Detecting Delirium Among Hospitalized Older Patients

Peter Pompei, MD; Marquis Foreman, RN, PhD; Christine K. Cassel, MD; Cathy Alessi, MD; Deon Cox, DO

Background: Delirium occurs commonly among older hospitalized patients and is frequently not recognized. In an effort to identify tools useful to clinicians in the diagnosis of delirium, test characteristics of four screening instruments were compared.

Methods: Patients 65 years of age or older who were admitted to one of four medical and surgical wards of a university teaching hospital were followed up prospectively. Potential subjects were excluded if unavailable for interviews or discharged within 48 hours of admission, or if judged too impaired to participate in the daily interviews. Research assistants administered four instruments used to detect delirium: Digit Span Test, Vigilance 'A' Test, Clinical Assessment of Confusion, and Confusion Assessment Method. Abnormal scores on these tests or suspicion of acute confusion prompted a referral to the clinician-investigators who then assessed the patient daily for delirium based on the *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition* criteria.

Results: Delirium occurred in 64 (14.8%) of 432 subjects. The positive likelihood ratios for all of the instruments were significantly more than 1. The instruments remained useful when applied to selected subgroups: subjects in whom acute mental status changes were documented, subjects on surgical services, and subjects with impaired cognitive status on admission. Combinations of any two instruments did not perform substantially better than the instrument with the best test characteristics: the Clinical Assessment of Confusion. All instruments were more useful at confirming delirium than in excluding it.

Conclusion: The four instruments studied, which are suitable for use at the bedside, can aid the clinician in identifying patients likely to be suffering from delirium.

(Arch Intern Med. 1995;155:301-307)

DELIRIUM OCCURS commonly among acutely ill hospitalized older persons. The reported incidence of this problem ranges from 12% to 80% depending on the population studied and the methods used to detect the syndrome.[1-14] Although some investigators have reported that delirium is often benign,[15-17] others have convincingly shown a strong association between delirium and increased morbidity and mortality,[18,19] longer hospitalizations,[18-20] and an increased intensity of nursing care.[21] In addition, impaired functional status and an increased likelihood of being admitted to a nursing home have been associated with acute confusional states.[4,14,22] Despite the clinically important consequences of delirium, it is unrecognized by physicians and nurses in about half the patients who develop this syndrome.[23-28]

There are several possible explanations for the failure to recognize delirium among older hospitalized patients. By definition, the syndrome is transient and the symptoms vary in intensity. Especially among older persons, the primary behavioral manifestation may be withdrawal rather than agitation. These clinical characteristics coupled with the sometimes brief interactions between patients and medical staff may make detection of the syndrome difficult. In addition, mental status changes may too often be ascribed to dementia, another common syndrome among hospitalized older persons. It may be difficult to recognize new onset of confusion unless a careful assessment of baseline mental status was done, and even then, changes recognized as new may be con-

See Patients and Methods on next page

From The University of Chicago (Ill). Dr Pompei is now with the Stanford (Calif) University School of Medicine.

PATIENTS AND METHODS

POPULATION AND SAMPLE

This was a prospective cohort study of delirium in hospitalized older persons performed at the University of Chicago (Ill) Hospitals between November 1989 and June 1991. Patients who were 65 years of age or older and who were admitted to one of four 24-bed wards were eligible for enrollment. Two of the wards were designated for patients on the general medicine services and two wards were designated for patients on the surgical service, primarily patients with orthopedic, general surgical, urologic, or vascular surgical problems. Informed consent was obtained from all study participants; subjects were excluded if they were unable to provide consent because of cognitive impairment, coma, aphasia, or inability to speak English. Patients were also excluded if they were considered too ill to tolerate the initial 40-minute interview, in protective isolation, discharged within 48 hours of admission, or unavailable for interview within 48 hours of admission.

MEASUREMENTS

Sociodemographic information was collected and assessments of function, cognitive status, and comorbidity were made. Physical function was assessed by the Activities of Daily Living scale[33]; patients who could perform all activities without help were considered to have full function. Cognitive status on admission was measured using the Folstein Mini-Mental State Examination.[34] The lower limit for what was considered a normal score was adjusted for level of education: for subjects with less than a high school education, the cut-off was 21 points; for those with high school experience, the cut-off was 23 points; and for subjects with a college education, the cut-off was 24 points.[35] Comorbidity was estimated by counting the total number of discharge diagnoses listed by medical record abstractors who were blinded to the purpose of the study and the group assignment of the subjects.

