

An Empirical Study of Different Diagnostic Criteria for Delirium Among Elderly Medical Inpatients

Martin G. Cole, M.D.
Nandini Dendukuri, Ph.D.
Jane McCusker, M.D., Dr.P.H.
Ling Han, M.D., M.Sc.

This study compared the sensitivity and specificity of DSM-IV criteria for delirium with the sensitivity and specificity of DSM-III and ICD-10 criteria among elderly medical inpatients with or without dementia. Secondary objectives were to examine the effect of changing the definition of criterion A on sensitivity and specificity and to compare the sensitivity and specificity of different numbers of symptoms of delirium. A total of 322 elderly patients who had been admitted from the emergency department to the medical services were classified into one of four groups using DSM-III-R criteria: delirium and dementia (n = 128), delirium only (n = 40), dementia only (n = 94), and neither (n = 60). The sensitivity and specificity of DSM-IV, DSM-III, and ICD-10 criteria were determined against DSM-III-R criteria using three definitions of criterion A (clouding of consciousness only, clouding of consciousness and inattention, clouding of consciousness or inattention). When criterion A was defined as clouding of consciousness or inattention, the sensitivity and specificity of DSM-IV, DSM-III, and ICD-10 criteria were 100% and 71%, 96% and 91%, and 61% and 91%, respectively. The results were similar among patients with or without dementia. The lower specificity of DSM-IV was accounted for by its inclusion of patients who did not show

disorganized thinking. DSM-IV criteria for delirium are the most inclusive criteria to date for elderly medical patients with or without dementia.

(The Journal of Neuropsychiatry and Clinical Neurosciences 2003; 15:200–207)

Delirium became a diagnostic entity in American psychiatry only in 1980, when it was formally recognized in DSM-III.¹ The history of the concept before that time and the evolution of diagnostic criteria since then have been reviewed by Tucker² and Liptzin.^{3,4}

DSM-IV diagnostic criteria,⁵ published in 1994, were based on a literature review,⁴ discussion of the limitations of DSM-III-R,⁶ and two data sets, one at the University of Pennsylvania⁷ and the other at Harvard University.⁸ DSM-IV attempted to identify the core symptoms of delirium, operationalize the diagnostic criteria, and account for the presence of dementia. However, the validity (sensitivity and specificity) of the new criteria have yet to be determined among patients with or without dementia.⁴

Because the validity of these criteria has implications for clinical practice and research, the primary objective

From the Department of Psychiatry, St. Mary's Hospital and McGill University (m.g.c.), and the Department of Clinical Epidemiology and Community Studies, St. Mary's Hospital and Department of Epidemiology and Biostatistics, McGill University (N.D., J.M., L.H.), Montreal, Quebec. Address correspondence to Dr. Cole, Department of Psychiatry, St. Mary's Hospital Center, 3830 Lacombe Avenue, Room 2508, Montreal, Quebec H3T 1M5. E-mail: martin.cole@ssss.gouv.qc.ca

Copyright © 2003 American Psychiatric Publishing, Inc.

of this study was to compare the sensitivity and specificity of DSM-IV diagnostic criteria for delirium with the criteria of DSM-III and ICD-10⁹ among elderly medical inpatients with or without dementia. Because there was a major change in the definition of criterion A from DSM-III-R to DSM-IV (inattention versus clouding of consciousness), a secondary objective of this study was to examine the effect that change had on sensitivity and specificity. Another secondary objective was to compare the sensitivity and specificity of different numbers of symptoms of delirium in the same population.

METHOD

Study Design

This study was a secondary analysis of data collected in two concurrent studies on delirium: a randomized controlled trial of management of delirium and a nonexperimental prospective study of the prognosis of delirium that included nondelirious subjects. The study was conducted at St. Mary's Hospital, a 400-bed primary acute care university-affiliated hospital in Montreal. A study nurse was responsible for patient screening and enrollment in the two studies. Only patients aged 65 years and over who were admitted from the emergency department to the medical services were included in the studies. We excluded patients with a primary diagnosis of stroke, those admitted to the oncology unit, those admitted to the intensive care unit or the cardiac monitoring unit unless they were transferred to a medical ward within 48 hours of admission, and those who did not speak English or French.

