

Early recognition of delirium

Marieke J. Schuurmans

Cover: Long Life by Rince de Jong 1999

Print: Academia Press, Gent , Belgium

Published in the series: Verplegingswetenschappelijke Studies / Studies in Nursing Science

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Early recognition of delirium/ Marieke J. Schuurmans

Thesis University Utrecht – With references – With a summary in Dutch
ISBN 90 – 393 – 2676 – 2

Marieke Schuurmans
University Medical Center Utrecht
Division of Nursing Science
P.O. Box 85060
3508 AB Utrecht
The Netherlands

© 2001 M.J. Schuurmans, Amsterdam, the Netherlands

All rights are reserved. No part of this publication may be reproduced or transmitted in any other form or by any means electronic or mechanical including photocopy recording, or any information storage and retrieval system without permission of the copyright owner.

Early recognition of delirium

Vroegtijdige herkenning van het delirium

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. dr. W.H. Gispen,
ingevolge het besluit van het College van Promoties
in het openbaar te verdedigen op
dinsdag 8 mei 2001 des middags om 4.15 uur

door

Marieke Joanne Schuurmans
geboren op 22 februari 1965 te Bunnik

Promotoren

Prof. dr. L.M. Shortridge-Baggett
Department of Nursing Science
University Medical Center Utrecht

Prof. dr. S.A. Duursma
Department of Geriatric Medicine
University Medical Center Utrecht

The studies described in this thesis were supported by grants from the National Center for Nursing and Care (Landelijk Centrum Verpleging en Verzorging).

The publication of this thesis was financially supported by:

- Stichting Alzheimer Nederland
- Roche Pharmaceuticals
- AstraZeneca

voor mijn ouders
voor Rolf

voor de liefde
het vertrouwen
de zin en het zijn

Contents

Introduction	1
Chapter 1 Early recognition of delirium: Review of the literature	11
Chapter 2 The measurement of delirium: Review of the literature	29
Chapter 3 The Delirium Observation Screening (DOS) Scale: A screening instrument for delirium	47
Chapter 4 The troublesome issue of interrater reliability: A study on nurses' observation of the symptoms of delirium	65
Chapter 5 Recognition of delirium by standard nursing observation	81
Chapter 6 Early symptoms of delirium observed by nurses in patients with a hip fracture	91
Chapter 7 Elderly patients with a hip fracture: the risk for delirium	109
Chapter 8 Low risk for delirium at a geriatric medicine department	125
Chapter 9 General Discussion and Conclusions	135
Summary	145
Samenvatting	151
Dankwoord / Acknowledgement	159
Curriculum Vitae	163

Introduction

Introduction

"They drove me in my bed all through the city. Than they brought me to a small room behind a well-known building, do not know the name right know. The room was number 3. What a coincidence I thought, in the hospital my room number was also 3. Is it just a coincidence or what else might it mean? Tean nurse J. came in. I was so surprised to see him there. I recognized him immediately, even there. I said: "what are you doing here?" He said that he came to help me. I did not believe him and said "no, no, you're lying, you're just like them, you have come to beat me up". Than nurse F. came in. Than I really got scared, if even she was in this plot.... I was so sure they were working for the Mafia. Wanting to steal the diamonds and pearls that were put in my hip during surgery."

Mr. K, 87 years, suffered from delirium after hip surgery, complicated by infection.

Delirium, a transient organic mental syndrome characterized by disturbances in consciousness, cognition and perception, is one of the most frequent forms of psychopathology in elderly patients. Incidences reported vary from 5% (1) to 52% (2), depending on the group of patients, the diagnostic criteria and the method of study. Development of delirium during hospital admission is independently associated with mortality, increased morbidity, functional decline, increased length of stay, intensive nursing care and nursing home placement after admission (3-8).

The term 'delirium' originates from the Greek word 'Lêros', which means 'fools talk' or 'nonsense' and the Latin word 'delirare' or 'delirare decedere', which means 'to go out of the furrow' (9, p.129). The first standardized diagnostic criteria for delirium were published in the Diagnostic Statistical Manual III of the American Psychiatric Association in 1980, based on the consensus of experts (10). Since then the criteria have been revised two times in the DSM-III-R in 1987 (11) and the DSM-IV in 1994. The DSM-IV criteria (12) are described in Table 1. Although the essential features of delirium are clearly described, the presentation of delirium differs from patient to patient (2, 13-15). Symptoms fluctuate and first symptoms occur mostly in the evening and during the night. An early diagnosis of delirium enables the search for diagnosis and treatment for the underlying etiological factor(s), and may prevent consequences of the disturbed behavior (13). The diagnosis of delirium is primarily clinical, based on careful observations and awareness of the diagnostic criteria (16).

Table 1 Criteria of DSM-IV for the diagnosis delirium

A	Disturbance of consciousness with reduced ability to focus, sustain or shift attention
B	Changed cognition or the development of a perceptual disturbance
C	Disturbance develops in a short period of time and fluctuates over the course of the day
D	There is evidence from history, physical examination or laboratory findings that the disturbance is: <ol style="list-style-type: none"> 1 Physiological consequence of general condition 2 Caused by intoxication 3 Caused by medication 4 Caused by more than one etiology

In order to diagnose delirium a relatively recent (hours to days) change of the patients' observable behavior and cognitive functioning need to be brought to the attention of the physician (13). Due to frequent, around the clock, contact with the patient, nurses are in a strategic position to observe changes in behavior of hospitalized elderly patients that indicate the onset of delirium (17-19). The diagnosis of delirium, however, is often overlooked or the symptoms are mistaken for dementia or depression (13, 20-24). Several studies showed that nurses have insufficient knowledge of cognitive deficits and do not recognize delirium (19, 25-27). Nevertheless, studies in which research nurses, who were not involved in regular care, used standard observations, showed promising results with regard to the recognition of delirium (28-30).

To improve early recognition of delirium in elderly hospitalized patients the studies described in this thesis were undertaken. The central question studied was:

- *Is it possible to come to the diagnosis of delirium with a limited number of standard observations by nurses?*

The perspective of these studies is the observation of patients by nursing staff during regular care. First step was to review delirium measurement scales. Most scales depend on testing, rather than on observing a patient, and are developed for use by trained physicians instead of nursing staff. None of the instruments developed for nurses' observation were based on the DSM criteria and due to methodological problems none of them was considered suitable to study the central question of this thesis. Therefore a new instrument, the Delirium

Observation Screening (DOS) Scale, was developed and tested for its psychometric qualities. An additional research question was formulated:

- *Is the Delirium Observation Screening Scale a valid and reliable instrument to detect delirium by nurses' observations during regular care?*

Six studies were performed in order to develop and test the DOS Scale and to answer the central research question. Two studies were prospective in high-risk groups of patients: geriatric patients and hip fracture patients. Data were collected on predisposing factors and precipitating factors of delirium. These data were used to answer an additional question with regard to early recognition of delirium by nurses. This question was formulated since a growing number of studies address the question of predictability of delirium (6, 31-37).

- *Is identification of risk factors useful for early recognition of delirium?*

Early recognition of delirium by nurses could be guided by two principles: the identification of patients at risk and a systematic observation of symptoms. Table 2 presents an overview and the characteristics of the six studies and the corresponding chapters in this thesis.

