

Challenges in Operationalizing the *DSM-IV* Clinical Significance Criterion

Janette Beals, PhD; Douglas K. Novins, MD; Paul Spicer, PhD; Heather D. Orton, MS; Christina M. Mitchell, PhD; Anna E. Barón, PhD; Spero M. Manson, PhD; and the AI-SUPERPPF Team

Background: An explicit clinical significance (CS) criterion was added to many *DSM-IV* diagnoses in an attempt to more closely approximate the clinical diagnostic process and reduce the proportion of false positives in epidemiological studies. The American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project (AI-SUPERPPF) offered a unique opportunity to examine the success of this effort.

Objective: To determine the impact of distress, impairment, and help-seeking reported in a lay structured interview on concordance with a clinical reappraisal. Further, to test the efficacy of 5 operationalizations of CS on the concordance and prevalence of *DSM-IV* lifetime disorders.

Design: Completed between 1997 and 2000, a cross-sectional probability sample survey with clinical reappraisal of approximately 10% of participants.

Setting: General community.

Participants: A population-based sample of 3084 members of 2 American Indian tribal groups, who were between the ages of 15 and 54 years and resided on or near their home reservations, were randomly sampled from the tribal rolls and participated in structured psychiatric interviews. Clinical reappraisals were conducted with approximately 10% of the lay-interview participants. The

response rate for the lay interview was 75%, and for the clinical reappraisal it was 72%.

Main Outcomes Measures: The AI-SUPERPPF Composite International Diagnostic Interview (CIDI), a culturally adapted version of the CIDI, University of Michigan version. Adapted to assess *DSM-IV* diagnoses, questions assessing the CS criterion were inserted in all diagnostic modules. The Structured Clinical Interview for *DSM-III-R* (SCID) was used in the clinical reappraisal.

Results: Most participants who qualified as having AI-SUPERPPF CIDI lifetime disorders reported at least moderate levels of distress or impairment. Evidence of increased concordance between the CIDI and the SCID was lacking when more restrictive operationalizations of CS were used; indeed, the CIDI was very likely to underdiagnose disorders compared with the SCID (false negatives). Concomitantly, the CS operationalizations affected prevalence rates dramatically.

Conclusion: The CS criterion, at least as operationalized to date, demonstrates little effectiveness in increasing the validity of diagnoses using lay-administered structured interviews.

Arch Gen Psychiatry. 2004;61:1197-1207

Author Affiliations: American Indian and Alaska Native Programs (Drs Beals, Novins, Spicer, Mitchell, and Manson, Ms Orton, and the AI-SUPERPPF Team) and Department of Preventive Medicine and Biometrics (Dr Barón), University of Colorado Health Sciences Center, Aurora. The names of the AI-SUPERPPF team authors are listed at the end of this article.

RECENT ADVANCES IN THE *Diagnostic and Statistical Manual of Mental Disorders (DSM)*¹ have focused on developing definitions of mental disorder that faithfully represent clinicians' experiences and can be consistently replicated among practitioners.² Although primarily a clinical tool, the *DSM*'s formal operationalizations of disorder also allow researchers to design and field structured and semistructured interview protocols, mak-

ing possible estimation of the prevalence and incidence of the more common mental disorders within populations. Recently, critiques of such estimates have shifted their focus from reliability to validity.³⁻⁶ In particular, many question whether structured lay interviews overestimate the rates of disorder.³ This "false positive" problem is probably best understood in the context of the largest psychiatric epidemiological studies in the United States to date. The Epidemiologic Catchment Area (ECA) stud-

ies⁷ and the National Comorbidity Surveys (NCS)^{8,9} suggest that, in any given year, almost 20% to 30% of the population experiences a mental or addictive disorder, while lifetime rates range between 32% and 49%.^{4,7} Clearly, the implications of such estimates for mental health policy in the United States are enormous and have called into question whether these rates accurately reflect the need for treatment.^{3,10-13}

Critiques of this epidemiological research focus on whether the diagnoses and the criteria defining them successfully differentiate disorder from “problems of living,”^{14,15} in other words, whether the thresholds for disorder in such instruments are too low. Hence, in the revisions that occurred between the *DSM-III-R*¹⁶ and the *DSM-IV*,¹⁷ an explicit clinical significance (CS) criterion was added to many diagnoses to address at least 1 type of overinclusion: those meeting symptomatic criteria for a disorder but for whom such problems were mild. Although worded somewhat differently across diagnoses, the CS criterion typically asserts that “the symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.”^{17(p1857)}

The conceptual utility and validity of the CS criterion have been much debated^{3,13,14,18-20}; however, empirical examinations of the CS criterion have only recently appeared in the literature. Narrow et al⁴ and Regier et al,²¹ in secondary analyses of the ECA and NCS studies, ascribed CS to participants who had either sought help for specific disorders or reported that these problems interfered with their lives or activities “a lot.” With this operationalization, admittedly limited by the data available in these surveys that predated the *DSM-IV*, past-year prevalence rates decreased 17% for the ECA (based on the *DSM-III*²²) and 32% for the NCS (*DSM-III-R*). Participants meeting both the CS and the symptomatic criteria were more likely than those meeting only symptomatic criteria to have sought services for mental health problems, to have reported either not being able to work or needing to cut down on work, and to have been suicidal. Slade and Andrews,²³ using data from the Australian National Survey of Mental Health and Well-Being (ANSMHWB),²⁴ found that the inclusion of significant self-reported distress or impairment also decreased the *DSM-IV* rates of past-month disorder, between 19% for major depressive disorder and 65% for obsessive-compulsive disorder. Controlling for sociodemographics and comorbidity, participants with significant distress/impairment were more likely to report help-seeking, distress, and impairment in other parts of the interview—but only for some diagnoses.

These reanalyses of the ECA, NCS, and ANSMHWB demonstrate that adding CS to the symptomatic criteria substantially lowers prevalence rates and may increase the severity of the resulting diagnoses. However, they also raise additional questions. Each study operationalized CS somewhat differently; also, none was able to compare the impact that multiple operationalizations of CS might have on prevalence. For instance, should only individuals reporting “a lot” of distress or impairment be considered to have met CS or might those reporting moderate or mild distress also qualify as disordered?²⁰ What role should help-seeking play in this calculus? And what is the de-

gree of overlap among distress, impairment, and help-seeking?⁵ Finally, none of these efforts used direct assessments of whether including the CS criterion would increase the concordance between lay- and clinician-administered interviews.²⁵

Data from the American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project (AI-SUPERPFP) provided us an opportunity to address these questions. Designed before the World Health Organization’s *DSM-IV* version of the Composite International Diagnostic Interview (WHO-CIDI 2.1,²⁶ used by Slade and Andrews²³) was widely available, the AI-SUPERPFP independently supplemented the CIDI, University of Michigan version (UM-CIDI,²⁷ used in the NCS) with items necessary to assess *DSM-IV* criteria in a format conducive to investigating the utility of different operationalizations of CS and the overlap among these constructs. A clinical reappraisal of more than 10% of the sample enabled us to investigate whether the concordance between the lay- and clinician-administered interviews increased with the inclusion of CS.

METHODS

AI-SUPERPFP LAY-INTERVIEW DESIGN AND SAMPLES

The AI-SUPERPFP sought to estimate the prevalence of psychiatric disorders and health service utilization in 2 American Indian reservation populations. The AI-SUPERPFP methods are described in detail elsewhere²⁸; the interview and training manual are available online at <http://www.uhsc.edu/ai/nciamnhr/presentresearch/superprj.htm>. The populations of inference were enrolled members of either 2 closely related Northern Plains tribes or a Southwest tribe who were 15 to 54 years old at the time of development of the sample frame (1997) and who lived on or within 20 miles of their reservations. Once located and found to be eligible, 73.7% (n=1446) from the Southwest tribe and 76.8% (n=1638) from the Northern Plains tribe agreed to participate. Tribal approvals were obtained prior to the project’s beginning. Informed consent was acquired from all participants; with minors, parental/guardian consent was obtained before adolescent assent.