The study nurse administered the Confusion Assessment Method (CAM)¹⁰ to subjects whose initial Short Portable Mental Status Questionnaire (SPMSQ)¹¹ score was 3 or more or whose nursing notes indicated symptoms of delirium. The study nurse used various sources to complete the CAM—chart, family, and nursing staff—and assessed the patient at several points in time if necessary. Delirium was diagnosed if the patient met DSM-III-R criteria for delirium. Those whose initial SPMSQ score was less than 3 and those whose initial score was 3 or more but who did not meet DSM-III-R criteria for delirium were rescreened with the SPMSQ daily for the following week. The CAM was readministered if the SPMSQ score increased or there was evidence from the nursing notes of symptoms of delirium. Nondelirious subjects were selected from among the patients who were screened for delirium and were found free of this condition. To balance the distributions of age and prior cognitive impairment among patients with and without delirium, the sampling method for nondelirious

patients took into account the patient's age and initial SPMSQ score. Thus, nondelirious subjects were selected from among the patients aged 70 years and over, and only a subsample of patients with SPMSQ scores of less than 3 were included. Prior to enrollment, cognitively intact patients were asked to provide informed consent, and patients with delirium or dementia were asked for their assent to participate in the study and a family member was asked to provide informed consent.

The Delirium Index (DI)¹² and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)¹³ were completed within 24 hours of the diagnosis by a research assistant who had no knowledge of the patient's diagnosis. The research assistant had been trained only to rate the presence and severity of the seven symptoms on the DI at one point in time.

Other measures completed by the study nurse or research assistant included the Clinical Severity of Illness,¹⁴ the Mini-Mental State Examination (MMSE),¹⁵ the Barthel Index,¹⁶ and, for premorbid data, the Instrumental Activities of Daily Living questionnaire from the Older Americans Resources and Services Instrument (OARS).¹⁷

Measures

The SPMSQ is a widely used, observer-rated, 10-item questionnaire that evaluates orientation, memory, and concentration. Scores range from 0 (no impairment) to 10 (severe impairment). The test-retest reliability is reported to be 0.8.¹¹ At a cut-off point of three or more errors, the instrument is reported to have a sensitivity of 0.84 and a specificity of 0.89 in identifying medical inpatients with organic brain syndromes.¹⁸

The CAM is a structured instrument that operationalizes the ten symptoms of delirium specified in DSM-III-R: acute onset, fluctuating course, inattention, disorganized thinking, altered level of consciousness, disorientation, memory impairment, perceptual disturbances, psychomotor agitation or retardation, and sleep/wake disturbance. The CAM diagnosis of delirium was previously validated against the clinical judgment of a psychiatrist and found to have a sensitivity of 97% and a specificity of 92%.¹⁰ In our study, the study nurse's diagnosis of delirium had a sensitivity of 0.89 and a specificity of 1.0 compared with a consensus diagnosis.¹⁹ In an earlier study,²⁰ interrater agreement (kappa) for the individual symptoms of delirium ranged from 0.64 to 1.0 ($n = 11$). In this study, they ranged from 0.28 to 1.0 ($n = 14$). The earlier study included only severe and typical cases of delirium, whereas this study included a broader range of cases, which accounts for the lower kappa values.

The DI is an instrument developed by our group for

the measurement of the severity of symptoms of delirium. It is based solely on observation of the individual patients, without additional information from family members, nursing staff, or the patient's medical chart. The first five questions of the MMSE constitute the minimum basis of observation. Seven of the ten symptoms assessed on the CAM (including disorders of attention, thought, consciousness, orientation, memory, perception, and psychomotor activity but excluding acute onset, fluctuation, and sleep/wake disturbance) are rated on the following impairment scale: 0 = absent, 1 = mild, 2 = moderate, 3 = severe. The total score ranges from 0 (no symptoms) to 21 (maximum severity). Patients who are unresponsive are scored at maximum severity on inattention, disorganized thinking, disorientation, and memory impairment. We established the interrater reliability, concurrent criterion validity, and sensitivity to change of the DI in 30 patients with delirium who were rated simultaneously and independently by one or two research assistants and a geriatric psychiatrist on up to four occasions. The concordance coefficient between DI ratings by psychiatrists and research assistants was 0.88, and between two research assistants it was 0.78. Pearson correlation coefficients were 0.84 between the DI and the Delirium Rating Scale (DRS) and 0.71 between change in the DI and the DRS. The interrater agreement (kappa) for the individual symptoms of delirium in this study ranged from 0.48 to 1.0 ($n = 43$).