Research assistants visited study subjects daily for the first week of hospitalization and then every other day until discharge. The following four tests were administered to subjects at each visit by the research assistants. The Digit Span Test was performed by asking the subject to listen carefully to a series of numbers and then to repeat them. The numbers were presented at a rate of one per second from lists of series of random numbers that were different for each visit. Beginning with a two-number sequence, each correctly repeated series was followed by a sequence with one additional digit. The Digit Span Test results were considered abnormal if the subject could not repeat at least five digits.

The Vigilance 'A' Test was used to measure concentration and sustained attention. This test consists of reading a series of 60 letters among which the letter *a* appears with greater than random frequency. The subject is instructed to indicate to the examiner each time the letter *a* is heard; the specific list of letters is read at a rate of one letter per second. Errors of omission (failure to indicate when the letter *a* is presented) and errors of commission (indication when a letter other than *a* is presented) are counted and summed; more than two errors is considered abnormal.[30]

The CAC was developed to determine the presence, pattern, and severity of confusion in hospitalized adults.[31] The CAC is a checklist of 25 psychomotor behaviors associated with varying degrees of confusion. The presence of more behaviors is associated with more severe confusion. The interviewer is requested to evaluate the patient

sidered expected among persons with dementia.

Another barrier to early detection of delirium is the lack of widely accepted neuropsychologic tests, practical for use at the bedside, to aid the clinician in detecting the syndrome. While the *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition* (*DSM-III-R*) of the American Psychiatric Association[29] criteria are recognized as useful in establishing the diagnosis, tests that are easy to perform and interpret would be more useful to many clinicians. Identifying and disseminating information about reliable and valid methods of case finding may increase the rate of detection of delirium that may, in turn, improve the care and outcomes of hospitalized older persons by reducing complications associated with this syndrome.

The purpose of this study was to examine the test characteristics of four instruments used by clinicians to identify patients with delirium. The four instruments selected for study were: Digit Span Test, Vigilance 'A' Test, Clinical Assessment of Confusion (CAC), and the Confusion Assessment Method (CAM). The Digit Span Test was selected because it is commonly used and has been shown to be a reliable test of attention, an aspect of cognitive function often impaired in delirium.[30] The Vigilance 'A' Test is another useful measure of concentration or sustained attention, and though not widely used by physicians, it is suitable for use at the bedside and has been shown to have few educational, intellectual, or socioeconomic biases.[30] The CAC was developed by nurses and is a checklist of behaviors commonly observed in acute confusional states.[31] The CAM is a standardized instrument, derived from the *DSM-III-R* criteria, that has proven useful in identifying patients with delirium.[32]

RESULTS

During the study 1168 patients 65 years of age and older were admitted to the four study wards. Of these, 278 were not eligible: 109 were so cognitively impaired they could not provide informed consent; 114 patients were discharged within 48 hours of admission; 47 patients could not communicate in English; and eight patients were unavailable to the research assistants because of protective isolation. Of the 890 eligible patients, about half were not enrolled: 306 refused, 107 were unavailable to the research team during the first 48 hours of hospitalization because they were off the wards for extended periods of time, and 45 patients were judged to be too ill to tolerate the initial 40-minute interview. Demographic and clinical characteristics of subjects who were eligible but not

on the basis of the presence or absence of each behavior. The score is the total number of observed behaviors; the presence of five or more of the behaviors was considered abnormal for this study. Foreman[36] has reported the following reliability and validity measures for the CAC test: test-retest reproducibility of 0.85, and interrater agreement (Cohen's κ) of 0.79.