AI-SUPERPFP CIDI

As explained in greater detail elsewhere,²⁸ carefully considered cultural adaptations had already been completed as part of a previous effort conducted between 1992 and 1995.²⁹ Because that process preceded release of the WHO-CIDI 2.1, this measure was independently augmented and adapted to render it consistent with the *DSM-IV*.

The present analyses included lifetime diagnoses of the AI-SUPERPFP disorders: panic disorder (PD), generalized anxiety disorder (GAD), posttraumatic stress disorder (PTSD), dysthymic disorder (DD), major depressive episode (MDE), substance abuse, and substance dependence. Substances included alcohol, sedatives, tranquilizers, stimulants, analgesics, inhalants, marijuana, cocaine, hallucinogens (including peyote), and heroin—each individually assessed and then combined into substance abuse or substance dependence. Additional aggregations included any anxiety disorder (those with GAD, PD, or PTSD), any depressive disorder (MDE or DD), any substance disorder (either abuse or dependence), any anxiety or depressive disorder, and any disorder.

	Anxiety Disorders			Mood Disorders		Substance Disorders	
	Panic Disorder (PD)	Generalized Anxiety Disorder (GAD)	Posttraumatic Stress Disorder (PTSD)	Major Depressive Episode (MDE)	Dysthymic Disorder (DD)	Dependence	Abuse
<i>DSM-IV Language</i>							
	None	<i>Separate Criterion:</i> "The Symptoms Cause Clinically Significant Distress or Impairment in Social, Occupational, or Other Important Areas of Functioning."			<i>Incorporated in Criterion A:</i> "A Maladaptive Pattern of Substance Use Leading to Clinically Significant Impairment or Distress."		
DIS Items Used in Narrow et al ⁴ Reanalyses of Epidemiologic Catchment Area (ECA) Studies							
Probe Flowchart: Syndrome Level		Diagnosis Not Assessed	Diagnosis Not Assessed	Help-Seeking, Medication Use, or a Lot of Impairment		Help-Seeking, Medication Use, or a Lot of Impairment	
Probe Flowchart: Symptom Level	Help-Seeking, Medication Use, or a Lot of Impairment	Diagnosis Not Assessed	Diagnosis Not Assessed		Help-Seeking, Medication Use, or a Lot of Impairment		
UM-CIDI Items Used in Narrow et al ⁴ Reanalyses of Baseline National Comorbidity Survey (NCS)							
Impairment in Important Areas of Functioning: Syndrome Level	How Much Did [Your Symptom(s) or Avoidance of Situations That Might Engender Symptoms] Interfere With Your Life or Activities—a Lot, Some, a Little, or Not at All?		Not Asked	How Much Did [Your Symptom(s)] Interfere With Your Life or Activities—a Lot, Some, a Little, or Not at All?			
Help-Seeking: Syndrome Level	Did You Ever Tell a, (a) Psychiatrist, (b) Mental Health Specialist, (c) Other Professional, or (d) Take Medications More Than Once Because of X? (4 Questions)			Did You Ever Tell a, (a) Psychiatrist, (b) Mental Health Specialist, (c) Other Professional, or (d) Take Medications More Than Once Because of X? (4 Questions)			
WHO-CIDI 2.1 Items Used in Slade and Andrews ²¹ Reanalyses of Australian National Survey of Mental Health and Well-Being (ANSMHWP)							
Distress	Not Asked	Have You Ever Been Very Upset With Yourself for Feeling Worried, Tense, or Anxious for Long Periods of Time?	Have You Ever Been Very Upset With Yourself for Having Problems That Occurred as a Result of [Event]?	Not Asked			
Impairment in Important Areas of Functioning	Not Asked	Did the Period of Worry, Tension, or Anxiety Ever Interfere With Your Life or Activities a Lot?	Have the Problems That Occurred as a Result of [Event] Ever Kept You From Going to a Party, Social Event, or Meeting?	Did Any Period Lasting 2 Weeks or Longer Seriously Interfere With Your Ability to Do Your Job, Take Care of Your House or Family, or Take Care of Yourself?	Not Asked		
Probe Flowchart: Syndrome Level	Help-Seeking, Medication Use, or a Lot of Impairment		Not Asked	Help-Seeking, Medication Use, or a Lot of Impairment		Not Asked	
Probe Flowchart: Symptom Level	Not Asked		Not Asked	Asked for 2 Screeners	Asked for 1 Screener	Not Asked	
American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project (AI-SUPERPPF) CIDI Items							
Distress	How Much Did X Ever Bother or Upset You? Would You Say Not at All, Some, or a Lot?						
Impairment in Important Areas of Functioning	How Much Did X Ever Interfere With Your Life or Activities? Would You Say Not at All, Some, or a Lot?						
Help-Seeking	Did You Ever Talk to a (a) Mental Health Specialist or (b) Medical Person? (c) Did You Ever Receive Treatment for X? Did You Ever Take Medicine That Was Prescribed by a Doctor for X? (d) Did You Ever Take Medicine You Got in a Store Without a Prescription for X?						
	(a) Did You Ever Talk to a Healer or Spiritual or Traditional Leader About X? (b) Did You Ever Take Medicine That Was Given to You by a Medicine Man or Woman for X? (c) Did You Ever Take Any Other Herbs, Roots, or Remedies for Problems Like These?						

Figure 1. Definitions of clinical significance criteria in *DSM-IV* and items used in 3 studies to operationalize clinical significance construct. Shaded cells indicate the additions to the diagnostic algorithms in Narrow et al⁴ and Slade and Andrews.²³ DIS indicates Diagnostic Interview Schedule; UM-CIDI, Composite International Diagnostic Interview, University of Michigan version; and WHO, World Health Organization.

ASSESSMENT OF CS CONSTRUCTS

Figure 1 summarizes the operationalizations of CS by Narrow et al⁴ and Regier et al²¹ using the Diagnostic Interview Schedule (DIS)³⁰ and the UM-CIDI, those of Slade and Andrews²³ using the WHO-CIDI 2.1 in the ANSMHWP, and, finally, the AI-SUPERPPF CIDI measures of distress, impairment, and help-seeking.

Because the ECA-DIS and UM-CIDI predated the *DSM-IV*, they did not assess CS directly. However, the ECA-DIS used the probe flowchart to distinguish symptoms and syndromes (groupings of symptoms) that were possibly psychiatric from others—a judgment that overlaps but is not synonymous with CS. This set of probes commenced with the question "Did you ever tell a doctor about your [problems just endorsed]?" and continued with questions about other help-seeking, use of medication, and impair-

ment in terms of the specific symptom/syndrome. When the probe flowchart was assessed at the symptom level, only individuals with probable psychiatric symptoms advanced in the diagnostic modules. As shown in Figure 1, the ECA-DIS used the probe flowchart at the symptom level for PD and DD, and thus the probe flowchart was a necessary component of these diagnoses. On the other hand, the probes for MDE and substance disorders were asked at the syndrome level and have been only recently introduced into the prevalence calculus in the work by Narrow et al⁴ and Regier et al.²¹ Similarly, the UM-CIDI assessed help-seeking, medication use, and impairment at the syndrome level; again, these data did not influence prevalence estimates until the recent secondary analyses. In the WHO-CIDI 2.1 algorithms, an instrument designed to assess *DSM-IV*, the probe flowchart was used at the symptom level only for MDE and DD; otherwise, it was used at the syndrome level. The distress and impairment questions specific to the *DSM-IV* were asked for GAD, PTSD, and MDE (impairment only).