The IQCODE assesses the presence of dementia prior to admission on the basis of the responses of an informant who has known the patient for at least 5 years; the score is an average of the 16 item scores, each rated from 1 (much improved) to 5 (much worse). In the original publication,¹³ in which a cutoff of 3.32 was used to identify patients with dementia, the sensitivity was 0.79 and the specificity was 0.82. In a later validation study of a French-language version of the questionnaire conducted in Quebec,²¹ a cutoff of 3.6 had a sensitivity of 0.75 and a specificity of 0.96. We used a cutoff of 3.51 to identify patients with dementia.

The Clinical Severity of Illness was assessed at baseline by the study nurse; scores range from 1 (minimal) to 9 (most severe). The MMSE measures cognitive function on a scale of 0 (poor) to 30 (excellent). The Barthel Index measures independence in personal care activities; we used a modified scoring system²² that ranges from 0 (dependent) to 100 (independent). The OARS Instrumental Activities of Daily Living questionnaire was administered to an informant to assess premorbid function (prior to current illness but not more than 1 month before admission); the scale ranges from 0 (dependent) to 16 (independent).

Statistical Analysis

Patients were classified into four groups according to their delirium and dementia status: both delirium and dementia, delirium only, dementia only, and neither delirium nor dementia. Descriptive statistics of demographic and clinical characteristics and the frequency of symptoms measured by the CAM and the DI were calculated separately for each of these groups.

The accuracy of each set of diagnostic criteria was estimated by two different approaches. In the first, we assumed arbitrarily that the DSM-III-R was the criterion standard and estimated the sensitivity, specificity, positive predictive value, and negative predictive value of DSM-III, DSM-IV, and ICD-10 criteria with respect to it. We also analyzed the data for criterion A—clouding of consciousness with inattention—in three different ways: clouding of consciousness alone; clouding of consciousness and inattention; or clouding of consciousness or inattention. We excluded from the analysis the criterion requiring a putative medical cause of delirium because this criterion was met by all of our elderly medical inpatients and therefore had no discriminant value.

In the second approach, we used a latent class model^{23,24} to simultaneously estimate the prevalence of delirium and the sensitivity and specificity of DSM-III-R and the three options for criterion A in DSM-III, DSM-IV, and ICD-10. Latent class models have been widely used to estimate prevalence and diagnostic accuracy from non-gold standard diagnostic tests. This approach does not require DSM-III-R to be arbitrarily selected as a criterion standard and hence is more realistic, since perfect tests seldom exist in practice.

We analyzed results from pairs of tests—DSM-III-R with each of the other tests—using a latent class random effects model described in Dendukuri and Joseph²⁵ that adjusts for the dependence between tests and uses a Bayesian approach for inference. The Bayesian approach is advantageous because it allows us to explicitly account for prior uncertainty in the accuracy of DSM-III-R in the analysis. Prior information on the sensitivity and specificity of the CAM assessment (using DSM-III-R criteria) for a similar population was available from a previous study by our group.¹⁹ The prior means (and 95% credible intervals) of the sensitivity and specificity of DSM-III-R were 0.83 (0.61, 0.97) and 0.94 (0.79, 0.99), respectively. An exact analytical solution is impossible for this model, so we used a Gibbs sampler, as described in Joseph et al.,²⁶ in which random samples are drawn from the posterior distributions of the parameters by simulation methods. After ensuring convergence, summary statistics of the parameters can be estimated on the basis of these random samples. We ran 10,500 iter-

ations of the Gibbs sampler, the first 500 to ensure convergence and the next 10,000 for inference.