The CAM was developed to assist clinicians without formal psychiatric training to quickly and accurately identify patients with delirium.[32] It was developed from *DSM-III-R* criteria for the syndrome. Patients are examined for the following four characteristics: acute change in mental status with a fluctuating course, inattention, disorganized thinking, and altered level of consciousness. The CAM algorithm for the diagnosis of delirium requires the presence of the first two criteria and either the third or the fourth. This instrument has been shown to have a sensitivity from 94% to 100% and a specificity between 90% and 95%. The interobserver reliability of the CAM is high (κ, 0.81 to 1.0).[32]

In addition to testing and observing the study subjects, research assistants sought evidence of acute confusion by reviewing the medical record and querying the nurses on a daily basis. Written documentation and verbal input from the nurses were used by the research assistants to complete the CAC and CAM instruments; because it was not uniformly available, information from family members and visitors was not routinely pursued. If acute confusion was described or if the subject had an abnormal score on any of the four screening instruments, he or she was referred to one of the clinician investigators for evaluation. The five clinician investigators (four geriatricians and one geriatric nurse specialist) evaluated the subject within 24 hours of referral and made an independent assessment of mental status. The diagnosis of delirium, on admission or at any time during hospitalization, was based on *DSM-III-R* criteria[29]; operational definitions for the individual items were modeled after work by Johnson and colleagues.[11] While the clinician investigators were not blinded to a subject's performance on the screening tests, these results were not used to establish the presence or absence of the individual diagnostic criteria according to the *DSM-III-R*.

Since subjects were tested throughout their hospital stay, a series of scores were available for analysis. For subjects with delirium, the scores from the day delirium was first diagnosed were selected for analysis. If scores for that day were missing, the scores on the day closest to and before the first day of delirium were selected. Scores were available for 61 of the 64 subjects with delirium; 55 subjects (90%) had a score on at least one of the four tests within the 48 hours before the diagnosis of delirium was made. For the subjects without delirium, the median score for each subject was selected and the effects of selecting the worst score and the best score were examined.

ANALYSIS

Data analysis was done using SAS-PC software.[37] Differences in clinical characteristics between subjects with and without delirium were assessed using the *t* test for continuous variables and the χ^2 for categorical variables. Results from the four instruments tested were considered normal or abnormal based on the criteria mentioned above. Sensitivity, specificity, and positive and negative likelihood ratios and their confidence intervals[38] were calculated from 2×2 tables. Further analyses were done to define the test characteristics in selected subgroups of subjects, and to determine whether combinations of any two instruments improved the test characteristics.

enrolled differed from those who enrolled only in that fewer were African American (54% vs 64%, $P<.01$). There were no differences with respect to gender, age, admitting service, or the total number of discharge diagnoses.

Of the 432 subjects enrolled, 263 (61%) were referred to the clinician investigators by the research assistants because of acute changes in mental status. Of these, 64 of the study sample (14.8%) were judged by the clinician investigators to meet the criteria for delirium according to the *DSM-III-R*. None of the subjects had delirium listed as a discharge diagnosis. The clinician investigators also followed up the hospital course of a 20% random sample of patients who were eligible but not enrolled. Although these patients could only be investigated through discussions with nursing staff and daily review of their medical records, 12 of the 80 patients manifested confusion or other changes in mental status consistent with delirium. This rate of 15% is not significantly different from what was observed in the study sample.

Results on all four instruments tested were missing for three subjects with delirium and one subject without delirium. These four subjects have been excluded from our analyses. Clinical characteristics of the subjects with and without delirium are shown in **Table 1**. The distribution of age, gender, and race was not different between the two groups. Subjects who developed delirium were more likely to have had evidence of cognitive impairment at the time of admission, after adjusting for years of education. Only about a third of the subjects with delirium were admitted to a surgical service and about half were fully functional in activities of daily living. The number of discharge diagnoses was greater among subjects with delirium.

Shown in **Table 2** are the sensitivity, specificity, and positive and negative likelihood ratios for the four instruments using the score closest to the day of delirium for the subjects with delirium and the median score for the subjects without delirium. Also shown are the posterior probabilities of delirium for three levels of prior probabilities: 10%, a rate estimate below the reported range of delirium in hospitalized older persons; 25%, moderately high suspicion for the syndrome; and 50%, high level of suspicion for the syndrome. The test characteristics of the two performance-based measures of attention, the Digit Span Test, and the Vigilance 'A' Test, are similar. The CAC performs better than the Vigilance 'A' Test: the point estimate of the likelihood ratio is almost three times higher and the confidence intervals do not overlap.