In the AI-SUPERPFP CIDI, identical items assessed self-reported distress, impairment, and help-seeking at the syndrome level across diagnoses. Impairment and help-seeking questions were patterned after those of the UM-CIDI but altered slightly to maximize cultural validity. For instance, whereas the UM-CIDI included “a little” as a response option in its impairment items, this was dropped in the AI-SUPERPFP because focus groups suggested that participants were unlikely to reliably differentiate between “some” and “a little.” Help-seeking questions reflected the service ecology of reservation residents by including specific types of service providers (eg, community health representatives as medical personnel) and directly assessing use of traditional healing resources.

OPERATIONALIZATIONS OF CS

Using the AI-SUPERPFP data, 5 operationalizations of CS were compared. The first (CS0) excluded the CS criterion. The next 3 operationalizations built upon one another, ranging from less (CS1) to more (CS3) restrictive, and focused on degrees of self-reported distress and impairment, adhering closely to the *DSM-IV* CS language. The final operationalization (CS4) included both help-seeking and impairment and most closely mimicked the probe flowchart.

Operationalization CS0: no assessment of clinically significant distress or impairment. Diagnoses were based on symptomatic criteria only.

Operationalization CS1: “a lot or some” impairment or distress. A disorder was considered clinically significant if the participant reported “a lot” or “some” distress or impairment. This follows Wakefield and Spitzer’s³ suggestion that those expressing either moderate or severe distress or impairment should be considered “true” positives.

Operationalization CS2: “a lot” of distress or impairment or “some” of both. As a variation of the definition of moderate disability, we included an operationalization whereby experiencing “a lot” of distress or impairment or some deleterious effects in multiple domains merited a diagnosis. This operationalization represented a middle ground between CS1 and CS3.

Operationalization CS3: “a lot” of distress or impairment. Here participants reporting “a lot” of distress or impairment were considered to meet the definition of CS. This operationalization was closest to that of Slade and Andrews.²³

Operationalization CS4: help-seeking or “a lot” of impairment.⁴ This operationalization most closely matches the work of Narrow et al⁴ and Regier et al,²¹ albeit at the syndrome level. Seeing or talking to a mental health provider or other medical personnel about the psychiatric symptoms or having been hospitalized for these problems constituted help-seeking.

CLINICAL REAPPRAISAL

The AI-SUPERPFP included a clinical reappraisal of approximately 10% (n=335) of participants, who were reinterviewed by psychiatrists or clinical psychologists. This component was designed to assess the concordance between the AI-SUPERPFP CIDI and the Structured Clinical Interview for *DSM-III-R*, nonpatient version (SCID).³¹ Approximately 75% of the clinical reappraisal sample was randomly chosen based on a positive CIDI diagnosis of the 3 most common disorders: MDE, PTSD, or alcohol abuse/dependence. The remaining sample, also randomly selected, did not qualify for any AI-SUPERPFP diagnosis. Because the greatest source of error between lay and clinical interviews is commonly found in those who have some but not all of the symptoms required for a diagnosis (subthreshold cases),³² approximately half of those in the no-disorder group endorsed significant levels of depressed, anxious, or irritable symptoms on a checklist independent of the CIDI, with the remainder having few or none of these symptoms. The SCID was adapted to allow for changes between the *DSM-III-R* and the *DSM-IV*, including the CS criterion. The 8 clinician interviewers had extensive clinical experience (more than 15 years on average) and had worked in American Indian communities. Before entering the field, each demonstrated a high level of interrater reliability ($\kappa \geq .80$) in a series of videotapes coded by an expert panel, and they also performed supervised interviews with members of the local American Indian community. Furthermore, all clinical reappraisals were audiotaped and reviewed by master clinicians for quality assurance purposes. The response rate for the clinical reappraisal sub-study was 72.3% and was similar to that for the CIDI-disordered and -nondisordered participants. An average of 120 days elapsed between lay and clinical interviews. There was no association between the time elapsed between interviews and agreement between the CIDI and the SCID. Clinicians were blind to participants’ CIDI diagnostic status.

STATISTICAL METHODS

Variable construction and noninferential analyses were completed using SAS,³³ SPSS,³⁴ and Stata.³⁵ First, to better understand the patterns of distress, impairment, and help-seeking among participants meeting symptom criteria for AI-SUPERPFP CIDI disorders, frequencies and associated confidence intervals for these constructs are presented in **Table 1** for each disorder. Two Venn diagrams (**Figure 2**) depict the overlap among these constructs for those qualifying for at least 1 disorder. As a second step, concordance between the AI-SUPERPFP CIDI and the SCID was evaluated by assessing the numbers of true and false positives and negatives generated by the CIDI when the SCID was considered the gold standard, as well as the following set of standard statistics: (1) Cohen κ ³⁶; (2) sensitivity and specificity; (3) positive and negative predictive values calculated using the Bayes rule³⁷; and (4) the McNemar χ^2 test (a measure of bias).³⁸ Although not an assessment of concordance, the Global Assessment of Functioning³⁹ scores, determined by the clinicians during the reappraisal, provided a measure of severity of disorder and are included in **Table 2**. Finally, **Table 3** presents the differential prevalences, with associated 95% confidence intervals, of specific and aggregated disorders across the 5 CS operationalizations using the AI-SUPERPFP CIDI. The results in Table 1 and Table 3 provide inferences to the populations and were conducted in Stata³⁵ using sample and nonresponse weights. Since the concordance analyses in Table 2 focus on relative functioning of instruments, unweighted estimates were deemed acceptable³² and offered the added benefit of being able to provide the actual numbers of participants in various cells in the table.

Table 1. Prevalence of Self-reported Distress, Impairment, and Help-Seeking Among Those Meeting Symptom Criteria for AI-SUPERPPF DSM-IV Lifetime Disorders

	No.	Distress		Impairment		Help-Seeking		
		How Much Did Feeling X Ever Bother or Upset You?		Interfere With Life or Activities		Did You Ever Talk to Y About X?		
		A Lot, % (95% CI)	Some or a Lot, % (95% CI)	A Lot, % (95% CI)	Some or a Lot, % (95% CI)	Biomedical, % (95% CI)	Traditional, % (95% CI)	Combined, % (95% CI)
Any disorder	1342	46.5 (43.7-49.4)	86.4 (84.4-88.3)	33.0 (30.3-35.7)	82.2 (80.0-84.3)	48.0 (45.2-50.8)	31.0 (28.5-33.6)	57.6 (54.9-60.4)
Anxiety disorders								
PD	105	49.6 (39.3-60.0)	90.5 (84.1-97.0)	30.5 (21.1-39.8)	80.5 (72.2-88.8)	58.3 (48.5-68.1)	50.0 (40.8-59.1)	76.4 (68.1-84.7)
GAD	89	46.7 (35.4-57.9)	92.2 (86.2-98.1)	35.0 (24.3-45.8)	89.0 (81.5-96.6)	53.8 (43.5-64.0)	49.2 (38.4-60.0)	71.3 (61.1-81.5)
PTSD	537	44.5 (40.1-48.9)	90.6 (88.0-93.2)	22.7 (19.1-26.3)	77.0 (73.2-80.7)	37.0 (32.8-41.3)	25.3 (21.5-29.0)	47.5 (43.1-52.0)
Any anxiety disorder	638	45.9 (41.9-50.0)	90.3 (87.8-92.7)	25.3 (21.8-28.8)	77.8 (74.4-81.2)	43.0 (39.0-47.0)	32.2 (28.5-35.9)	55.4 (51.4-59.4)
Depressive disorders								
MDE	320	38.0 (32.4-43.6)	94.5 (91.9-97.2)	30.6 (25.3-35.9)	89.5 (86.1-93.0)	45.8 (40.1-51.5)	36.5 (31.1-41.8)	57.5 (51.9-63.2)
DD	103	45.4 (35.2-55.5)	95.7 (91.5-99.9)	36.0 (26.1-45.9)	90.3 (84.4-96.1)	44.5 (34.2-54.9)	44.8 (34.8-54.8)	60.4 (50.3-70.5)
Any depressive disorder	436	38.5 (33.1-44.0)	94.7 (92.1-97.3)	29.8 (24.6-35.0)	89.5 (86.0-93.0)	45.2 (39.6-50.8)	36.9 (31.7-42.0)	57.2 (51.7-62.7)
Any depressive or anxiety disorder	794	46.2 (42.6-49.8)	91.6 (89.5-93.6)	28.6 (25.3-31.8)	82.0 (79.2-84.8)	44.9 (41.3-48.5)	35.0 (31.6-38.3)	57.5 (53.9-61.0)
Substance disorders								
Substance dependence	436	55.9 (50.8-61.1)	93.1 (90.7-95.6)	43.4 (38.2-48.6)	94.2 (91.7-96.6)	52.8 (47.6-58.0)	23.6 (19.4-27.9)	59.9 (54.8-65.0)
Substance abuse	505	25.8 (21.7-30.0)	68.5 (64.1-72.9)	17.6 (14.1-21.2)	65.1 (60.5-69.6)	29.4 (25.2-33.8)	19.4 (15.6-23.2)	38.1 (33.6-42.7)
Any substance disorder	941	39.8 (36.4-43.3)	80.0 (77.2-82.7)	29.6 (26.4-32.8)	78.6 (75.7-81.5)	40.3 (36.9-43.8)	21.4 (18.6-24.2)	48.2 (44.8-51.7)