RESULTS

During the study enrollment period, there were 4,085 medical admissions, of which 1,552 (38.0%) were screened for delirium. The reasons for exclusion were: admission to oncology (452), admission to intensive care or coronary care units (377), transfer to long-term care (332), language barrier (301), stroke (289), not sampled or missed (181), refused screening (164), previously enrolled in study (127), transferred or discharged (113), communication problem (105), residence outside geographic area (69), died (20), and other (3). Of the 1,552 patients screened, 187 met DSM-III-R criteria for delirium; 174 nondelirious subjects were also enrolled in the study. We excluded 19 patients with delirium and 20 nondelirious subjects because data on dementia status were missing because of failure to interview an informant, leaving 168 patients with delirium and 154 nondelirious subjects in the study sample.

At enrollment, there were significant differences between the four patient groups with respect to age, gender, living arrangements, and the measures of function and illness severity (Table 1). Patients with delirium

only were more likely to be male and living at home with others. Patients with delirium, with or without dementia, had higher clinical severity and lower Barthel Index scores than the two groups without delirium. Patients with dementia only were more likely to be female.

The frequencies of symptoms of delirium using the CAM or the DI are presented in Tables 2 and 3, respectively. All symptoms of delirium (except memory impairment among patients with dementia) are more frequent among patients with delirium than patients without delirium, although the frequencies and differences in frequencies between groups with or without delirium are smaller when measured by the DI.

The sensitivity and specificity of DSM-III, DSM-IV, and ICD-10 diagnostic criteria (compared with those of DSM-III-R) are similar among patients with or without dementia, although the specificity of DSM-IV is lower (66% vs. 78%) among patients with dementia when criterion A is defined as clouding of consciousness or inattention (Table 4). The sensitivity and specificity of DSM-III, DSM-IV, and ICD-10 criteria in assessing the whole sample are greatly affected by the interpretation of criterion A involving clouding of consciousness and inattention. When clouding of consciousness alone is a required symptom or when both clouding of consciousness and inattention are required symptoms, the sensitivity of the different sets of diagnostic criteria are low.

TABLE 1. Selected characteristics of patients at enrollment, by presence of delirium or dementia

Characteristics	Delirium and Dementia (<i>n</i> = 128)		Delirium Only (<i>n</i> = 40)		Dementia Only (<i>n</i> = 94)		Neither (<i>n</i> = 60)		<i>P</i> ^a
	<i>N</i> or mean	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	
Age (years)									0.064
65–74	15	11.7	13	32.5	5	5.3	5	8.3	
75–84	65	50.8	27	67.5	22	23.4	18	30.0	
85 or older	86	67.2	18	45.0	29	30.9	19	31.7	
Sex									0.004
Female	72	56.3	21	52.5	73	77.7	38	63.3	
Male	56	43.8	19	47.5	21	22.3	22	36.7	
Living arrangements prior to admission ^b									<0.001
Home alone	33	25.8	13	32.5	29	30.9	27	45.0	
Home with others	53	41.4	24	60.0	34	36.2	16	26.7	
Senior residence or foster home	20	15.6	2	5.0	26	27.7	13	21.7	
Nursing home	22	17.2	1	2.5	5	5.3	3	5.0	
Education									0.978
Less than secondary	84	51.2	29	51.8	29	54.7	22	52.4	
Secondary or more	80	48.8	27	48.2	24	45.3	20	47.6	
Age, mean (SD) years	84.3 (6.82)		80.8 (8.2)		84.4 (6.9)		82.3 (6.9)		0.015
MMSE, mean score (SD)	12.6 (7.3)		18.1 (6.2)		18.3 (5.6)		22.2 (5.6)		<0.001
Barthel Index, mean score (SD)	34.8 (28.3)		49.0 (32.5)		49.8 (26.4)		59.1 (27.1)		<0.001
IADL (premorbid), mean score (SD)	5.4 (3.3)		10.4 (2.5)		6.3 (3.3)		9.1 (3.1)		<0.001
Clinical Severity of Illness, mean score (SD)	5.6 (1.4)		5.5 (1.1)		4.2 (1.5)		4.3 (1.3)		<0.001

^a*P* value for chi-square test in the case of categorical variables and for a 2-tailed one-way ANOVA in the case of interval variables. MMSE = Mini-Mental State Examination; IADL = Instrumental Activities of Daily Living questionnaire

^bData on living arrangements missing for one subject without delirium or dementia.

However, when either clouding of consciousness or inattention is a required symptom, the sensitivity is markedly higher.