To examine the stability of the likelihood ratios over the range of scores obtained throughout hospitalization for the subjects without delirium, the analysis was repeated using the best and worst test scores for each individual. For the subjects with delirium, the score closest to and before the day of delirium was retained. These results are shown in **Table 3**. When the worst scores for the subjects without delirium are considered, the two performance-based tests of attention, if abnormal, do not significantly increase the probability of delirium because the positive likelihood ratios are approximately 1. When the worst test scores for the subjects without delirium are used, the usefulness of the two observer-rated instruments is less. However, even in this worst case scenario, abnormal results on these instruments still approximately double the odds of delirium. When the best scores for the subjects without delirium are used, the likelihood ratios approximately double for the two performance-based measures of attention and approximately triple for the two observer-rated instruments compared with the results when median scores are used.

The performance characteristics of the instruments for selected subgroups of the study sample are shown in **Table 4**. For these analyses, the median score for the subjects without delirium was used. First, the group of subjects with delirium was limited to those individuals for whom test results within the 48 hours prior to delirium were available. The positive likelihood ratios were only slightly reduced in this analysis. Second, the group of subjects without delirium was limited to those individuals who had been referred to the clinician investigators by the research assistants because of suspected acute changes in mental status. For this analysis, there is almost no change in the likelihood ratios compared with those of the entire study population. Third, the analysis was limited to subjects on the surgical services. All tests continued to perform well in this subgroup where the incidence of delirium was 10%. The point estimate of the positive likelihood ratio for the CAC was substantially increased, but the precision of the estimate was poor as reflected in the wide confidence intervals. Fourth, the analysis was limited to subjects who had impaired cognitive status at the time of admission. As expected, all tests performed less well in this subgroup of subjects where identifying delirium superimposed on dementia can be extremely difficult. Nevertheless, the positive likelihood ratios for all the instruments remained more than 1.

Finally, we examined the effects of requiring abnormal scores on two tests as an indicator of delirium. For these analyses, the median score for subjects without delirium was used. The positive likelihood ratios were calculated for all combinations of the tested instruments taken two at a time. As shown in **Table 5**, requiring an abnormal result on both the Vigilance 'A' Test and the Digit Span Test raised the positive likelihood ratio of either test taken separately; the test characteristics of this combination approximates that of the best test, CAC considered alone. The test characteristics of the observer-rated instruments were not substantially enhanced by adding a performance-based measure of attention. Similarly, requiring abnormalities on both the CAM and the CAC did not result in a substantial improvement over the CAC alone.

Table 1. Clinical Characteristics of Subjects With and Without Delirium

Clinical Characteristics	Subjects Without Delirium (n=367)	Subjects With Delirium (n=61)
Age, y (mean±SD)	74.2±6.7	74.8±7.3
Percent		
Women	56	56
African Americans	66	62
Education beyond 12th grade*	30	21
Full function in activities of daily living before admission	62	52
Impaired cognition on admission*	32	64
Admitted to a surgical service*	48	32
No. of discharge diagnoses (mean±SD)*	4.2±2.2	6.4±2.3

*P<.05.

Table 2. Sensitivity, Specificity, and Positive and Negative Likelihood Ratios of the Four Instruments*

Test	Sensitivity/Specificity	+Likelihood Ratio (95% CI) −Likelihood Ratio (95% CI)	Posterior Probabilities of Delirium Given a Prior Probability, %		
			10	25	50
Digit Span Test†	0.34/0.90	3.5 (2.2-5.7)	28	54	78
		0.7 (0.6-0.9)	7	19	42
Vigilance 'A' Test‡	0.61/0.77	2.7 (2.0-3.6)	23	47	73
		0.5 (0.4-0.7)	5	14	33
Clinical Assessment of Confusion§	0.36/0.95	7.8 (4.4-13.8)	46	72	89
		0.7 (0.6-0.8)	7	18	40
Confusion Assessment Method§	0.46/0.92	5.4 (3.5-8.4)	37	64	84
		0.6 (0.5-0.7)	6	16	37

*Also shown are the posterior probabilities of delirium given an abnormal result in the screening instrument (+ likelihood ratio) and a normal result on the screening instrument (− likelihood ratio) for illustrative prior probabilities of 10%, 25%, and 50%. CI indicates confidence interval.

†Scores available for 56 subjects with delirium and 363 subjects without delirium.