Table 2 Characteristics of the six studies on early recognition of delirium

Study	Sample (n)		Aim of study	Chapter
	Professionals	Patients		
1	Multidisciplinary experts (7)		Content validity DOS Scale	3
2		Geriatric patients (82)	Psychometric qualities DOS Scale Risk factors for delirium	3 8
3		Geriatric patients (36)	Interrater reliability DOS Scale	4
4	Nurses (44) Nursing experts (5)		Interrater reliability DOS Scale Interrater reliability DOS Scale	4 4
5		Patients referred for psychiatric consultation (57) Nurses (68)	Interrater reliability DOS Scale Psychometric qualities DOS Scale Practical evaluation DOS Scale	4 5 5
6		Hip fracture patients (92)	Psychometric qualities DOS Scale Limitation of number of standard observations Risk factors for delirium	3 6 7

This thesis comprises nine chapters. The chapters 1-8 are written and submitted as articles. Therefore some repetition of information regarding the literature and methods of investigation could not be avoided. The advantage, however, is that all chapters can be read independently.

The first two chapters describe the background of this investigation on delirium.

- **Chapter 1** gives an overview of the literature on early recognition of delirium. Specific emphasis is given to explore the definition and symptoms of delirium, its causes, risk factors, consequences and the role of nurses in early recognition.
- **Chapter 2** presents twelve delirium instruments with regard to their psychometric qualities and their use in clinical practice. Relevant aspects of instrument development in this field are discussed.

The research question: *'Is the Delirium Observation Screening Scale a valid and reliable instrument to detect delirium by nurses' observations during regular care?* will be answered in the chapters 3, 4 & 5.

- **Chapter 3** reports of the development and psychometric testing of the DOS Scale. Experts tested the content validity. Prospectively collected data on geriatric and hip fracture patients were analyzed to study reliability (internal consistency) and validity (criterion and construct) of the DOS Scale.
- **Chapter 4** discusses the interrater reliability of the DOS Scale, tested in three separate studies: in psychiatric consultation patients, in geriatric patients and in a videotaped simulant elderly delirious patient. Methods, results and considerations with regard to interrater reliability in observational scales are discussed.
- **Chapter 5** shows the results of the DOS Scale, tested in a population of patients referred for psychiatric consultation. In this study nurses were invited to evaluate the use of this instrument for their clinical practice. The methods and results of this study are described.

The central question of this thesis: *'Is it possible to come to the diagnosis of delirium with a limited number of standard observations by nurses?'*, will be answered in chapter 6.

- **Chapter 6** is based on the DOS Scale ratings of the hip fracture patients. These ratings are further analyzed in order to come to a limited number of observations that leads to the diagnosis of delirium. An algorithm with a limited number of items of the DOS Scale is developed and tested.

The additional question *'Is identification of risk factors useful for early recognition of delirium?'*, will be answered in the chapters 7 and 8.

- **Chapter 7** gives insight in the risk factors for delirium found in the population of hip fracture patients studied. The occurrence of risk factors will be related to the development of delirium.

- **Chapter 8** has the same perspective as chapter 7 for the group of geriatric patients studied.

The final chapter, **Chapter 9**, is a retrospective view on the findings of all studies, the methodological issues and the implications for clinical practice and research.

References

1. Johnson JC, Gottlieb GL, Sullivan E, Wanich C, Kinosiosian B, Forciea MA, et al. Using the DSM-III criteria to diagnose delirium in elderly general medical patients. *Journal of Gerontology: Medical Sciences* 1990;45(3):M113-119.
2. Williams MA, Campbell EB, Raynor WJ, Musholt MA, Mlynarczyk SM, Crane LF. Predictors of acute confusional states in hospitalized elderly patients. *Research in Nursing and Health* 1985;8:31-40.
3. Cole MG, Primeau FJ. Prognosis of delirium in elderly hospital patients. *Canadian Medical Association Journal* 1993;149(1):41-46.
4. Francis J, Kapoor WN. Prognosis after hospital discharge of older medical patients with delirium. *Journal of the American Geriatric Society* 1992;40:601-606.
5. Brännström B, Gustafson Y, Norberg A, Winblad B. ADL performance and dependency on nursing care in patients with hip fractures and acute confusion in a task allocation care system. *Scandinavian Journal of Caring Sciences* 1991;5(1):3-11.
6. Pompei P, Foreman M, Rudberg MA, Inouye SK, Braund V, Cassel CK. Delirium in hospitalized older persons: outcomes and predictors. *Journal of the American Geriatric Society* 1994;42:809-815.
7. O'Keeffe S, Lavan L. The prognostic significance of delirium in older hospital patients. *Journal of the American Geriatric Society* 1997;45:174-178.
8. Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P. Does delirium contribute to poor hospital outcomes. *Journal of General Internal Medicine* 1998;13:234-242.
9. Godderis JL, Ven Lvd, Wils V. *Handboek geriatrische psychiatrie*. Leuven/Apeldoorn: Garant; 1992.
10. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. Washington: American Psychiatric Association; 1980.
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. Washington D.C.: American Psychiatric Association; 1987.
12. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
13. Lipowski ZJ. *Delirium: acute confusional states*: Oxford University Press; 1990.
14. Rudberg MA, Pompei P, Foreman MD, Ross RE, Cassel CK. The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing* 1997;26:169-174.
15. Sandberg O, Gustafson Y, Brännström B, Bucht G. Clinical profile of delirium in older patients. *Journal of the American Geriatric Society* 1999;47:1300-1306.
16. Inouye SK. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *The American Journal of Medicine* 1994;97:278-288.
17. Foreman MD. Acute confusional states in hospitalized elderly: a research dilemma. *Nursing Research* 1986;35(1):34-38.