Abbreviations: AI-SUPERPPF, American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project; CI, confidence interval; DD, dysthymic disorder; GAD, generalized anxiety disorder; MDE, major depressive episode; PD, panic disorder; PTSD, posttraumatic stress disorder.

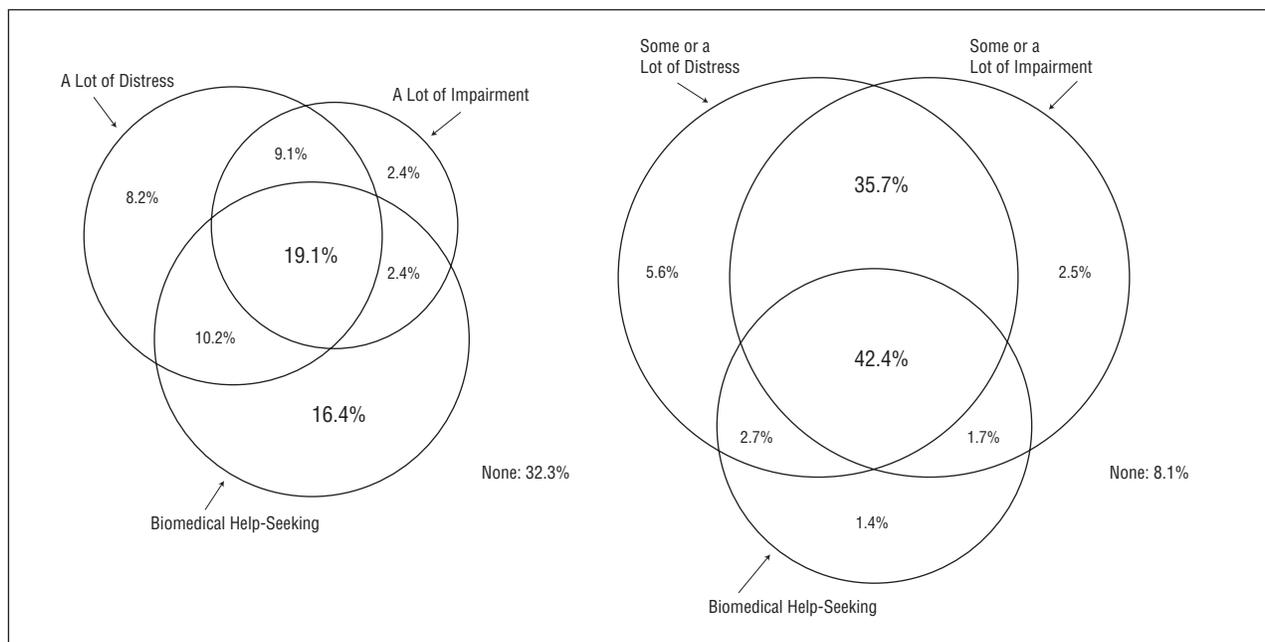


Figure 2. Any American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project disorder: overlap between distress, impairment, and help-seeking.

RESULTS

CS CONSTRUCTS

Table 1 presents the self-reported distress, impairment, and help-seeking for participants meeting the symptom

criteria for DSM-IV lifetime disorders in AI-SUPERPPF. When considering distress, among those with any AI-SUPERPPF CIDI disorder, 46.5% reported being upset or bothered “a lot” (distressed) by their symptoms. Between 35% and 50% of those with depressive or anxiety disorders reported such levels of distress. Among those

Table 2. Concordance Between the AI-SUPERPPF CIDI and the SCID for Lifetime Disorders*

		No.				Cohen κ	Sensitivity, %	Specificity, %	PPV, † %	NPV, † %	Bias ‡	GAF §
		True Positives	False Negatives	False Positives	True Negatives							
		Positive SCID and Positive CIDI	Positive SCID and Negative CIDI	Negative SCID and Positive CIDI	Negative SCID and Negative CIDI							
No Clinical Significance Criterion												
CS0	No assessment of clinically significant distress or impairment	206	49	22	51	0.45	81	70	89	55	35.8	65.2
Constructs That Have Been Used to Define Clinical Significance												
Distress	A lot	128	127	8	65	0.25	50	89	93	37	117.6	64.8
	Some or a lot	193	62	15	58	0.45	76	79	92	52	49.1	65.2
Impairment	A lot	90	165	8	65	0.14	35	89	91	31	155.5	62.8
	Some or a lot	182	73	15	58	0.40	71	79	91	48	59.5	64.2
Help-seeking	Biomedical	117	138	7	66	0.22	46	90	93	36	129.5	62.4
	Traditional	68	187	5	68	0.10	27	93	92	30	180.2	64.5
	Either biomedical or traditional	140	115	7	66	0.30	55	90	94	40	106.6	63.5
Operationalizations of Clinical Significance												
CS1	"A lot" of or "some" distress or impairment	197	58	17	56	0.45	77	77	91	53	44.6	65.0
CS2	"A lot" of distress or impairment or "some" of both	184	71	15	58	0.41	72	79	91	49	57.6	64.4
CS3	"A lot" of distress or impairment	135	120	10	63	0.26	53	86	92	38	109.0	64.1
CS4¶	Help-seeking or "a lot" of impairment	143	112	10	63	0.28	56	86	92	40	101.1	63.4

Abbreviations: AI-SUPERPPF, American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project; CIDI, Composite International Diagnostic Interview; CS, clinical significance; GAF, Global Assessment of Functioning; NPV, negative predictive value; PPV, positive predictive value; SCID, Structured Clinical Interview for *DSM-III-R*; WHO-CIDI 2.1, World Health Organization's *DSM-IV* version of the CIDI.

*Statistical tests between operationalizations were inappropriate because they do not represent independent samples. Similar conclusions were reached whether or not the data were weighted back to the lay sample or if the data were restricted to those disorders for which *DSM-IV* includes specific CS language; for ease of presentation, the unweighted results of all disorders are reported.

†Calculated using the Bayes rule (that is, using the SCID prevalence weighted back to the full population).

‡Calculated using the McNemar χ^2 test.

§Mean score for those with a disorder under this operationalization.

||The AI-SUPERPPF approximation of WHO-CIDI 2.1 operationalization.

¶The AI-SUPERPPF approximation of Narrow and Regier operationalization.

with substance problems, a majority with dependence reported high levels of distress compared with about one quarter with substance abuse. When reports of "some" distress were included, more than 90% of participants who were assigned a diagnosis other than substance abuse reported being upset or bothered by their symptoms. Overall, an additional 39.9% (95% confidence interval, 36.4%-46.5%) with any disorder reported being bothered or upset "some" when compared with "a lot," with the difference somewhat larger for depressive disorders (56.2%) than either anxiety disorders (44.4%) or substance disorders (40.2%).