‡Scores available for 54 subjects with delirium and 367 subjects without delirium.

§Scores available for 61 subjects with delirium and 367 subjects without delirium.

COMMENT

With the ultimate goal of improving the identification of delirium among hospitalized older persons, we characterized the usefulness of four diagnostic tools. The instruments studied were selected based on our judgment regarding their acceptability to subjects, nurses, and physicians in the hospital setting. For all the instruments tested, an abnormal result increases the posterior probability of delirium more than a normal result reduces that probability. Abnormal results on either of the two observer-rated tests substantially increase the posterior probability of delirium. If either the CAC or CAM is used for screening (eg, a prior probability of about 10%), an abnormal result would raise the likelihood of delirium above a threshold warranting further investigation. Conversely, if the clinician had a moderately high index of suspicion (eg, a prior probability of 25%), a normal result would only reduce the likelihood of delirium to the low end of the range of reported rates of this syndrome among hospitalized older persons.

Since these instruments are more likely to be used in patients for whom concerns about mental status changes are raised, it was useful to find that the performance characteristics of the tests were not significantly altered when only the subgroup of subjects referred to the clinician investigators by the research assistants was examined. As expected, all instruments performed less well when applied to subjects with cognitive impairment. We can only speculate that the performance of these instruments would be poor in the 9% of potential subjects excluded because of severe cognitive impairment and inability to provide informed consent. The challenge of identifying delirium superimposed on dementia may better be addressed by examining changes in scores over time rather than comparing single scores with threshold values developed for individuals without cognitive impairment.

In the few instances when requiring abnormal results on two tests simultaneously enhanced the positive likelihood ratio of the better test, it was at the expense of loss of precision as reflected in the widened confidence intervals. Combinations of tests are worth further exploration, especially since some instruments may be preferred by nurses while others may become part of the physicians' assessment. Certainly, given the fluctuating course and multitude of manifestations of this syndrome, improved detection of delirium will require integration of input from everyone involved in caring for the patient.

When screening instruments are administered once daily, there is a risk of missing a transient syndrome like delirium. The observer-rated instruments (CAC and CAM) are more likely to incorporate observations over time and to identify an interval prevalence of delirium, while the performance-based measures (Digit Span Test and Vigilance 'A' Test) measure only point prevalence. This difference may explain, in part, the superior test char-

Table 3. Positive Likelihood Ratios and Sensitivity/Specificity for the Four Instruments Using the Worst, Median, and Best Score for Subjects Without Delirium*

Score Selected for Subjects Without Delirium	Digit Span Test	Vigilance 'A' Test	Clinical Assessment of Confusion	Confusion Assessment Method
Worst score	1.0 (0.7-1.5) [0.34/0.66]	1.4 (1.1-1.8) [0.61/0.56]	2.4 (1.6-3.6) [0.36/0.85]	1.8 (1.3-2.5) [0.46/0.75]
Median score	3.5 (2.2-5.7) [0.34/0.90]	2.7 (2.0-3.6) [0.61/0.77]	7.8 (4.4-13.8) [0.36/0.95]	5.4 (3.5-8.4) [0.46/0.92]
Best score	6.8 (3.8-12.2) [0.34/0.95]	6.6 (4.5-9.7) [0.61/0.90]	26.5 (10-67) [0.36/0.99]	16.8 (8.6-32.9) [0.46/0.97]

**Numbers within parentheses and brackets indicate 95% confidence intervals and sensitivity/specificity, respectively.*

Table 4. Positive Likelihood Ratios and Sensitivity/Specificity for the Four Instruments Applied in Selected Subgroups of the Study Sample*

Subgroup	Digit Span Test	Vigilance 'A' Test	Clinical Assessment of Confusion	Confusion Assessment Method
Subjects with scores within 48 hours of delirium†	2.0 (1.3-3.3) [0.31/0.90]	1.6 (1.2-2.2) [0.57/0.77]	5.1 (2.8-9.4) [0.41/0.95]	3.4 (2.2-5.3) [0.47/0.92]
Subjects referred because of acute changes in mental status‡	3.2 (1.9-5.5) [0.34/0.83]	2.5 (1.8-3.5) [0.61/0.63]	8.8 (5.0-15.5) [0.36/0.93]	5.6 (3.6-8.7) [0.46/0.86]
Subjects admitted to surgical services§	4.8 (2.4-9.9) [0.52/0.91]	3.1 (1.9-5.0) [0.58/0.81]	19.9 (6.7-58.8) [0.45/0.98]	7.2 (3.4-15.3) [0.45/0.94]
Subjects with impaired cognitive status on admission‖	2.3 (1.3-4.1) [0.38/0.83]	1.4 (1.0-2.0) [0.59/0.59]	3.6 (2.0-6.6) [0.44/0.88]	2.6 (1.6-4.1) [0.54/0.79]