18. Yeaw EMJ, Abbate JH. Identification of confusion among the elderly in an acute care setting. *Clinical Nurse Specialist* 1993;7(4):192-197.
19. Morency CR, Levkoff SE, Dick KL. Research considerations delirium in hospitalized elders. *Journal of Gerontological Nursing* 1994;20(8):24-30.
20. Foreman MD. The cognitive and behavioral nature of acute confusional states. *Scholarly Inquiry for Nursing Practice: an International Journal* 1990;5(1):3-20.
21. Rockwood K, Cosway S, Stolee P, Kydd D, Carver D, Jarret P, et al. Increasing the recognition of delirium in elderly patients. *Journal of the American Geriatric Society* 1994;42:252-256.
22. Lewis LM, Miller DK, Morley JE, Nork MJ, Lasater LC. Unrecognized delirium in ED geriatric patients. *American Journal of Emergency Medicine* 1995;13(2):142-145.
23. Farrel K, Ganzini L. Misdiagnosing delirium as depression in medically ill elderly patients. *Archives of Internal Medicine* 1995;155:2459-2464.
24. Johnson J. Identifying and recognizing delirium. *Dementia and Geriatric Cognitive Disorders* 1999;10(5):353-358.
25. Palmateer LM, McCartney JR. Do nurses know when patients have cognitive deficits? *Journal of Gerontological Nursing* 1985;11(2):6-16.
26. Morency CR. Mental status change in the elderly: recognizing and treating delirium. *Journal of Professional Nursing* 1990;6(6):356-365.
27. Treloar AJ, Macdonald AJD. Recognition of cognitive impairment by day and night nursing staff among acute geriatric patients. *Journal of the Royal Society of Medicine* 1995;88:196-198.
28. Williams MA, Ward SE, Campbell EB. Confusion: testing versus observation. *Journal of Gerontological Nursing* 1988;14(1):25-30.
29. Neelon VJ, Champagne MT, McConnel E, Carlson J, Funk SG. Use of the NEECHAM Confusion Scale to assess acute confusional states of hospitalized older patients. In: *Key aspects of elder care; managing falls, incontinence and cognitive impairment*; 1992. p. 278-289.
30. Neelon VJ, Champagne MT, Carlson JR, Funk SG. The NEECHAM Confusion Scale: construction, validation, and clinical testing. *Nursing Research* 1996;45(6):324-330.
31. Schor JD, Levkoff SE, Lipsitz LA, Reilly CH, Cleary PD, Rowe JW, et al. Risk factors for delirium in hospitalized elderly. *JAMA* 1992;267(6):827-831.
32. Williams-Russo P, Urquhart BL, Sharrock NE, Charlson ME. Post-operative delirium: predictors and prognosis in elderly orthopedic patients. *Journal of the American Geriatric Society* 1992;40:759-767.
33. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of Internal Medicine* 1993;119(6):474-480.

34. Inouye SK, Charpentier PA. Precipitating factors for delirium in hospitalized elderly persons; a predictive model and interrelationship with baseline vulnerability. *JAMA* 1996;275(11):852-857.
35. O'Keeffe ST. Predicting delirium in elderly patients: development and validation of a risk-stratification model. *Age and Ageing* 1996;25:317-321.
36. Eden BM, Foreman MD, Sisk R. Delirium: comparison of four predictive models in hospitalized critically ill elderly patients. *Applied Nursing Research* 1998;11(1):27-35.
37. Martin NJ, Stones MJ, Young JE, Bédard M. Development of delirium: a prospective cohort study in a community hospital. *International Psychogeriatrics* 2000;12(1):117-127.

chapter 1

Early recognition of delirium: Review of the literature

Marieke J. Schuurmans
Lillie M. Shortridge-Baggett
Sijmen A. Duursma

This article is an translation and up-date of:
Verpleegkundigen en oude patiënten met een delier. Verpleegkunde, 1998; 14(2): 86-98.
Submitted (current version)

Abstract

Delirium is a transient organic mental syndrome characterized by disturbances in consciousness, cognition or perception. Incidence in elderly hospitalized patients is about 25%. The causes of delirium are multifactorial; risk factors include high age, cognitive impairment and severity of illness. The consequences of delirium include high morbidity and mortality, lengthened hospital stay and nursing home placement. Delirium develops in a short period and symptoms fluctuate, therefore nurses are in a key position to early recognize symptoms. Delirium, however, is often overlooked or misdiagnosed due to lack of knowledge and awareness of nurses and doctors. To improve early recognition of delirium emphasize should be given to terminology, vision and knowledge toward health in aging and delirium as a potential medical emergency and to instruments for systematic screening of symptoms.

Introduction

Delirium, a frequent form of psychopathology in elderly hospitalized patients, is associated with high morbidity and mortality (1-4). Early recognition of delirium enables diagnosis and treatment of the underlying cause and can prevent negative outcomes (5, 6). Many studies on the clinical presentation, etiology, risk factors and consequences of delirium have been published. The contribution of nursing to this body of knowledge is relatively small, though few nurses, Neelon (7-9), Foreman (10-15) and Williams (16-19), have published on this topic over the years.

Notwithstanding the growing number of studies on delirium, the diagnosis is often overlooked or mistaken for dementia or depression (11, 20-23).

The aims of this review are:

1. To describe a definition of delirium, causes , incidence and risk factors and the importance of this diagnosis in elderly patients because of its' consequences.
2. To focus on early recognition of delirium by nurses' observation of symptoms.

Material for this review was drawn from searches conducted on MEDLINE (1990-2000) and CINAHL. The main sources of this review are research articles, however, some review articles were used as well.

Definition of delirium

Delirium is a transient organic mental syndrome of acute onset, characterized by global impairment of cognitive functions, reduced level of consciousness, attention abnormalities, increased or decreased psychomotor activity and a disordered sleep-wake cycle (20). The term 'delirium' originates from the Greek word 'Lêros', which means 'fools talk or nonsense' and the Latin word 'delirare' or 'delirare decedere' which means 'to go out of the furrow' (24, p. 129).

The first standardized diagnostic criteria for delirium were published in the Diagnostic Statistical Manual III (DSM) in 1980 (25), based on the consensus of experts. Until 1980 a wide range of terms was used to describe delirium, the DSM diagnosis introduced consistency that improved communication and facilitated research (20). The semantic problem, however, is still alive. In nursing literature delirium has been defined by a number of terms such as acute confusional state (14, 19, 26), post-operative confusion (27, 28) or intensive care unit psychosis (29, 30). Some authors note that delirium and acute confusion are not synonymous (31, 32).

The DSM criteria for delirium have been adjusted in 1987 (DSM-III-R) and in 1994 (DSM-IV) (33, 34). In the DSM III delirium was classified as an organic psycho-syndrome: a complex of psychiatric symptoms, caused by a somatic disturbance. The term organic psycho-syndrome disappeared in the DSM-IV (34), because it suggests that functional psychiatric disturbances do not have a biological origin (35). Disturbances of the sleep/wake cycle and psychomotor activities were dropped as diagnostic criteria in the DSM-IV since they do not differentiate delirium from other syndromes (25, 36, 37). Table 1 shows the DSM-IV criteria.

Although the essential features of delirium have been clearly described in the DSM criteria, the presentation of delirium differs from patient to patient (38). In the study of Rudberg et al. (39) 15% of hospitalized patients over 65 years of age became delirious. A day by day follow-up showed no pattern of symptoms or pattern of duration of symptoms.

Two different subtypes of delirium have been described (20, 38, 40-42). In hyperactive delirium the patient is overly alert and has an increased psychomotor activity and responsivity to stimuli. A patient with hypoactive delirium has reduced psychomotor activity and alertness. A patient may change from hyperactive to hypoactive in the course of a day or a delirious episode. Sandberg et al. (38) studied 315 delirious patients in hospitals, nursing homes, residential homes and at home. In this sample 22% was classified as hyperactive, 26% as hypoactive, 42% as mixed and 11% neither hyper- nor hypoactive. In a study of 94 delirious patients admitted to an acute geriatric department

(42) findings were comparable: 21% hyperactive delirium, 29% hypoactive delirium, 43% mixed type and 7% no psychomotor disturbances. The type of delirium has been related to the cause (43) and to the prognosis of delirium (42, 44).