Similarly, for impairment, one third of participants with any disorder reported that their symptoms interfered with their lives and activities "a lot." Those with substance dependence ranked highest, while those with substance abuse ranked lowest. The difference between "a lot" and

"some" was 49.2% overall and followed patterns similar to those seen for distress.

Turning to help-seeking, almost half of the participants qualifying for 1 of the AI-SUPERPPF disorders had sought biomedical help for their symptoms; about one third had sought help from traditional healing sources. Generally, the pattern of seeking help from traditional sources mirrored that from biomedical sources. Combining traditional with biomedical sources of help-seeking increased the rates by 9.6%.

CONCORDANCE ANALYSES

Overall, 78% of the clinical reappraisal sample was judged by the SCID interviewers to have a disorder; without considering CS (that is, CS0), the CIDI designated 70% of this select sample as having at least 1 *DSM-IV* disorder.

Table 3. Lifetime Prevalence of *DSM-IV* Disorders Across Operationalizations of Distress and Impairment

	Operationalizations of CS, % (95% CI)				
	CS0	CS1	CS2	CS3*	CS4†
	No Assessment of Clinically Significant Distress or Impairment	"A Lot" of or "Some" Distress or Impairment	"A Lot" of Either Distress or Impairment or "Some" of Both	"A Lot" of Either Distress or Impairment	Help-Seeking or "A Lot" of Impairment
Any disorder‡	45.5 (43.6-47.3)	40.2 (38.4-42.0)	35.6 (33.8-37.4)	22.8 (21.2-24.4)	26.4 (24.8-28.1)
Imposed on relevant disorders only§		43.5 (41.7-45.3)	43.3 (41.4-45.1)	38.2 (36.4-40.0)	32.5 (30.7-34.3)
Anxiety disorders‡					
PD	3.4 (2.7-4.0)	3.2 (2.5-3.8)	2.7 (2.1-3.3)	1.7 (1.3-2.2)	2.2 (1.7-2.8)
GAD	2.9 (2.2-3.5)	2.8 (2.2-3.3)	2.4 (1.9-3.0)	1.5 (1.1-2.0)	1.9 (1.4-2.4)
PTSD	18.5 (17.0-19.9)	17.2 (15.7-18.6)	15.0 (13.7-16.4)	8.8 (7.7-9.8)	9.0 (7.9-10.0)
Any anxiety disorder‡	21.0 (19.5-22.4)	19.5 (18.0-20.9)	16.3 (15.0-17.7)	10.3 (9.2-11.4)	11.4 (10.2-12.5)
Imposed on relevant disorders only§		19.6 (18.1-21.0)	17.4 (16.0-18.8)	11.4 (10.2-12.5)	11.9 (10.8-13.2)
Depressive disorders‡					
MDE	10.5 (9.4-11.6)	10.2 (9.1-11.3)	9.2 (8.1-10.2)	4.7 (4.0-5.5)	6.3 (5.4-7.1)
DD	3.3 (2.6-3.9)	3.2 (2.6-3.8)	2.9 (2.3-3.5)	1.7 (1.3-2.2)	1.8 (1.4-2.3)
Any depressive disorder‡	11.0 (9.8-12.0)	10.6 (9.5-11.8)	9.5 (8.5-10.6)	4.9 (4.1-5.7)	6.4 (5.5-7.3)
Any depressive or anxiety disorder‡	26.1 (24.5-27.7)	24.6 (23.0-26.2)	21.4 (19.9-22.9)	13.0 (11.8-14.3)	14.9 (13.6-16.2)
Imposed on relevant disorders only§		24.6 (23.0-26.1)	22.1 (20.6-23.6)	14.0 (12.8-15.3)	13.0 (11.8-14.2)
Substance disorders‡					
Substance dependence	15.3 (14.0-16.7)	13.4 (12.1-14.7)	12.7 (11.4-14.0)	8.6 (7.5-9.7)	9.6 (8.5-10.8)
Substance abuse	17.1 (15.7-18.5)	12.3 (11.1-13.5)	9.3 (8.2-10.4)	4.8 (4.0-5.7)	6.1 (5.2-7.0)
Any substance disorder‡	32.4 (30.7-34.1)	26.1 (24.5-27.8)	22.4 (20.8-24.0)	13.6 (12.4-15.0)	16.0 (14.6-17.4)

Abbreviations: AI-SUPERPFP, American Indian Service Utilization, Psychiatric Epidemiology, Risk and Protective Factors Project; CI, confidence interval; CS, clinical significance; GAD, generalized anxiety disorder; PD, panic disorder; PTSD, posttraumatic stress disorder; WHO-CIDI 2.1, World Health Organization's *DSM-IV* version of the Composite International Diagnostic Interview.

*The AI-SUPERPFP approximation of WHO-CIDI 2.1 operationalization.

†The AI-SUPERPFP approximation of Narrow and Regier operationalization.

‡Aggregated disorders.

§Aggregated disorders restricted to relevant disorders only (that is, excluding panic and substance disorders). These aggregates include the CS0 operationalizations of PD and substance disorders because the *DSM-IV* does not explicitly specify CS for these disorders.

||The CS operationalizations presented here are for demonstration purposes only because the *DSM-IV* does not explicitly specify CS for PD or substance disorders.

The concordance between the AI-SUPERPFP CIDI and clinical reappraisals with the SCID is presented in Table 2. Similar to others' reports,^{38,40} agreement between these clinical and lay methods of case ascertainment was modest. However, the hypothesis tested here was whether inclusion of the CS operationalizations in the lay-interview data would more closely approximate clinical diagnoses.

Focusing first on distress, impairment, and help-seeking constructs, the κ values were lower for the more restrictive response patterns. Sensitivity assesses the ability of the CIDI to record a positive diagnosis for SCID-defined cases; here, a 26% decrease in sensitivities arose when we designated those having "a lot" of distress as meeting CS compared with including "some or a lot." In contrast, the specificity or ability of the CIDI to identify the SCID-defined noncases increased 10% between the same definitions. The positive predictive values indicated that the vast majority of the CIDI-defined cases also received SCID diagnoses. The negative predictive values revealed that as the CS criteria became more restrictive, a greater percentage of CIDI-defined noncases actually received SCID diagnoses. The bias was more severe as the requirements increased for being labeled distressed, impaired, and a help-seeker. Finally, although the Global Assessment of Functioning scores decreased with the more restrictive operationalizations, the differences between them were minimal.

To this point, we have discussed distress, impairment, and help-seeking as separate constructs. Figure 2 depicts the overlap among distress, impairment, and help-seeking for participants qualifying for any of the AI-SUPERPFP disorders. When reports of "a lot" of either distress or impairment were required, almost one third (32.3%) of participants meeting symptomatic criteria failed to report sufficient distress/impairment or help-seeking. Less than 20% reported all 3, and 16.4% reported help-seeking only. Considering the overlap when participants who reported "some" or "a lot" of distress or impairment were included, many fewer reported no distress, impairment, or help-seeking (8.1%). Here, the largest categories included participants who reported all 3 indicators of CS (42.4%) and those who reported both distress and impairment (35.7%). Thus, the extent of this overlap varied dramatically by the inclusiveness of the responses considered as markers of distress and impairment.