**Numbers within parentheses and brackets indicate 95% confidence interval and sensitivity/specificity, respectively.*
†Scores available for 55 subjects with delirium and 367 subjects without delirium.
‡Scores available for 61 subjects with delirium and 199 subjects without delirium.
§Scores available for 20 subjects with delirium and 177 subjects without delirium.
‖Scores available for 39 subjects with delirium and 116 subjects without delirium.

Table 5. Sensitivity, Specificity, and Positive Likelihood Ratios for Selected Combinations of Instruments (Abnormal Results Required on Both Tests)

Combinations of Test Instruments*	Sensitivity/ Specificity	+Likelihood Ratios (95% Confidence Intervals)
Vigilance 'A' Test and Digit Span Test†	0.26/0.97	8.6 (4.1-17.9)
Vigilance 'A' Test and Confusion Assessment Method‡	0.31/0.95	5.8 (3.2-10.3)
Digit Span Test and Clinical Assessment of Confusion§	0.14/0.99	12.9 (4.0-41.5)
Digit Span Test and Confusion Assessment Method§	0.20/0.97	7.1 (3.2-16.0)
Clinical Assessment of Confusion and Confusion Assessment Method‖	0.33/0.97	10.0 (5.2-19.4)

**All combinations use the median scores for the patients without delirium.*

†Scores available for 54 subjects with delirium and 363 subjects without delirium.

‡Scores available for 54 subjects with delirium and 366 subjects without delirium.

§Scores available for 56 subjects with delirium and 362 subjects without delirium.

‖Scores available for 61 subjects with delirium and 367 subjects without delirium.

acteristics of the observer-rated instruments. In this study, there was considerable variability in the likelihood ratios of all instruments calculated from the worst, median, and best scores of the subjects without delirium. Nevertheless, even in the analysis that would bias the results against their usefulness, using the worst score for the subjects without delirium, three of the four instruments still had positive likelihood ratios significantly more than 1.

This study confirmed that delirium occurs commonly among older hospitalized persons and that individuals with significant cognitive impairment are at increased risk for developing the syndrome. It also confirmed that despite the high incidence of delirium, the syndrome is significantly underreported. Confusion was frequently mentioned in the medical record, most often by the nurses, but unless this is specifically recognized as an acute change with an organic cause, perilous delays in treatment of reversible conditions can result.

There are various limitations of this study. The performance characteristics of only four instruments were examined. Certainly, others are available that could have been compared with the tests we used.[39,40] In addition, we limited the analysis to the published guidelines defining normal and abnormal test scores; no attempt was made to maximize the test performance by selecting alternative scores to define an abnormal result. We acknowledge that the performance characteristics of observer-rated instruments such as the CAC and CAM will depend, in part, on how zealously informants of behaviors are pursued. Agreement among observers was not examined.

Several potential sources of bias have been considered. The high incidence of impaired cognition on admission among subjects with delirium raised the possibility of referral bias. However, it is well recognized that dementia is a risk factor for delirium[7,10,12,13]; subjects with delirium would be expected to have a high rate of cognitive impairment. In studies of delirium, precise and accurate diagnosis is always a challenge. We used *DSM-III-R* criteria as applied by experienced geriatricians. Preferably, the clinician investigators should have been blinded to results of the screening tests; however, due to the transient nature of delirium and the inadequate documentation in the medical record of subtle behavioral changes among patients, we found it necessary to provide the geriatricians with as much information as possible for their assessment. Nevertheless, we did require the clinician investigator to perform an independent assessment of each subject referred by the research assistants and to document the specific elements of semistructured *DSM-III-R* criteria that were met if delirium was diagnosed. These safeguards and the fact that only 64 of the 263 subjects referred to the clinician investigators by the research assistants were diagnosed with delirium reduces the likelihood of significant incorporation bias. Since only subjects referred by the research assistants were evaluated by the clinician investigators, there is a possibility of work-up bias. This seems unlikely since the daily screening process was designed to enrich the sample of subjects with delirium, and even so, the majority of subjects (61%) were referred to the clinician investigators. In addition, the rate of delirium in this study sample was within the range reported in other prospective studies of similar populations.[7-13] Nevertheless, if we assume that 5% of the subjects not referred to the clinician investigators experienced delirium, the sensitivities of the instruments would decrease, but the specificities would remain about the same; point estimates of the positive and negative likelihood ratios would all remain within the confidence limits illustrated in Table 2.