Table 1 Diagnostic criteria of DSM-IV for the diagnosis of delirium

A	Disturbance of consciousness with reduced ability to focus, sustain or shift attention
B	Changed cognition or the development of a perceptual disturbance
C	Disturbance develops in a short period of time and fluctuates over the course of the day
D	There is evidence from history, physical examination or laboratory findings that the disturbance is: <ol style="list-style-type: none"> 1 Physiological consequence of general condition 2 Caused by intoxication 3 Caused by medication 4 Caused by more than one etiology

Causes and incidence

In ancient times it was known that a patient could lose his or her mind due to severe fever (20). In the nineteenth century delirium was recognized as originating "from a certain physiological condition of the brain" (20). In 1901 Bonhoeffer (in 20, p. 29) described delirium as an "acute exogenous reaction type", implying that the origin of the disorder is in the body, outside the brain. The pathophysiology, however, is still not clear. A disturbance in oxidative metabolism, causing a change in cholinergic transmission, appears to explain the mechanism (45, 46). The DSM IV describes four etiological categories (see Table 1). Underlying conditions commonly associated with delirium are central nervous system disorders, metabolic disorders, cardiopulmonary disorders and systemic illnesses (47).

Foreman (14) noted that intoxication by medications, infections, dehydration and other metabolic disturbances are the main causes of delirium in elderly hospitalized patients. These causes are confirmed in recent studies. In a study of 94 elderly delirious patients (48) infection was found a major cause (35%), followed by hypoxia-ischemia (32%) and medication toxicity (15%). A study of 171 elderly delirious patients (49) showed infection in 34%. Medication use (42%), metabolic abnormality (22%) and infection (8%) were found

as suspected cause in a study of 64 delirious elderly patients (39). All authors state that the development of delirium is multifactorial.

Incidences of delirium in elderly hospitalized patients vary from 5.0% (50) to 51.5% (17). A percentage of about 25% is found in several studies of medical and surgical patients (3, 45, 51-53). The sample of patients (age, general and medical condition), the method of study (case finding) and the diagnostic criteria used influence the incidence reported.

Risk factors

Some patients are more vulnerable to develop delirium than others are. Inouye et al. (51, 54) have divided risk factors for delirium into predisposing and precipitating factors. Predisposing factors are present at the time of hospital admission and reflect the baseline vulnerability of a patient. Precipitating factors are noxious insults or hospital related factors that contribute to the development of delirium. Patients with a high vulnerability to delirium may develop delirium with any precipitating factor, even if it is mild. But patients with low vulnerability are resistant to the development of delirium even with noxious insults. Several studies have tried to identify predisposing and precipitating factors in elderly patients (2, 51, 52, 54-58). These studies have concluded that the most powerful risks for developing delirium are present before admission. In nearly all studies, pre-existing cognitive impairment was identified as a predisposing factor for delirium. Severe illness, dehydration (51, 56, 57) and high age (52, 55, 57, 58), put patients at risk for developing delirium. Male gender (52, 55), alcohol abuse (2), and comorbidities (2, 57) contributed significantly. Risk factors also include a fracture (52), depression (2) and vision impairment (51). Elie et al. (59) reviewed 27 publications on risk factors for delirium, published between 1966 and 1995. Dementia, advanced age and medical illness were consistent risk factors.

The development of delirium in a study of Inouye and Charpentier (54) was predicted by five precipitating factors: use of physical restraints, malnutrition, more than three medications newly prescribed, use of bladder catheter and any iatrogenic event. Precipitating factors found by Schor et al. (52) were infection and the use of neuroleptics and narcotics. Martin et al. (58) found a high number of medications administered during hospitalization, surgery, a high number of procedures during early hospitalization (e.g., x-rays and blood tests) and intensive care treatment to precipitate delirium. Environmental factors or hospital related factors, such as sleep deprivation, unfamiliar environment, frequent reallocation and lack of windows have also been studied as factors in the development of delirium. These factors, however, did not predict delirium

(10, 58, 60). Psychosocial factors such as stress and lack of social support did not correlate with the occurrence of delirium (10, 58, 60).

Consequences

The consequences of delirium have been studied in several samples of elderly patients (61). The findings show death (2, 4, 62), prolonged stay in the hospital (2, 3), functional decline (3, 4, 63) and new nursing home placement after admission (3, 4) as independent outcomes of delirium. Brännström et al. (1) and Murray et al. (63) found that dependency in activities of daily living is high after a delirious episode, even six months after hospital discharge. Rockwood et al. (64) followed 38 elderly patients with delirium and 148 patients with no delirium or dementia, for a period of 33 months. The follow-up of these patients showed that delirium is an important indicator for dementia and death, even in those patients without prior cognitive or functional impairment.

The experience of being delirious has been studied by Andersson et al. (65) Laitinen (66) and Schofield (67). Laitinen interviewed ten coronary bypass patients who experienced acute confusion during their post operative stay on the ICU. At the time of the interview the patients were in good physical and mental condition. They remembered having been confused and spoke about it spontaneously. They expressed embarrassment when they remembered the experience. Patients described deep loneliness, huge anxiety and fear and helplessness when they found themselves between consciousness and unconsciousness. Patients said that in their stage of confusion small signals or signs from the environment had a huge meaning, good or bad, calming or inducing chaos. The way nurses responded to them had a great impact. Schofield interviewed 19 elderly hospitalized patients after an episode of delirium. Over half of the patients described perceptual disturbances like illusions and hallucinations. Their experiences varied from pleasing and entertaining, to bad and frightening. Some patients described a feeling of being detached from their surroundings. Some experienced memory loss and disorientation. Some patients were reluctant to discuss matters that had been bewildering and were not completely comprehended. Whether they did not remember these things or did not want to talk about them was not clear. In a case study by Andersson et al., the patient said that during the acute confusional state 'she felt as if she had spoken about things that were like fantasies and that she knew that she had mixed various things'.

Early recognition of delirium by nurses

Because of its fluctuating nature and different presentations, the diagnosis of delirium is difficult to make (11, 26). The fluctuating nature limits the sensitivity of once-daily interviews (68). Nurses have frequent contacts with the patient and are in a strategic position to observe changes at an early stage (11, 51, 52, 69, 70). Francis (71) who studied recognition of delirium by physicians and nurses, concluded that nurses may be more likely to recognize delirium because they recognize subtle changes and use information from spouses.

Observation of symptoms by nurses

The knowledge and awareness of nurses regarding symptoms of delirium have been studied from various perspectives (53, 69, 70, 72-77). Palmateer and McCartney (72) and Treloir and McDonald (76) found that because nurses have insufficient knowledge of cognitive deficits, many patients suffering from delirium are not recognized. Morency et al. (70) studied nurses' ability to recognize symptoms of delirium in elderly acute care patients. Each patient was interviewed daily by a trained assistant using the Delirium Symptom Interview (DSI) (78). The day nurses were asked for their own observations and the presence and absence of DSM-III symptoms daily. The two sets of observations were compared. Thirty eight percent of the patients had a delirium on admission or developed delirium during the hospital stay. Of all symptoms of delirium nurses assessed disorientation best (sensitivity compared with DSI: 0.81). Disturbances in sleep-wake cycle, disturbances of consciousness and increased or decreased psychomotor activity were correctly recognized in about two thirds of patients. The nurses only observed fluctuating behavior and speech disturbances in half of patients. Recognition of perceptual disturbances showed a sensitivity of 0.41. The researchers observed that nurses were protective of their patients when questioned about speech disturbances and offered explanations for the symptom. Morency et al. concluded that nurses tend to focus on orientation and fail to recognize the behavioral aspects of delirium. To assess perceptual disturbances it is important to ask the patients' experiences. The authors (p 29) state that 'nurses should be aware of the dangers of normalizing behaviors as lethargy ("simply tired") and tangentiality ("likes to tell stories")'. Yeaw and Abbate (69) interviewed registered nurses after a day shift in which the nurses had cared for a confused, non-demented patient. They also concluded that nurses recognized disorientation best. Rasmussen and Creason (75) found that six of ten interviewed nurses described confusion and disorientation as the same phenomenon. They used a variety of descriptions of signs and symptoms 'unique to each nurse'.