When the sensitivity and specificity of DSM-III, DSM-IV, and ICD-10 criteria in the whole sample are compared when criterion A is interpreted as requiring either clouding of consciousness or inattention, DSM-IV criteria are the most sensitive (100%) and ICD-10 criteria are the least sensitive (61%); DSM-III and ICD-10 criteria are the most specific (91%) and DSM-IV criteria are the least specific (71%). The relatively low sensitivity of ICD-10 can probably be explained by its requiring five rather than three (DSM-IV) or four (DSM-III) psychopathology criteria for a diagnosis of delirium. The relatively low specificity of DSM-IV can be accounted for by 44 patients diagnosed with delirium according to DSM-IV criteria but not diagnosed with delirium according to DSM-III-R criteria, because of the absence of disorganized thinking. Most of these 44 patients were hypoactive and probably had little verbal production.

The results of latent class analysis for the whole sample are presented in Table 5. The sensitivity and speci-

ficity of DSM-III-R remain quite high even though the test is not perfect. However, DSM-III-R has a lower sensitivity compared with DSM-III and DSM-IV, which are more inclusive. The computed sensitivity and specificity of the remaining tests were very similar to the values presented in Table 4. This is because most subjects with inattention alone, who had delirium according to DSM-III-R but not according to the other standards, were classified as truly having delirium according to the latent class model.

The sensitivity and specificity of the number of symptoms of delirium (as determined by the CAM or the DI), irrespective of the type of symptom, are presented in Table 6. The presence of seven or more symptoms on the CAM yields a sensitivity and specificity of 98% and 76%, respectively, among patients with dementia and a sensitivity and specificity of 95% and 83%, respectively, among patients without dementia. An examination of Table 1 suggests that among these seven or more symptoms, inattention, disorganized thinking, disorientation, memory impairment, fluctuation, and acute onset were those most often present.

TABLE 2. Proportions of patients with symptoms identified by the Confusion Assessment Method in patient groups with or without delirium or dementia

Symptom	Patients Without Dementia		Patients With Dementia	
	With Delirium (n = 40)	Without Delirium (n = 60)	With Delirium (n = 128)	Without Delirium (n = 94)
	%	%	%	%
Inattention	100.0	35.0	100.0	45.7
Disorganized thinking	100.0	1.7	100.0	6.4
Altered consciousness	47.5	15.0	54.7	20.2
Disorientation	100.0	80.0	100.0	92.6
Memory impairment	100.0	80.0	100.0	100.0
Perceptual disturbances	30.0	5.0	25.0	7.4
Psychomotor agitation	65.0	20.0	84.4	30.9
Psychomotor retardation	57.5	21.7	60.9	28.7
Fluctuation	100.0	35.0	100.0	51.1
Altered sleep/wake cycle	72.5	16.7	60.2	18.1
Acute onset	100.0	35.0	100.0	47.9

TABLE 3. Proportions of patients with symptoms identified by the DI in patient groups with or without delirium or dementia

Symptom	Patients Without Dementia		Patients With Dementia	
	With Delirium (n = 40)	Without Delirium (n = 60)	With Delirium (n = 128)	Without Delirium (n = 94)
	%	%	%	%
Inattention	70.0	50.0	80.5	67.0
Disorganized thinking	47.5	13.3	69.5	25.5
Altered consciousness	25.0	8.3	35.9	12.8
Disorientation	72.5	46.7	95.3	79.8
Memory impairment	95.0	81.7	97.7	94.7
Perceptual disturbances	27.5	13.3	21.9	13.8
Psychomotor agitation	12.5	8.3	21.9	10.6
Psychomotor retardation	47.5	15.0	46.9	22.3

TABLE 4. Diagnostic accuracy (%) of DSM-III, DSM-IV, and ICD-10 criteria for delirium against DSM-III-R