Other studies should be done to characterize the clinical usefulness of these and other tests in other settings. In addition, more work needs to be done on the reliability of the instruments and on their validity when other definitions of delirium are used. Although most would be willing to accept the assumption that improved detection of this syndrome is important to the patient, the family, and the clinician, studies that demonstrate the benefits of improved recognition are still needed.

Accepted for publication June 16, 1994.

This study was funded from a grant by the John A. Hartford Foundation Inc, New York, NY.

We thank the patients, nurses, and physicians from the University of Chicago Hospitals who participated in this study. We are grateful to the research assistants who worked tirelessly on this project, to Ruth Ross, PhD, for research and data coordination and management, and to Bong Joo Lee, PhD, for programming and initial data analyses. We also thank Theodore Karrison, PhD, for advice regarding methods and statistics, and Joseph Francis, MD, MPH, and Mark Rudberg, MD, MPH, for their helpful reviews of the manuscript.

Reprints not available.

REFERENCES

1. Cavanaugh S. The prevalence of emotional and cognitive dysfunction in a general medical population: using the MMSE, GHQ, and BDI. *Gen Hosp Psychiatry.* 1983;5:15-24.
2. Chisolm SE, Deniston OL, Ingrisan RM, Barbus AJ. Prevalence of confusion in elderly hospitalized patients. *J Gerontol Nurs.* 1982;8:87-96.
3. Gillick MR, Serrell NA, Gillick LS. Adverse consequences of hospitalization in the elderly. *Soc Sci Med.* 1982;16:1033-1038.
4. Lipowski ZJ. The need to integrate liaison psychiatry and geropsychiatry. *Am J Psychiatry.* 1983;140:1003-1005.
5. Williams MA, Holloway JR, Winn MC, et al. Nursing activities and acute confusional states in elderly hip-fractured patients. *Nurs Res.* 1979;28:25-35.
6. Williams MA, Campbell EB, Raynor WJ Jr, Musholt MA, Mlynarczyk SM, Crane LF. Predictors of acute confusional states in hospitalized elderly patients. *Res Nurs Health.* 1985;8:31-40.
7. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Ann Intern Med.* 1993;119:474-481.
8. Erkinjuntti T, Wikstrom J, Palo J, Autio L. Dementia among medical inpatients: evaluation of 2000 consecutive admissions. *Arch Intern Med.* 1986;146:1923-1926.
9. Cameron DJ, Thomas RI, Mulvihill M, Bronheim H. Delirium: a test of the *Diagnostic and Statistical Manual III* criteria on medical inpatients. *J Am Geriatr Soc.* 1987;35:1007-1010.
10. Francis J, Martin D, Kapoor WN. A prospective study of delirium in hospitalized elderly. *JAMA.* 1990;263:1097-1101.
11. Johnson JC, Gottlieb GL, Sullivan E, et al. Using *DSM-III* criteria to diagnose delirium in elderly general medical patients. *J Gerontol.* 1990;45:M113-119.
12. Rockwood K. Acute confusion in elderly medical patients. *J Am Geriatr Soc.* 1989;37:150-154.
13. Schor JD, Levkoff SE, Lipsitz LA, et al. Risk factors for delirium in hospitalized elderly. *JAMA.* 1992;267:827-831.
14. Levkoff SE, Evans DA, Liptzin B, et al. Delirium: the occurrence and persistence of symptoms among elderly hospitalized patients. *Arch Intern Med.* 1992; 152:334-340.
15. Lipowski ZJ. Transient cognitive disorders (delirium, acute confusional states) in the elderly. *Am J Psychiatry.* 1983;140:1426-1436.
16. Lipowski ZJ. Delirium (acute confusional states). *JAMA.* 1987;258:1789-1792.