Table 2 also examines the concordance statistics for 4 methods of combining these distress, impairment, and help-seeking measures with varying levels of response inclusiveness, informed by previous work in this area. Thus, CS3 ("a lot" of distress or impairment) approximated the definition found in Slade and Andrews,²³ while CS4 (help-seeking or "a lot" of impairment) was closest to the approaches of Narrow et al¹ and Regier et al.²¹ Operation-

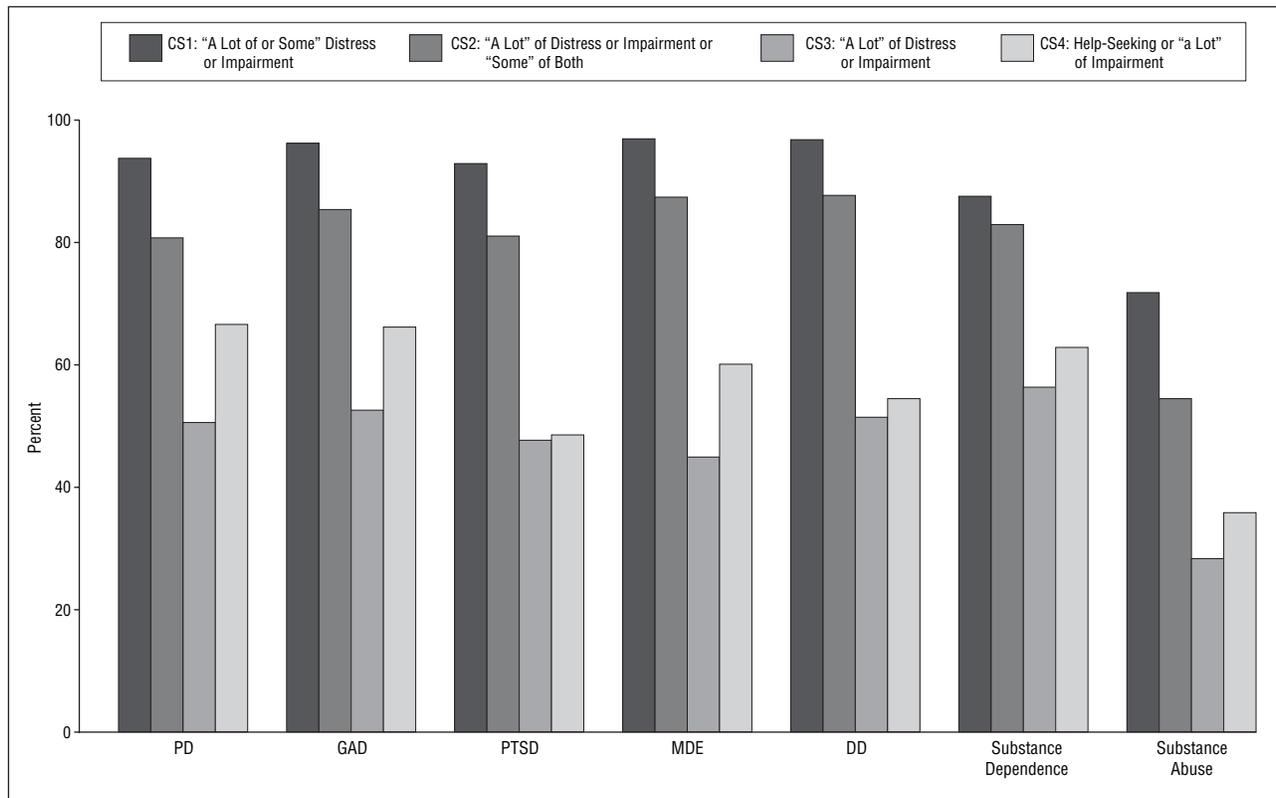


Figure 3. Decrease in lifetime prevalence by operationalization. CS indicates clinical significance; DD, dysthymic disorder; GAD, generalized anxiety disorder; MDE, major depressive episode; PD, panic disorder; and PTSD, posttraumatic stress disorder.

alizations CS1 ("some" distress or impairment) and CS2 ("a lot" of distress or impairment or "some" of both) followed Wakefield and Spitzer's⁵ suggestions and included moderate levels of distress and impairment. Once again, as the operationalizations became more restrictive, the κ values decreased, the sensitivities decreased more dramatically than the specificities increased, the positive predictive values remained stable while the negative predictive values decreased dramatically, the biases increased, and the Global Assessment of Functioning scores remained quite stable. These results indicate that CS was not the source of disagreement between the AI-SUPERPFP CIDI and the SCID.

PREVALENCE ACROSS DIFFERENT OPERATIONALIZATIONS OF CLINICALLY SIGNIFICANT DISTRESS OR IMPAIRMENT

Next we turn to the effect the various 5 operationalizations of CS had on prevalence estimates, focusing first on "any disorder." Operationalization CS3 is the most conservative: the prevalence rate was 50.1% of that when no CS criterion was used (22.8% using CS3 compared with 45.5% with no CS criterion). Similarly, the relative rate for CS4 was 58.0%, 77.4% for CS2, and 88.3% for CS1. When the "any disorder" category was restricted to only diagnoses with an explicit CS criterion in the *DSM-IV* (GAD, PTSD, MDE, and DD), the relative rates were less dramatic and ranged between 71.4% for CS3 and 95.6% for CS1. **Figure 3** demonstrates the relative diminution by disorder across the various operationalizations.

Imposing the CS criterion had the greatest impact for substance abuse. Although the *DSM-IV* implies that CS is less relevant to defining substance dependence and PD, the patterns in these instances were quite similar to those of other disorders. When compared with the secondary analyses of the NCS of Narrow et al⁴ and Regier et al²¹ (see Table 2 of Narrow et al⁴), the AI-SUPERPFP prevalence rates applying CS4 were reduced to a greater extent than was observed in the NCS (PD reduced 35.3% [$100\% - (2.2\%/3.4\%) = 35.3\%$] compared with 22.7%; GAD, 34.5% compared with 17.6%; MDE, 40.0% compared with 36.6%; DD, 45.5% compared with 28.0%; and substance disorders, 50.6% compared with 33.9% for the AI-SUPERPFP and the NCS, respectively).

COMMENT

The concern that psychiatric epidemiological methods, especially those relying on structured protocols administered by lay interviewers, should reflect the clinical expertise inherent in the diagnostic process of the *DSM* was a driving force behind the deliberate inclusion of the CS criterion in the *DSM-IV*. Subsequently, 2 basic questions have informed epidemiological work on CS. First, were the unexpectedly high rates reported by the ECA and the NCS inflated by "false positives," that is, by participants' meeting the symptomatic criterion for disorder but for whom such problems were relatively inconsequential?^{4,7} Second, given that a central goal of studies like the ECA and the NCS was to project treatment need,

did adding CS to the diagnostic algorithm render the estimates more appropriate for mental health policy planning?^{7,20} The data presented here, and our experiences in cross-cultural settings generally, provide additional insight into these questions and further inform the debate about CS in preparation for the *DSM-V*.

The AI-SUPERPFP *DSM-IV* rates of any lifetime disorder were 45.5% when CS was not considered. When the *DSM-III-R* rates were compared directly with those of the NCS, the prevalence of disorders was similar across the 3 samples, although the American Indian lifetime rates were somewhat higher (a range of 2% to 10% based on tribe and sex) than those in the NCS.⁴¹ The next question, then, was whether false positives inflated these rates. Using these methods within this cultural context, false negatives rather than false positives were the major source of discrepancy between lay- and clinician-administered interviews. This finding does not appear to be specific to the AI-SUPERPFP, having also been reported with the ECA data.^{32,38,42} Within the baseline NCS, MDE exhibited significant levels of false positives when compared with the SCID; the bias for most disorders was in the opposite direction, although statistically significant only for simple phobias.²⁷ The recent NCS-Replication appears to have addressed the issue of false positives for lifetime but not current MDE when compared with clinical reappraisals.⁹ Thus, with the exception of the baseline NCS MDE rates, lay interviews appear to underestimate rather than overestimate lifetime disorder when compared with clinical interviews—the reverse of the bias the CS criterion was designed to address. Indeed, Spitzer and Wakefield^{17,20} anticipated that the CS criterion might dramatically increase the false negatives while having little influence on the false positives. These findings support their hypothesis.