Diagnostic Criteria	Patients Without Dementia (<i>n</i> = 100)				Patients With dementia (<i>n</i> = 222)				Whole Sample (<i>n</i> = 322)			
	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
DSM-III, with criterion A:												
Only clouding of consciousness	47.5	98.3	95.0	73.8	53.1	94.7	93.2	59.7	51.8	96.1	93.6	64.6
Both clouding of consciousness and inattention	47.5	98.3	95.0	73.8	53.1	94.7	93.2	59.7	51.8	96.1	93.6	64.6
Either clouding of consciousness or inattention	95.0	95.0	92.7	96.6	96.9	88.3	91.9	95.4	96.4	90.9	92.1	95.9
DSM-IV, with criterion A:												
Only clouding of consciousness	47.5	90.0	76.0	72.0	54.7	84.0	82.4	57.7	53.0	86.4	80.9	62.7
Both clouding of consciousness and inattention	47.5	90.0	76.0	72.0	54.7	85.1	83.3	58.0	53.0	87.0	81.7	62.9
Either clouding of consciousness or inattention	100.0	78.3	75.5	100.0	100.0	66.0	80.0	100.0	100.0	70.8	78.9	100.0
ICD-10, with criterion A:												
Only clouding of consciousness	32.5	96.7	86.7	68.2	37.5	95.7	92.3	52.9	36.3	96.1	91.0	58.0
Both clouding of consciousness and inattention	32.5	96.7	86.7	68.2	37.5	95.7	92.3	52.9	36.3	96.1	91.0	58.0
Either clouding of consciousness or inattention	67.5	91.7	84.4	80.9	58.6	90.4	89.3	61.6	60.7	90.9	87.9	68.0
Confusion Assessment Method	100.0	98.3	97.6	100.0	100.0	96.8	97.7	100.0	100.0	97.4	97.7	100.0

Note: PPV = positive predictive value; NPV = negative predictive value. PPV and NPV are calculated assuming the proportion of DSM-III-R delirium in the sample to be representative of that in the source population.

The presence of five or more symptoms on the DI yields a sensitivity and specificity of 61% and 85%, respectively, among patients with dementia; three or more symptoms yield a sensitivity and specificity of 83% and 63%, respectively, among patients without dementia. An examination of Table 2 suggests that among these five or more symptoms, memory impairment, disorientation, and inattention were those most often present.

DISCUSSION

The primary purpose of this study was to compare the sensitivity and specificity of DSM-IV criteria for delirium with DSM-III and ICD-10 criteria. We found that DSM-IV criteria were more sensitive than those of DSM-III or ICD-10 but less specific, although the lower specificity was accounted for by the inclusion of patients who had been excluded by DSM-III-R criteria because they were hypoactive/nonverbal and did not demonstrate disorganized thinking. Thus, DSM-IV criteria seem to be the most inclusive. This finding applied to elderly medical inpatients with or without dementia.

In addition, our results suggest that criterion A, clouding of consciousness with inattention, is best interpreted as either clouding of consciousness or inattention in DSM-IV, DSM-III, and ICD-10. When the criterion is interpreted as clouding of consciousness alone or both clouding of consciousness and inattention, the sensitivity of the three sets of criteria decreases markedly.

The requirements for seven or more symptoms on the CAM among patients with or without dementia or five or more symptoms on the DI among patients with dementia are probably not useful. However, the requirement for only three or more DI symptoms among patients without dementia suggests that further simplification of the criteria for delirium may be possible in this population.

These results have implications for clinical practice and research. In clinical practice, use of DSM-IV criteria would probably minimize false negatives (particularly among patients with hypoactive symptoms) and encourage treating clinicians to look for possible causes of delirium that might include underlying medical illness or the toxic effects of medications. In research studies where it is important that the study population be as homogeneous as possible, there may be concerns that the use of DSM-IV criteria would lead to the inclusion of false positive cases; however, our examination of the 44 patients diagnosed with delirium according to DSM-IV but not DSM-III-R criteria suggests that the use of DSM-IV criteria identifies only patients with delirium

but is more likely to include patients who are hypoactive, nonverbal, and do not show disorganized thinking.

This study has two strengths. First, the numbers of patients with or without delirium or dementia in the study were relatively large. Second, the symptoms of delirium were rated independently using two different valid and reliable instruments.