Components of confusion indicated by all nurses were disorientation, memory impairment, altered awareness of own situation and safety needs, aggressive behavior and altered ability to interact. All nurses stressed the importance of knowing the patient in order to be able to recognize small changes in behavior.

Milisen (53) studied the documentation of delirium in nursing records of elderly hip fracture patients. Patients were assessed by trained nurses using the Confusion Assessment Method (CAM) (79) on post operative days. After discharge the patients' nursing records were reviewed for references to the diagnoses of delirium, or synonyms or symptoms of delirium. The overall incidence of delirium determined with the CAM was 20%. Comparison of the CAM results with nursing records showed no documentation of delirium in 87.5% of the delirious patients on the first postoperative day, on the twelfth day this was still the case in 50%. The documentation by nurses showed a great variety of descriptions. Most referred to the level of psychomotor activity and alertness of the patient. The symptom 'altered attention', essential for the diagnosis of delirium, was not documented at all. Conclusion was that documentation was poor, fragmentary and inconsistent, and delirium was underdiagnosed. Souder and O'Sullivan (77) studied the documentation of nurses versus standardized assessment of cognitive status in 42 medically hospitalized patients. The standardized assessment showed impaired performance in 24 to 67% of all measures. Chart review showed no documentation on impaired cognitive status in any of these patients. Souder and O'Sullivan concluded that nurses limit their assessment to orientation and therefore miss many cases of cognitive problems in their patients. Brady (80) studied recognition of confusion by nursing staff in long term care. Disturbing behavior was often described as confusion. Not only do nurses lack knowledge of delirium; studies on recognition of delirium by doctors have shown similar results (22, 71, 81, 82).

Improvement of early recognition of delirium

McCarthy (74) analyzed factors that led to the failure of nurses to detect and manage acute confusional states in elderly hospitalized patients. The study revealed that recognition of acute confusion was influenced by the philosophical orientation of the nurse towards health in aging. Three orientations were identified: (1) decline perspective (regards health in aging as decremental), (2) vulnerable perspective (regards health in aging either positive or negative) and (3) healthful perspective (regards "good health" in aging as normative). Nurses using the healthful perspective were most likely to differentiate between acute and chronic confusion in elderly hospitalized patients. Nurses using the decline perspective were least likely to detect acute confusion. The author suggests that in order to detect acute confusion, nursing students should be

trained to accept and act according to the healthful perspective. Rasmussen and Creason (75) and Morency et al. (70) concluded that nurses should be taught more about normal aging. The fact that cognitive decline is accepted as normal contributes to delay or neglect in the diagnosis of delirium. In patients with known dementia changes in cognition and behavior are seldom seen as important (60, 83).

Another improvement in recognition of delirium is to overcome the semantic problem in the definition of delirium (6). The word 'confusion' is unclear (53, 69, 72, 73, 84). Milisen states 'everyone knows what confusion means, yet no one does, because of the many different definitions among health care workers' (p. 70). Nurses should use behavioral descriptions instead of 'patient is confused' (72, 73).

Since nurses are mainly focussed on disorientation most authors recommend a combination of education and the implementation of standard systematic cognitive tests (5, 53, 72, 73, 77). Standard cognitive tests, however, are influenced by age and education and do not measure all aspects of delirium (85). For successful implementation of delirium screening, according to Williams and Neelon (8, 9, 18) nurses need instruments that are based on observation, used at the bedside during regular care, repetitively and without respondent burden.

Simon et al. (86) describe a process improvement project in an orthopedic unit in which assessment of risk, systematic cognitive screening and observation of behavior, lead to improved detection of delirium. They emphasize the importance of the collaboration between nurses and physicians in this process. A study of Bowler et al. (81) showed, improvement of detection of psychiatric disorders in elderly medical inpatients, when doctors and nurses pooled their observations. Miller et al. (87) successfully implemented standard assessment by use of the NEECHAM Confusion Scale (88). They concluded, however, that nurses gathered information but did not consistently incorporate this information in nursing care and interventions. Lacko et al. (89) studied the effect of staff nurses using a standardized protocol to identify delirium in elderly hospitalized patients. Nurses on the intervention unit were taught about normal changes of aging, the process of delirium and methods for differentiating and diagnosing this condition. A test of orientation, memory and concentration and the CAM, were used by nurses on a daily basis. On the intervention unit 8 out of 32 patients developed delirium, nurses detected and documented all cases correctly. Rapp et al. (90) implemented a protocol with a combination of three instruments: the Mini Mental State Examination (91), the NEECHAM Confusion Scale and the CAM. Nurses were trained in the use of the instruments. At first they reported to be uncomfortable to ask patients about their mental abilities, but after some practice, improved assessment skills, increased recognition and

better communication with other professionals were reported. Nurses were positive about the protocol, however, were concerned about the time screening takes.

Screening instruments

Several instruments have been developed to screen patients for symptoms of delirium. These instruments have been reviewed from different perspectives (14, 90, 92-94). Many instruments have been tested in small groups of patients with relatively modest emphasis on study of reliability and validity. The Clinical Assessment of Confusion A (CAC-A) (31, 95, 96), the Confusion Rating Scale (CRS) (16, 18, 19) and the NEECHAM Confusion Scale (NEECHAM) (8, 9, 87) were developed for use in nursing clinical practice. All were developed based on criteria identified by nurses as symptoms of acute confusion. The CAC-A, a 25 item observational scale showed low sensitivity (0.36) for delirium measured by the DSM-II-R criteria (97). The CRS, a 4 item observational scale, show modest criterion validity and sensitivity. According to the developers use of the CRS in clinical practice should be considered premature (19). The NEECHAM Confusion Scale is a 9 item observation rating scale with good validity and reliability, however, it does not differentiate between chronic and acute cognitive impairment. The Confusion Assessment Method (CAM) (51, 79) was developed for use by trained nurses and physicians. Nevertheless, the CAM showed a sensitivity of 0.13 (98), of 0.68 (21) and of 0.89 (99) when used by nurses. Sensitivity in studies with physicians was excellent (0.81-1.0) (79, 98).

Discussion

Delirium is a serious problem in elderly patients. Knowledge of delirium, presentation, causes and consequences, have increased by the introduction of standardized criteria which facilitated research. In clinical practice use of the standardized criteria should facilitate recognition of symptoms and multidisciplinary communication. Knowledge of causes and risk factors should enable nurses to focus on patients at risk. Findings on outcomes of delirium stress the importance of nurses identifying patients at risk and recognizing delirium symptoms early.