17. Russo-Williams P, Urquhart BL, Sharrock NE, Charlson ME. Post-operative delirium: predictors and prognosis in elderly orthopedic patients. *J Am Geriatr Soc.* 1992;40:759-767.
18. Thomas RI, Cameron DJ, Fahs MC. A prospective study of delirium and prolonged hospital stay. *Arch Gen Psychiatry.* 1988;45:937-940.
19. Weddington WW. The mortality of delirium: an underappreciated problem. *Psychosomatics.* 1982;23:1232-1235.
20. Berggren D, Gustafson Y, Eriksson B, et al. Postoperative confusion after anesthesia in elderly patients with femoral neck fractures. *Anesth Analg.* 1987; 66:497-504.
21. Williams MA, Ward SE, Campbell EB. Issues in studying confusion in older hospitalized patients. In: Stinson SM, Kerr JC, Giovanetti P, Field PA, MacPhail J, eds. *New Frontiers in Nursing Research: Proceedings of the International Nursing Research Conference.* Edmonton, Alberta: University of Alberta; 1986: 390. Abstract.
22. Levkoff SE, Besdine RW, Wetle T. Acute confusional states (delirium) in the hospitalized elderly. In: Eisendorfer C, ed. *Annual Review of Gerontology and Geriatrics.* New York, NY: Springer Publishing Co Inc; 1986;6:1-26.
23. Knights EB, Folstein MF. Unsuspected emotional and cognitive disturbance in medical patients. *Ann Intern Med.* 1977;87:723-724.
24. Palmateer LM, McCartney JR. Do nurses know when patients have cognitive deficits? *J Gerontol Nurs.* 1985;11:6-16.
25. Levkoff SE, Safran C, Cleary PD, Gallop J, Phillips RS. Identification of factors associated with the diagnosis of delirium in elderly hospitalized patients. *J Am Geriatr Soc.* 1988;36:1099-1104.
26. Levkoff SE, Besdine RW, Wetle T. Acute confusional states (delirium) in the hosptialized elderly. *Ann Rev Gerontol Geriatr.* 1986;6:1-26.
27. Williams MA, Holloway JR, Winn MC, et al. Nursing activities and acute confusional states in elderly hip-fractures patients. *Nurs Res.* 1979;28:25-35.
28. Francis J, Strong S, Martin D, Kapoor W. Delirium in elderly general medical patients: common but often unrecognized. *Clin Res.* 1988;36:711. Abstract.
29. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition Revised.* Washington, DC: American Psychiatric Association; 1987.
30. Strub RL, Black FW. *The Mental Status Evaluation in Neurology.* 2nd ed. Philadelphia, Pa: FA Davis Co Publishers; 1985:42-43.
31. Vermeersch PEH. The clinical assessment of confusion. *Appl Nurs Res.* 1990; 3:128-133.
32. Inouye SK, Van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. *Ann Intern Med.* 1990;113:941-948.
33. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function. *JAMA.* 1963;185:914-919.
34. Folstein MF, Folstein SE, McHugh PR. The Folstein Mini-Mental State Examination: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12:189-198.
35. Uhlmann RF, Larson EB. Effect of education on the Mini-Mental State Examination as a screening test for dementia. *J Am Geriatr Assoc.* 1991;39:876-880.
36. Foreman MD. A comparison of the reliability and validity of mental status questionnaires in hospitalized patients. *Nurs Res.* 1987;36:216-220.
37. SAS Institute. *SAS Software System, Version 6.03.* Cary, NC: SAS Institute Inc; 1984.
38. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol.* 1991;44:763-770.
39. Albert MS, Levkoff SE, Reilly C, et al. The delirium symptom interview: an interview for the detection of delirium symptoms in hospitalized patients. *J Geriatr Psychiatry Neurol.* 1992;5:14-21.
40. Levkoff S, Liptzin B, Cleary P, Reilly CH, Evans D. Review of research instruments and techniques used to detect delirium. *Int Psychogeriatr.* 1991;3:253-272.