Several limitations are worth noting. First, the study was a secondary analysis of data collected for a randomized trial and a prognosis study. Second, our exclusion of patients admitted to the oncology unit or the intensive care unit and those with a diagnosis of stroke (related to the purposes of the two studies on which the analysis is based) may limit the generalizability of the results. Third, our criterion standard, DSM-III-R criteria for delirium, may be controversial; however, the results of the Bayesian analysis that accounted for the imperfect nature of the criterion standard were similar to those of the principal analysis. Fourth, the interrater agreement

(kappa) for some symptoms on the CAM and the DI were low; nonetheless, they were similar to values reported for the Delirium Symptom Interview. Fifth, we used the IQCODE to determine dementia status. The sensitivity and specificity of this instrument are high, but it has not been validated with patients presenting with delirium; nevertheless, the reported demographic and clinical characteristics were consistent with the classification into the four groups. Finally, the prevalence of delirium in this population was 52% (168 of 322 patients). Although this high prevalence is unlikely to have affected sensitivity or specificity, it could be expected to raise the positive predictive value and lower the negative predictive value.

In conclusion, among elderly medical inpatients, DSM-IV criteria (when criterion A is interpreted as either clouding of consciousness or inattention) seem to be the most inclusive set of criteria yet proposed. Moreover, these criteria seem to be equally useful whether or not patients have dementia.

TABLE 5. Diagnostic accuracy (%) of DSM-III, DSM-IV, ICD-10, and DSM-III-R criteria for delirium based on latent class analysis of whole sample ($n = 322$)

Diagnostic Criteria	Sensitivity		Specificity		PPV		NPV	
	Median	95% CI	Median	95% CI	Median	95% CI	Median	95% CI
DSM-III, with criterion A:								
Only clouding of consciousness	54.3	45.4, 73.1	98.2	93.9, 99.9	97.1	89.5, 99.9	66.3	57.3, 85.5
Both clouding of consciousness and inattention	54.6	45.5, 76.4	98.2	94.0, 99.9	97.0	89.4, 99.9	66.7	57.0, 87.5
Either clouding of consciousness or inattention	97.5	93.5, 99.9	96.5	89.5, 99.8	97.1	90.9, 99.9	97.0	92.0, 99.8
DSM-IV, with criterion A:								
Only clouding of consciousness	55.2	46.6, 67.9	92.1	83.6, 99.4	89.7	76.9, 99.2	61.9	50.6, 77.9
Both clouding of consciousness and inattention	55.1	46.5, 68.2	93.1	84.5, 99.6	91.1	77.8, 99.5	61.6	50.3, 78.5
Either clouding of consciousness or inattention	100.0	97.9, 100.0	84.5	69.9, 99.0	90.6	79.3, 99.5	99.3	96.3, 100.0
ICD-10, with criterion A:								
Only clouding of consciousness	37.9	30.0, 49.4	98.1	93.6, 99.9	95.9	85.4, 99.8	57.9	48.5, 72.2
Both clouding of consciousness and inattention	37.9	30.0, 49.5	98.1	93.6, 99.10	95.9	85.4, 99.9	57.9	48.5, 72.3
Either clouding of consciousness or inattention	63.3	54.4, 80.8	95.9	88.9, 99.8	94.7	84.6, 99.8	68.8	58.9, 88.1
DSM-III-R	93.8	87.5, 98.6	95.6	81.5, 99.9	92.5	68.1, 99.8	94.3	87.1, 98.8
Confusion Assessment Method	100.0		97.4		97.7		100.0	

Note: PPV = positive predictive value; NPV = negative predictive value. PPV and NPV are calculated assuming the proportion of DSM-III-R delirium in the sample to be representative of that in the source population.

TABLE 6. Accuracy of Confusion Assessment Method (CAM) or Delirium Index (DI) classification of delirium by number of symptoms of delirium compared with DSM-III-R criteria

Cut-Off Number of Symptoms	Patients Without Dementia ($n = 100$)				Patients With Dementia ($n = 222$)			
	Sensitivity	95% CI	Specificity	95% CI	Sensitivity	95% CI	Specificity	95% CI
CAM								
>2	1.00	0.91, 1.00	0.53	0.40, 0.66	1.00	0.97, 1.00	0.29	0.20, 0.39
>3	1.00	0.91, 1.00	0.60	0.47, 0.72	1.00	0.97, 1.00	0.47	0.36, 0.57
>4	1.00	0.91, 1.00	0.65	0.52, 0.77	1.00	0.91, 1.00	0.51	0.41, 0.62
>5	1.00	0.91, 1.00	0.72	0.59, 0.83	1.00	0.97, 1.00	0.61	0.50, 0.71
>6	0.95	0.83, 0.99	0.83	0.71, 0.92	0.98	0.94, 1.00	0.76	0.66, 0.84
DI								
>2	0.83	0.67, 0.93	0.63	0.50, 0.75	0.91	0.85, 0.96	0.30	0.21, 0.40
>3	0.60	0.43, 0.75	0.82	0.70, 0.90	0.77	0.69, 0.84	0.57	0.47, 0.68
>4	0.43	0.27, 0.59	0.92	0.82, 0.97	0.61	0.52, 0.69	0.85	0.76, 0.92