Thus, if the SCID were considered a true gold standard, the conclusion that the CS criterion should be ignored would be reasonable, for the AI-SUPERPFP at least. However, we are not yet asserting the primacy of one method of case ascertainment over the other. Rather, we consider both lay- and clinician-administered interviews to be potentially biased. For instance, underreporting may be more common in the lay interviews, while clinicians may be more prone to attribute disorder to “normal” behaviors.^{43,44} Furthermore, biases inherent in the measurement of *DSM*-defined diagnoses in these cultural contexts may differentially affect lay- and clinician-administered interviews. While in-depth investigations of the relative validity of the lay- and clinician-administered protocols will be a focus of our investigative energies in coming years, the current analyses have implications for others using *DSM-IV* definitions of disorder.

As seen in Figure 1, the operationalizations of CS to date have differed considerably. Neither the ECA-DIS nor the UM-CIDI was designed to assess *DSM-IV* disorders; thus, their differences preceded the explicit inclusion of the CS criterion in the *DSM-IV*. At the same time, their probe-flowchart data and the resulting definitions of “probable” psychiatric symptoms drove much of the pre-*DSM-IV* debate about CS. The WHO-CIDI 2.1 was designed for the *DSM-IV* and, in many senses, represents a

compromise between the ECA-DIS and the UM-CIDI approaches with its use of the probe flowchart, mostly at the syndrome level, and with individual items also assessing clinical significance for GAD, PTSD, and MDE. Figure 1 illustrates considerable variation in the measurement of CS; further, the items often do not closely match the *DSM-IV* language. Survey methodologists have consistently demonstrated the large impact of even small differences in wording.⁴⁵ As a limitation to the joint ECA/NCS analyses, Narrow et al⁴ pointed out that the ECA impairment question included “a lot” in the stem whereas NCS did not. Thus, it is unknown whether participants answering “some” to the NCS question might choose “yes” or “no” in the ECA version.⁴ As shown in Table 1 and Table 3, the AI-SUPERPFP data suggest that such differences may be substantial.

Figure 1 also highlights the differential inclusion of help-seeking in the diagnostic calculus across instruments. Table 1 and Figure 2 provide data on the prevalence of help-seeking in the AI-SUPERPFP samples and the overlap among help-seeking, distress, and impairment. Individuals seeking help for their symptoms are often distressed or impaired; however, the inclusion of help-seeking may carry with it an assumption that adequate services are available and known to be efficacious and acceptable to community members.⁴⁶ In preparation for the AI-SUPERPFP, both the ECA-DIS and the UM-CIDI were submitted to focus group review, and concerns were raised about the use of the ECA-DIS probe flowchart with its help-seeking stem question. In particular, our informants suggested that many American Indians with emotional problems have learned from hard experience that local service providers are few in number and often lack the expertise or training to treat such matters. Also, many American Indians who suffer from mental disorders seek treatment from traditional healing sources. Even at this earlier stage in our research, therefore, serious concerns were raised about the use of a help-seeking question as a conditional definition of probable symptoms.

Before concluding, limitations of the current work deserve mention. The samples from which these data were derived limit the inferences drawn. These data were restricted to American Indian participants and, even then, represented only 3 of more than 300 federally recognized American Indian tribes; participants were restricted to members living on or near their reservations and covered a limited age range. Further, the analyses were limited; in particular, the concordance analyses required an assumption that the SCID be considered a gold standard, and thus the estimations of the sensitivities and specificities were likely biased to some degree.⁴⁷ Finally, we did not assess the viability of other constructs such as “harmful dysfunction”⁶ to explain the remaining false positives; an investigation of the cultural definitions of such constructs in American Indian communities is strongly recommended.

Even with these limitations, the current work informs ongoing debates about the CS criterion and other definitions of probable psychiatric symptoms. As others have noted,^{5,17} the lack of consistency with which the CS criterion is applied in the *DSM-IV* is unsettling and may

reflect the ambivalence of the field about this construct. In the absence of biological markers, most diagnoses must superimpose a threshold on dimensions of psychopathology.^{48,49} Further inclusion of thresholds based on disability threatens to make the diagnostic calculus unmanageable and, on the basis of the data reported here, may have limited value. As previously argued,⁵⁰ and as operationalized now in the *International Classification of Diseases, 10th Revision*,²⁵ we suggest the authors of the DSM seriously consider uncoupling assessments of disability from diagnosis, which would serve to emphasize, in a slightly different manner than do CS criteria, that diagnosis should not in itself be equated with medical necessity.^{3,12}

Submitted for Publication: October 20, 2003; final revision received December 30, 2003; accepted April 21, 2004.

Correspondence: Janette Beals, PhD, American Indian and Alaska Native Programs, University of Colorado Health Sciences Center, MS F800, PO Box 6508, Aurora, CO 80045-0508 (jan.beals@uchsc.edu).

Additional Authors/The AI-SUPERPPF Team: Cecelia K. Big Crow, Dedra Buchwald, MD, Buck Chambers, Michelle L. Christensen, PhD, Denise A. Dillard, PhD, Karen DuBray, Paula A. Espinoza, PhD, Candace M. Fleming, PhD, Ann Wilson Frederick, Diana Gurley, PhD, Lori L. Jervis, PhD, Shirlene M. Jim, Carol E. Kaufman, PhD, Ellen M. Keane, Suzell A. Klein, Denise Lee, Monica C. McNulty, Denise L. Middlebrook, PhD, Laurie A. Moore, Tilda D. Nez, Ilena M. Norton, MD, Carlette J. Randall, Angela Sam, James H. Shore, MD, Sylvia G. Simpson, MD, and Lorette L. Yazzie.

Funding/Support: This study was supported by the following grants from the National Institutes of Health (NIH), Bethesda, Md: R01 MH48174 (Dr Manson) and P01 MH42473 (Dr Manson). Manuscript preparation was supported by NIH grants R01 DA14817 (Dr Beals) and R01 AA13420 (Dr Beals).

Acknowledgment: The AI-SUPERPPF would not have been possible without the significant contributions of many people. The following interviewers and computer/data management and administrative staff supplied energy and enthusiasm for an often difficult job: Amelia T. Begay, Cathy A. E. Bell, Mary Cook, Helen J. Curley, Mary C. Davenport, Rhonda Wiegman Dick, Marvine D. Duville, Geneva Emhoolah, Fay Flame, Roslyn Green, Billie K. Greene, Jack Herman, Tamara Holmes, Shelly Hubing, Cameron R. Joe, Louise F. Joe, Cheryl L. Martin, Jeff Miller, Robert H. Moran, Jr, Natalie K. Murphy, Ralph L. Roanhorse, Margo Schwab, PhD, Jennifer Settlementire, Donna M. Shangreaux, Matilda J. Shorty, Selena S. S. Simmons, Jennifer Truel, Lori Trullinger, Jennifer M. Warren, Theresa (Dawn) Wright, Jenny J. Yazzie, and Sheila A. Young. We would also like to acknowledge the contributions of the Methods Advisory Group: Margarita Alegria, PhD, Evelyn J. Bromet, PhD, Dedra Buchwald, MD, Steven G. Heeringa, PhD, Ronald Kessler, PhD, Peter Guarnaccia, PhD, R. Jay Turner, PhD, and William A. Vega, PhD. William E. Narrow, MD, Tim Slade, PhD, and Gavin Andrews, MD, are gratefully acknowledged for excellent suggestions based on a review of the manuscript