In clinical practice, however, delirium is often overlooked or misdiagnosed. Studies show that nurses lack knowledge and awareness of this problem.

Based on this review several recommendations can be made to change clinical practice with regard to early recognition of delirium.

First, the word confusion should be dropped; it causes misunderstanding in the process of observation and diagnosis. Use of exact behavioral descriptions should replace the word confusion.

Second, emphasis should be given to nurses' view and knowledge with regard to health in elderly patients. As long as nurses view cognitive decline as a normal consequence of aging, and delirium is not understood as a potential medical emergency, there is a world to win. Knowledge about delirium, and its causes and risk factors should be part of basic training of nurses.

Third, instruments for systematic screening of symptoms should be further studied with regard to reliability and validity and especially with regard to feasibility in clinical practice.

Acknowledgement

This review was funded by a grant from the National Center for Nursing and Care (LCVV).

References

1. Brännström B, Gustafson Y, Norberg A, Winblad B. ADL performance and dependency on nursing care in patients with hip fractures and acute confusion in a task allocation care system. *Scandinavian Journal of Caring Sciences* 1991;5(1):3-11.
2. Pompei P, Foreman M, Rudberg MA, Inouye SK, Braund V, Cassel CK. Delirium in hospitalized older persons: outcomes and predictors. *Journal of the American Geriatric Society* 1994;42:809-815.
3. O'Keeffe S, Lavan L. The prognostic significance of delirium in older hospital patients. *Journal of the American Geriatric Society* 1997;45:174-178.
4. Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P. Does delirium contribute to poor hospital outcomes. *Journal of General Internal Medicine* 1998;13:234-242.
5. Bostrom AC. Early identification of confusion in hospitalized elderly patients. *Michigan Nurse* 1992;-(3):11-12.
6. Johnson J. Identifying and recognizing delirium. *Dementia and Geriatric Cognitive Disorders* 1999;10(5):353-358.
7. Neelon VJ. Postoperative confusion. *Critical Care Nursing Clinics of North America* 1990;2(4):579-587.
8. Neelon VJ, Champagne MT, McConnel E, Carlson J, Funk SG. Use of the NEECHAM Confusion Scale to assess acute confusional states of hospitalized older patients. In: *Key aspects of elder care; managing falls, incontinence and cognitive impairment*; 1992; p. 278-289.
9. Neelon VJ, Champagne MT, Carlson JR, Funk SG. The NEECHAM Confusion Scale: construction, validation, and clinical testing. *Nursing Research* 1996;45(6):324-330.
10. Foreman MD. Acute confusional states in hospitalized elderly: a research dilemma. *Nursing Research* 1986;35(1):34-38.
11. Foreman MD. The cognitive and behavioral nature of acute confusional states. *Scholarly Inquiry for Nursing Practice: an International Journal* 1990;5(1):3-20.
12. Foreman MD. Acute confusion in the elderly. *Annual Review of Nursing Research* 1991;11:3-30.
13. Foreman MD, Grabowski R. Diagnostic dilemma: cognitive impairment in the elderly. *Journal of Gerontological Nursing* 1992;18(9):5-12.
14. Foreman MD. Acute confusion in the elderly. *Annual Review of Nursing Research* 1993;11:3-30.
15. Foreman MD, Fletcher K, Mion LC, Simon L. Assessing cognitive function. *Geriatric Nursing* 1996;17(5):228-232.
16. Williams MA, Hooloway JR, Winn MC, Wolanin MO, Lawler ML, Westwick CR, et al. Nursing activities and acute confusional states in hospitalized elderly patients. *Research in Nursing and Health* 1979;28(1):25-35.

17. Williams MA, Campbell EB, Raynor SM, Mlynarczyk SM, Ward SE. Reducing acute confusional states in elderly patients with hipfractures. *Research in Nursing and Health* 1985;8:329-337.
18. Williams MA, Ward SE, Campbell EB. Confusion: testing versus observation. *Journal of Gerontological Nursing* 1988;14(1):25-30.
19. Williams MA. Delirium/Acute Confusional States: evaluation devices in nursing. *International Psychogeriatrics* 1991;3(2):301-308.
20. Lipowski ZJ. *Delirium: acute confusional states*: Oxford University Press; 1990.
21. Rockwood K, Cosway S, Stolee P, Kydd D, Carver D, Jarret P, et al. Increasing the recognition of delirium in elderly patients. *Journal of the American Geriatric Society* 1994;42:252-256.
22. Lewis LM, Miller DK, Morley JE, Nork MJ, Lasater LC. Unrecognized delirium in ED geriatric patients. *American Journal of Emergency Medicine* 1995;13(2):142-145.
23. Farrel K, Ganzini L. Misdiagnosing delirium as depression in medically ill elderly patients. *Archives of Internal Medicine* 1995;155:2459-2464.
24. Godderis JL, Ven Lvd, Wils V. *Handboek geriatrische psychiatrie*. Leuven/Apeldoorn: Garant; 1992.
25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. Washington: American Psychiatric Association; 1980.
26. Rasin JH. Confusion. *Nursing Clinics of North America* 1990;25(4):909-918.
27. Gardner J. Post-operative confusion in elderly hip surgical patients. *The Australian Journal of Advanced Nursing* 1984;1(3):6-10.
28. Platzer H. Post-operative confusion in the elderly - a literature review. *International Journal of Nursing Studies* 1989;26(4):369-379.
29. Geary SM. Intensive care unit psychosis revisited: understanding and managing delirium in the critical care setting. *Critical Care Nursing Quarterly* 1994;17(1):51-63.
30. Granberg A, Engberg IB, Lundberg D. Intensive care syndrome: a literature review. *Intensive Critical Care Nursing* 1996;6(12):173-182.
31. Vermeersch PEH. Response to 'the cognitive and behavioral nature of acute confusional states'. *Scholarly Inquiry for Nursing Practice: an International Journal* 1991;5(1):17-20.
32. Schwarz P, Kooij CHvd. Verpleegkundige richtlijnen voor acuut optredende verwardheid. *Tijdschrift voor Ziekenverpleging* 1992;-(1):9-12.
33. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. Washington D.C.: American Psychiatric Association; 1987.
34. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
35. Hertogh CMPM, Eikelenboom P. Delirium bij ouderen en de DSM-IV: is er nu helderheid over verwardheid? *Tijdschrift voor Gerontologie en Geriatrie* 1995(26):51-56.
36. Tucker GJ. DSM-IV: proposals for revision of diagnostic criteria for delirium. *International Psychogeriatrics* 1991;3(2):197-208.

