been concerned with the validity of clinicians’ diagnoses and assessments (Garb, 2005). Even direct surveys of clinicians’ opinions concerning a revision to the DSM (e.g., user acceptability surveys) will often concern matters of validity rather than utility. For example, during the construction of the ICD-10 (Sartorius et al., 1993) and DSM–IV (Setterberg et al., 1991), clinicians were surveyed as to whether they supported various proposed revisions. The support provided by these clinicians could reflect simply their opinion regarding validity rather than clinical utility. The authors of the DSM–III avoided such opinion surveys because “no one wanted to repeat the scene of the general membership voting on a presumably ‘scientific’ issue, as was done in 1973 on the issue of the elimination of homosexuality from the DSM-II classification” (Spitzer et al., 1980, p. 152). Surveys of user acceptability or support can be informative, particularly with regard to anticipating likely objections, concerns, and problems in the implementation of a diagnostic revision, but validity questions are perhaps best answered through more traditional construct validity research (Smith & Combs, in press). Decisions should be informed by a fair hearing of the diversity of perspectives, and these viewpoints and perspectives should be systematically and enthusiastically solicited. Nevertheless, the most scientifically valid decision may at times be “politically incorrect.” The authors of the DSM should have the authority to make innovative decisions that are scientifically justified even when they are contrary to general clinical consensus (Widiger & Clark, 2000).

What would be useful are surveys and studies that directly concern matters of clinical utility, more specifically, ease of usage, communication, and treatment. For example, a number of clinical utility studies have been conducted with regard to the proposal to replace the DSM–IV–TR personality disorder diagnostic categories with the five-factor dimensional model (FFM) of general personality structure. The first head-to-head comparison of the clinical utility of dimensional systems for personality diagnosis was provided by Sprock (2003). She reported that clinicians found the diagnostic categories to be more useful for communication and treatment planning than the FFM. A fundamental limitation of that study, however, is that the case vignettes were written in terms of the DSM diagnostic criterion sets. Samuel and Widiger (2006) obtained higher clinical utility ratings for the FFM than for the DSM–IV–TR using case vignettes of actual persons (e.g., Ted Bundy). As noted earlier, Spitzer et al. (2008) reported higher utility ratings for the prototypal matching procedure of Shedler and Westen (2004) than for an FFM profile description when describing patients currently being seen in clinical practice. However, this study confounded the dimensional model with the method of assessment. The FFM assessment required the consideration and completion of five to six pages of material, whereas the profile matching could be done in just 1–2 min (Westen et al., 2006). Love and Widiger (in press) reported no differences in the clinical utility ratings for the FFM and the Shedler-Westen Assessment Procedure-200 (Shedler & Westen, 2004) when the constructs of these alternative dimensional models were assessed in a comparable manner. Mullins-Swheat, Smith, Verheul, Oldham, and Widiger (2009) reported in a direct survey of psychologists and psychiatrists from the United States and the Netherlands that clinicians considered abnormal personality constructs to be relatively more useful in clinical treatment than normal personality constructs, but they did still nevertheless identify a substantial number of normal personality constructs that they felt should be included in a future edition of the DSM.

It would also be useful to conduct more naturalistic studies of how the diagnostic manual is being used in general clinical practice. A purported strength of the prototypal matching approach is that it is said to “conform to the way clinicians think about psychiatric diagnoses” (First & Westen, 2007, p. 473). However, there has in fact been little research to support this hypothesis. There are also data to suggest that clinicians instead reach a diagnostic decision on the basis of just one or two individual diagnostic criteria rather than matching to a global prototype (e.g., Garb, 2005; Herkov & Blashfield, 1995; Miller, 2008). When provided with a narrative paragraph describing a prototypic case, consisting of 15–20 sentences, clinicians may not in fact consider the global gestalt, but may instead focus on just one or two key features that they consider to be particularly salient for the respective mental disorder.

Naturalistic studies on treatment planning and decisions would also be useful, particularly with regard to existing and proposed diagnostic categories and modifiers. As suggested by First, Pincus, et al. (2004), “improvement in clinical outcome can be assessed by measuring reduction in symptom severity, measuring improvement in functioning, or documenting the prevention of a negative future outcome (e.g., reduction in relapse rates over a period of time)” (p. 951). In other words, would the revision to the DSM actually have a meaningful benefit for the treatment of patients? There are, of course, quite a few controlled treatment outcome studies, but very few naturalistic studies of treatment practice in general clinical care, and even fewer that test the impact of proposed revisions to the diagnostic manual. In summary, surpris-
ingly, very little is really known about how effectively the diagnostic manual is actually impacting clinical care.

Conclusions

Clinical utility and construct validity are not entirely independent concepts. An entirely invalid or illusory diagnostic construct would have no clinical utility, and an entirely valid diagnostic construct would have at least some clinical utility even if (for instance) there were no viable or remotely effective treatment implications. There would still be some utility (for instance) in being able to communicate with the patient, public health agencies, and the family concerning the existence of the untreatable disorder. Nevertheless, validity and utility also fail to be synonymous concepts, and it was the purpose of the present article in part to indicate the importance in not conflating them, particularly when making decisions concerning the construction of the DSM.

A further purpose of this article was to attempt to draw more direct attention to and interest in matters of clinical utility. All else being equal, improving the validity of the diagnostic manual should trump improving its clinical utility. However, it is never really the case that for any particular decision regarding the DSM that all else is in fact equal. Clinical utility may not weigh as importantly as validity, but it should still be a significant part of the equation (Rorer et al., 1966). For instance, an entirely valid DSM could actually have relatively little clinical utility if it is not feasible for usage in general clinical practice (First, 2005). Of central importance to the authors of the DSM, composed primarily of researchers, have been matters of construct validity. Of central importance to clinicians, largely excluded from the construction of the DSM, are matters of clinical utility. The function of the DSM is not simply a matter of addressing a scientific interest in understanding and explaining psychopathology; its ultimate purpose is to help reduce pain and suffering within the general population—more specifically, to facilitate the practice of clinicians administering clinical care.

References