before submission. We are also indebted to the ARCHIVES reviewers, whose comments greatly improved the manuscript. Finally, we thank the tribal members who so generously answered all the questions asked of them.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association; 1994.
2. Spitzer RL. Values and assumptions in the development of *DSM-III* and *DSM-III-R*: an insider's perspective and a belated response to Sadler, Hulgus, and Agich's "On values in recent American psychiatric classification." *J Nerv Ment Dis*. 2001; 189:351-359.
3. Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Arch Gen Psychiatry*. 1998;55: 109-115.
4. Narrow WE, Rae DS, Robins LN, Regier DA. Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Arch Gen Psychiatry*. 2002;59:115-123.
5. Wakefield JC, Spitzer RL. Why requiring clinical significance does not solve epidemiology's and *DSM's* validity problem: response to Regier and Narrow. In: Helzer JE, Hudziak JJ, eds. *Defining Psychopathology in the 21st Century: DSM-V and Beyond*. Washington, DC: American Psychiatric Publishing Inc; 2002:31-40.
6. Wakefield JC. Diagnosing *DSM-IV*, I: *DSM-IV* and the concept of disorder. *Behav Res Ther*. 1997;35:633-649.
7. Robins LN, Regier DA. *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. New York, NY: The Free Press; 1991.
8. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. Lifetime and 12-month prevalence of *DSM-III-R* psychiatric disorders in the United States: results from the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994;51:8-19.
9. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003;289:3095-3105.
10. Andrews G, Henderson AS, eds. *Unmet Need in Psychiatry*. Cambridge, England: Cambridge University Press; 2000.
11. Regier DA, Narrow WE, Rupp A, Rae DS, Kaelber CT. The epidemiology of mental disorder treatment need: community estimates of "medical necessity." In: Andrews G, Henderson S, eds. *Unmet Need in Psychiatry: Problems, Resources, Responses*. New York, NY: Cambridge University Press; 2000:41-58.
12. Ford WE. Medical necessity: its impact in managed mental health care. *Psychiatr Serv*. 1998;49:183-184.
13. Cooper B, Singh B. Population research and mental health policy: bridging the gap. *Br J Psychiatry*. 2000;176:407-411.
14. Wakefield JC. *DSM-IV*: are we making diagnostic progress? *Contemp Psychology*. 1996;41:646-652.
15. Frances A. Problems in defining clinical significance in epidemiological studies. *Arch Gen Psychiatry*. 1998;55:119.
16. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition*. Washington, DC: American Psychiatric Association; 1987.
17. Spitzer RL, Wakefield JC. *DSM-IV* diagnostic criterion for clinical significance: does it help solve the false positives problem? *Am J Psychiatry*. 1999;156: 1856-1864.
18. Pincus HA, Zarin DA, First M. "Clinical significance" and *DSM-IV*. *Arch Gen Psychiatry*. 1998;55:1145.
19. Kendler KS. Setting boundaries for psychiatric disorder. *Am J Psychiatry*. 1999; 156:1845-1848.
20. Wakefield JC, Spitzer RL. Lowered estimates—but of what? *Arch Gen Psychiatry*. 2002;59:129-130.
21. Regier DA, Narrow WE. Defining clinically significant psychopathology with epidemiologic data. In: Helzer JE, Hudziak JJ, eds. *Defining Psychopathology in the 21st Century: DSM-V and Beyond*. Washington, DC: American Psychiatric Publishing Inc; 2002:19-30.
22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. Washington, DC: American Psychiatric Association; 1980.
23. Slade T, Andrews G. Empirical impact of *DSM-IV* diagnostic criterion for clinical significance. *J Nerv Ment Dis*. 2002;190:334-337.
24. Andrews G, Henderson S, Hall W. Prevalence, comorbidity, disability and ser-

- vice utilisation: overview of the Australian National Mental Health Survey. *Br J Psychiatry*. 2001;178:145-153.
25. Ustun TB, Chatterji S, Rehm J. Limitations of diagnostic paradigm: it doesn't explain "need." *Arch Gen Psychiatry*. 1998;55:1145-1148.
 26. Andrews G, Peters L. The psychometric properties of the Composite International Diagnostic Interview. *Soc Psychiatry Psychiatr Epidemiol*. 1998;33:80-88.
 27. Kessler RC, Wittchen H-U, Alelson JM, McGonagle KA, Kendler KS, Knauper B, Zhao S. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. *Int J Methods Psychiatr Res*. 1998;7:33-55.
 28. Beals J, Manson SM, Mitchell CM, Spicer P; the AI-SUPERPPF Team. Cultural specificity and comparison in psychiatric epidemiology: walking the tightrope in American Indian research. *Cult Med Psychiatry*. 2003;27:259-289.
 29. Beals J, Manson SM, Shore JH, Friedman M, Ashcraft M, Fairbank JA, Schlenger WE. The prevalence of posttraumatic stress disorder among American Indian Vietnam veterans: disparities and context. *J Trauma Stress*. 2002;15:89-97.
 30. Robins LN, Helzer JE, Croughan J, Ratliff KS. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Arch Gen Psychiatry*. 1981;38:381-389.
 31. Spitzer R, Williams J, Gibbon M. *Structured Clinical Interview for DSM-III-R, Version NP*. New York, NY: New York Psychiatric Institute, Biometrics Research Department; 1987.
 32. Eaton WW, Neufeld K, Chen LS, Cai G. A comparison of self-report and clinical diagnostic interviews for depression: diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore epidemiologic catchment area follow-up. *Arch Gen Psychiatry*. 2000;57:217-222.
 33. *Language SAS* [computer program]. Version 8.2. Cary, NC: SAS Institute; 2001.
 34. *SPSS* [computer program]. Version 11.0. Chicago, Ill: SPSS Inc; 2001.
 35. Kish L. *Survey Sampling*. New York, NY: John Wiley and Sons; 1965.
 36. Cohen J. A coefficient for agreement of nominal scales. *Educ Psychol Meas*. 1960;20:37-46.
 37. Rosner B. *Fundamentals of Biostatistics*. 5th ed. Pacific Grove, Calif: Duxbury; 2000.
 38. Helzer JE, Robins LN, McEvoy LT, Spitznagel EL, Stoltzman RK, Farmer A, Brockington IF. A comparison of clinical and diagnostic interview schedule diagnoses: physician reexamination of lay-interviewed cases in the general population. *Arch Gen Psychiatry*. 1985;42:657-666.
 39. Endicott J, Spitzer RL, Fleiss JL, Cohen J. The global assessment scale: a procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry*. 1976;33:766-771.
 40. Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE. A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychol Med*. 2001;31:1001-1013.
 41. Beals J, Manson SM, Whitesell NR, Spicer P, Novins DK; the AI-SUPERPPF Team. Prevalence of *DSM-IV* disorders and attendant help-seeking in 2 American Indian reservation populations. *Arch Gen Psychiatry*. In press.
 42. Robins LN. Epidemiology: reflection on testing the validity of psychiatric interviews. *Arch Gen Psychiatry*. 1985;42:918-924.
 43. Berkson J. Limitations of the application of fourfold table analyses to hospital data. *Biometric Bull*. 1946;2:47-53.
 44. Cohen P, Cohen J. The clinician's illusion. *Arch Gen Psychiatry*. 1984;41:1178-1182.
 45. Schuman H, Presser S. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, Calif: Sage Publications Inc; 1996.
 46. Rubio-Stipec M, Canino G, Robins LN, Wittchen HU, Sartorius N, Miranda CT. The somatization schedule of the Composite International Diagnostic Interview: the use of the probe chart in 17 different countries. *Int J Methods Psychiatr Res*. 1993;3:129-136.
 47. Buck AA, Gart JJ. Comparison of a screening test and a reference test in epidemiologic studies, I: indices of agreement and their relation to prevalence. *Am J Epidemiol*. 1966;83:586-592.
 48. Kraemer HC, Noda A, O'Hara R. Categorical versus dimensional approaches to diagnosis: methodological challenges. *J Psychiatr Res*. 2004;38:17-25.
 49. Kessler RC. The categorical versus dimensional assessment controversy in the sociology of mental illness. *J Health Soc Behav*. 2002;43:171-188.
 50. Lehman AF. Mental disorders and disability: time to reevaluate the relationship. In: Kupfer DJ, First MB, Regier DA, eds. *A Research Agenda for DSM-V*. Washington, DC: American Psychiatric Association; 2002:201-218.