References

1. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 3rd edition. Washington, DC, American Psychiatric Association, 1980
2. Tucker GJ: The diagnosis of delirium and DSM-IV. *Dement Geriatr Cogn Disord* 99; 10:359–363
3. Liptzin B, Levkoff S, Gottlieb G, et al: Delirium. *J Neuropsychiatry Clin Neurosci* 1993; 5:154–160
4. Liptzin B: What criteria should be used for the diagnosis of delirium? *Dement Geriatr Cogn Disord* 1999; 10:364–367
5. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th edition. Washington, DC, American Psychiatric Association, 1994
6. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 3rd edition, revised. Washington, DC, American Psychiatric Association, 1987
7. Johnson J, Gottlieb G, Sullivan E, et al: Using DSM-III criteria to diagnose delirium in elderly general medical patients. *J Gerontol* 1990; 45:M113–M119
8. Levkoff S, Evans D, Liptzin B, et al: Delirium: the occurrence and persistence of symptoms among elderly hospitalized patients. *Arch Intern Med* 1992; 152:334–340
9. World Health Organization: The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. Geneva, World Health Organization, 1993
10. Inouye S, VanDyck C, Alessi C, et al: Clarifying confusion: the confusion assessment method: a new method for detection of delirium. *Ann Intern Med* 1990; 113:941–948
11. Pfeiffer E: A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc* 1975; 23:433–441
12. McCusker J, Cole M, Bellavance F, et al: Reliability and validity of a new measure of severity of delirium. *Int Psychogeriatr* 1998; 10:421–433
13. Jorm A: A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. *Psychol Med* 1994; 24:145–153
14. Charlson M, Sax F, MacKenzie R, et al: Assessing illness severity: does clinical judgment work? *J Chron Dis* 1986; 39:439–452
15. Folstein MF, Folstein SE, McHugh PR: “Mini-Mental State”: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12:189–198
16. Mahoney FI, Barthel DW: Functional evaluation: the Barthel Index. *MD State Med J* 1965; 14(2):61–65
17. Fillenbaum G: Multidimensional Functional Assessment: The OARS Methodology: A Manual, 2nd edition. Durham, NC, Duke University, Center for the Study of Aging and Human Development, 1978
18. Erkinjuntti T, Sulkava R, Wikstrom J, et al: Short Portable Mental Status Questionnaire as a screening test for dementia and delirium among the elderly. *J Am Geriatr Soc* 1987; 35:412–416
19. Zou Y, Cole M, Primeau F, et al: Detection and diagnosis of delirium in the elderly: psychiatrist diagnosis, Confusion Assessment Method or consensus diagnosis? *Int Psychogeriatr* 1998; 10:303–308
20. Cole M, Primeau F, Bailey R, et al: Systematic intervention for elderly inpatients with delirium: a randomized trial. *Can Med Assoc J* 1994; 151:965–970
21. Law S, Wolfson C: Validation of a French version of an informant-based questionnaire as a screening test for Alzheimer’s disease. *Br J Psychiatr* 1995; 167:541–544
22. Shah S, Vanclay F, Cooper B: Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol* 1989; 42:703–709
23. Faraone S, Tsuang M: Measuring diagnostic accuracy in the absence of a “gold standard.” *Am J Psychiatry* 1994; 151:650–657
24. Walter S, Irwig L: Estimation of test error rates, disease prevalence, and relative risk from misclassified data: a review. *J Clin Epidemiol* 1988; 41:923–937
25. Dendukuri N, Joseph L: Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; 57:158–167
26. Joseph L, Gyorkos T, Coupal L: Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol* 1995; 141:263–272