37. Tucker GJ. The diagnosis of delirium and DSM-IV. *Dementia and Geriatric Cognitive Disorders* 1999;10(5):359-363.
38. Sandberg O, Gustafson Y, Brännström B, Bucht G. Clinical profile of delirium in older patients. *Journal of the American Geriatric Society* 1999;47:1300-1306.
39. Rudberg MA, Pompei P, Foreman MD, Ross RE, Cassel CK. The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing* 1997;26:169-174.
40. Francis J, Martin D, Kapoor WN. A prospective study of delirium in hospitalized elderly. *JAMA* 1990;263(8):1097-1101.
41. Liptzin B, Levkoff SE. An empirical study of delirium subtypes. *British Journal of Psychiatry* 1992(161):843-845.
42. O'Keeffe ST, Lavan JN. Clinical significance of delirium subtypes in older people. *Age and Ageing* 1999;28:115-119.
43. Meagher DJ, O'Hanlon D, O'Mahoney E, Casey PR, Trzepacz PT. Relationship between etiology and phenomenologic profile in delirium. *Journal of Geriatric Psychiatry and Neurology* 1998;11:146-149.
44. O'Keeffe ST. Clinical subtypes of delirium in the elderly. *Dementia and Geriatric Cognitive Disorders* 1999;10:380-385.
45. Francis J. Delirium in older patients. *Journal of the American Geriatric Society* 1992;40:829-838.
46. Eikelenboom P, Rozemuller JM, Muiswinkel FLv. Inflammation and Alzheimer's Disease: relationships between pathogenic mechanisms and clinical expression. *Experimental Neurology* 1998;154:89-98.
47. American Psychiatric Association. Practice guidelines for the treatment of patients with delirium. *The American Journal of Psychiatry* 1999;156(5 (supplement)):1-20.
48. O'Keeffe ST, Lavan J. Causes of delirium in a prospective study of elderly hospital patients. *Age and Ageing* 1997;26(1 (suppl)):P29.
49. George J, Bleasdale S, Singleton SJ. Causes and prognosis of delirium in elderly patients admitted to a district general hospital. *Age and Ageing* 1997;26:423-427.
50. Johnson JC, Gottlieb GL, Sullivan E, Wanich C, Kinosisian B, Forciea MA, et al. Using the DSM-III criteria to diagnose delirium in elderly general medical patients. *Journal of Gerontology: Medical Sciences* 1990;45(3):M113-119.
51. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of Internal Medicine* 1993;119(6):474-480.
52. Schor JD, Levkoff SE, Lipsitz LA, Reilly CH, Cleary PD, Rowe JW, et al. Risk factors for delirium in hospitalized elderly. *JAMA* 1992;267(6):827-831.
53. Milisen K. An intervention study for delirium in elderly hip fracture patients [PhD]. Leuven: Katholic University Leuven; 1999.

54. Inouye SK, Charpentier PA. Precipitating factors for delirium in hospitalized elderly persons; a predictive model and interrelationship with baseline vulnerability. *JAMA* 1996;275(11):852-857.
55. Williams-Russo P, Urquhart BL, Sharrock NE, Charlson ME. Post-operative delirium: predictors and prognosis in elderly orthopedic patients. *Journal of the American Geriatric Society* 1992;40:759-767.
56. O'Keeffe ST. Predicting delirium in elderly patients: development and validation of a risk-stratification model. *Age and Ageing* 1996;25:317-321.
57. Eden BM, Foreman MD, Sisk R. Delirium: comparison of four predictive models in hospitalized critically ill elderly patients. *Applied Nursing Research* 1998;11(1):27-35.
58. Martin NJ, Stones MJ, Young JE, Bédard M. Development of delirium: a prospective cohort study in a community hospital. *International Psychogeriatrics* 2000;12(1):117-127.
59. Elie M, Cole MG, Primeau FJ, Bellavance F. Delirium risk factors in elderly hospitalized patients. *Journal General Internal Medicine* 1998;13:204-212.
60. Inouye SK. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *The American Journal of Medicine* 1994;97:278-288.
61. Cole MG, Primeau FJ. Prognosis of delirium in elderly hospital patients. *Canadian Medical Association Journal* 1993;149(1):41-46.
62. Stevens LE, Moore GMd, Simpsom JM. Delirium in hospital: does it increase length of stay? *Australian and New Zealand Journal of Psychiatry* 1998;32:805-808.
63. Murray AM, Levkoff SE, Wetle TT, Beckett L, Cleary PD, Schor JD, et al. Acute delirium and functional decline in the hospitalized elderly patient. *Journal of Gerontology: Medical Sciences* 1993;48(5):M181-M186.
64. Rockwood K, Cosway S, Carver D, Jarret P, Stadnyk K, Fisk J. The risk of dementia and death after delirium. *Age and Ageing* 1999;28:551-556.
65. Andersson EM, Knutsson IK, Hallberg IR, Norberg A. The experience of being confused: a case study. *Geriatric Nursing* 1993;14(5):242-247.
66. Laitinen H. Patient's experience of confusion in the intensive care unit following cardiac surgery. *Intensive and Critical Care Nursing* 1996;12:79-83.
67. Schofield I. A small exploratory study of the reaction of older people to an episode of delirium. *Journal of Advanced Nursing* 1997;25:942-952.
68. Marcantonio ER, Goldman L, Mangione CM, Ludwig LE, Muraca B, Haslauer CM, et al. A clinical prediction rule for delirium after elective noncardiac surgery. *JAMA* 1994;271(2):134-139.
69. Yeaw EMJ, Abbate JH. Identification of confusion among the elderly in an acute care setting. *Clinical Nurse Specialist* 1993;7(4):192-197.
70. Morency CR, Levkoff SE, Dick KL. Research considerations delirium in hospitalized elders. *Journal of Gerontological Nursing* 1994;20(8):24-30.

71. Francis J. When do physicians and nurses recognize and document delirium? *Journal of the American Geriatric Society* 1992;40 (supplement 10):SA 10.
72. Palmateer LM, McCartney JR. Do nurses know when patients have cognitive deficits? *Journal of Gerontological Nursing* 1985;11(2):6-16.
73. Morency CR. Mental status change in the elderly: recognizing and treating delirium. *Journal of Professional Nursing* 1990;6(6):356-365.
74. McCarthy MC. Interpretation of confusion in the aged: conflicting models of clinical reasoning among nurses [PhD]. San Francisco: University of California; 1991.
75. Rasmussen BH, Creason NS. Nurses' perception of the phenomenon confusion in elderly hospitalized patients. *Vard i Norden* 1991;21(11):5-12.
76. Treloar AJ, Macdonald AJD. Recognition of cognitive impairment by day and night nursing staff among acute geriatric patients. *Journal of the Royal Society of Medicine* 1995;88:196-198.
77. Souder E, O'Sullivan PS. Nursing documentation versus standardized assessments of cognitive status in hospitalized medical patients. *Applied Nursing Research* 2000;13(1):29-36.
78. Albert MS, Levkoff SE, Reilly C, Liptzin B, Pilgrim D, Cleary PD, et al. The Delirium Symptom Interview: an interview for the detection of delirium symptoms in hospitalized patients. *Journal of Geriatric Psychiatry and Neurology* 1992;5(1):14-21.
79. Inouye SK, Dyck CH, Alessi CA, Balkin S, Siegel AP, Horwitz RI. Clarifying confusion: the Confusion Assessment Method. *Annals of Internal Medicine* 1990;113(12):941-948.
80. Brady PF. Labeling of confusion in the elderly. *Journal of Gerontological Nursing* 1987;13(6):29-32.
81. Bowler C, Boyle A, Branford M, Cooper SA, Harper R, Lindesay J. Detection of psychiatric disorders in elderly medical inpatients. *Age and Ageing* 1994;23:307-311.
82. Harwood DMJ, Hope T, Jacoby R. Cognitive impairment in medical inpatients. II: Do physicians miss cognitive impairment? *Age and Ageing* 1997;26:37-39.
83. Inouye SK, Schlesinger MJ, Lydon TJ. Delirium: a symptom of how hospital care is failing older persons and a window to improve quality of hospital care. *The American Journal of Medicine* 1999;106(May):565-571.
84. Simpson CJ. Doctors and nurses use of the word confusion. *British Journal of Psychiatry* 1984;145:441-443.
85. Anthony JC, LeResche L, Niaz U, Von Körf MR, Folstein MF. Limits of the 'Mini-Mental-State' as a screening test for dementia and delirium among hospital patients. *Psychological Medicine* 1982;12:397-408.
86. Simon L, Jewell N, Brokel J. Management of acute delirium in hospitalized elderly: a process improvement project. *Geriatric Nursing* 1997;18(4):150-154.
87. Miller J, Neelon V, Champagne M, Baily D, Ng'andu N, Belyea M, et al. The assessment of acute confusion as part of nursing care. *Applied Nursing Research* 1997;10(3):143-151.

88. Champagne MT, Neelon VJ, McConnel ES, Funk S. The NEECHAM Confusion Scale: assessing acute confusion in hospitalized and nursing home elderly. *The Gerontologist* 1987;27(suppl):4A.
89. Lacko L, Bryan Y, Cellasega C, Salerno F. Changing clinical practice through research. *Clinical Nursing Research* 1999;8(3):235-250.
90. Rapp CG, Wakefield B, Kundrat M, Menten J, Tripp-Reimer T, Culp K, et al. Acute confusion assessment instruments: clinical versus research usability. *Applied Nursing Research* 2000;13(1):37-45.
91. Folstein MG, Folstein SE, McHugh PR. "Mini-Mental State" a practical guide for grading the cognitive state of patients for clinicians. *Journal of Psychiatric Research* 1975;12:189-198.
92. Levkoff S, Liptzin B, Cleary P, Reilly CH, Evans D. Review of research instruments and techniques used to detect delirium. *International Psychogeriatrics* 1991;3(2):253-271.
93. McDougall GJ. A review of screening instruments for assessing cognition and mental status in older adults. *Nurse Practitioner* 1990;15(11):18-28.
94. Smith MJ, Breitbart WS, Platt MM. A critique of instruments and methods to detect, diagnose and rate delirium. *Journal of Pain and Symptom Management* 1995;10(1):35-77.
95. Vermeersch PEH. The Clinical Assessment of Confusion-A. *Applied Nursing Research* 1990;3(3):128-133.
96. Vermeersch P, Henly SJ. Validation of the structure for the "Clinical Assessment of Confusion-A". *Nursing Research* 1997;46(4):208-213.
97. Pompei P, Foreman M, Cassel CK, Alessi C, Cox D. Detecting delirium among hospitalized older patients. *Archives of Internal Medicine* 1995;155(3):301-7.
98. Rolfson DB, McElhaney JE, Jhangri GS, Rockwood K. Validity of the Confusion Assessment Method in detecting postoperative delirium in the elderly. *International Psychogeriatrics* 1999;11(4):431-438.
99. Zou Y, Cole MG, Primeau FJ, McCusker J, Bellavance F, Laplante J. Detection and diagnosis of delirium in the elderly: psychiatrist diagnosis, Confusion Assessment Method, or consensus diagnosis? *International Psychogeriatrics* 1998;10(3):303-308.

chapter 2

The measurement of delirium: Review of scales

Marieke J. Schuurmans

Patricia I. Deschamps

Susan W. Markham

Lillie M. Shortridge-Baggett

Sijmen A. Duursma

Submitted

Abstract

This review describes the characteristics and evaluates the psychometric qualities (process of testing and results) of twelve delirium instruments. Delirium instruments differ in goal (diagnosis, screening symptoms severity), type of data on which the rating is based (observation, interview or test of patients), the rater qualities required, the number of items and the rating time needed. Most instruments are based on the Diagnostic Statistical Manual criteria and measure signs and symptoms as described by these criteria. Reliability of delirium instruments show good to excellent results. Validity of the delirium instruments is overall fair to good. Differences exist, however, in the degree to which reliability and validity were tested and the quality of the testing procedures. Conclusion of this review is that most delirium instruments show promising results but need further testing. Testing is needed in different samples and on a broader range of aspects with regard to reliability and validity. Much emphasis should be given to the procedures used in future studies. Only the CAM, the DRS, the MDAS and the NEECHAM Confusion Scale can be seen as 'ready to use' instruments, since they are well tested in more than one sample with good results. The CAM, DRS and MDAS are physician rated, based on observation, interview and test of a patient. The NEECHAM Confusion Scale is nurses' rated, based on observation of a patient and test of vital functions.

Introduction

Delirium is an acute confusional state that develops over the course of a few hours to days. The major symptoms are disturbances of consciousness and cognition or perception (1). Symptoms fluctuate during the day (2) and presentation of symptoms can differ from patient to patient (3, 4). In elderly patients delirium is a symptom of acute somatic illness. Infections, electrolyte disturbances and dehydration frequently induce delirium (5). More serious conditions such as a heart attack can show in elderly patients no other symptoms but delirium. The incidence of delirium has been found to range from 5.0 % (6) to 51.5 % (7), depending on the population of patients and the method of research. Due to its fluctuating nature and its different presentation delirium is difficult to diagnose. The use of instruments can aid in recognition of signs and symptoms. As Inouye et al. stated in 1990 (8), 'a major reason for the underdiagnosis of and high complication rate associated with delirium has been the absence of a clinically useful measure'.

The development of instruments is hindered by several problems (9):

- changing definitions of delirium over time;
- changing focus on the core symptoms of delirium;
- fluctuation of the symptoms during the day;

- different presentations of delirium: hypoactive, hyperactive, mixed;
- wide variety of conditions that may influence the presentation;
- resemblance to the symptoms of dementia (specially in cases where delirium is superimposed on dementia).

Notwithstanding these facts, several instruments to assess delirium have been developed, described and reviewed in the literature. The reviews have had different goals: to demonstrate the need for a new instrument (9, 10), as an introduction to a test of clinical usability of the instruments (11, 12) or to advice clinicians or researchers on choice of instruments based on the literature (13-15). The number and the type of instruments reviewed are diverse.

The aim of this review is to describe the characteristics and evaluate the psychometric qualities (process of testing and results) of delirium instruments. This description and evaluation can aid students, clinicians and researchers to assign value to the instruments and guide their choice of instrument.

Methods

To detect instruments developed specific for delirium, computerized searches were done on MEDLINE and CINAHL (1986- 2000). References were also traced back from current reports and earlier reviews. Instruments were included in the review when at least one research publication was found. Instruments that were described but not tested, studies in which an instrument was used for data collection but not tested and studies with translated instruments, were not included in the review.

Each instrument will be described with regard to goal (diagnosis, screening severity of symptoms), the diagnostic criteria on which it is based, the type of data on which the rating is based (observation, interview or test of patients), the rater qualities required and the number of items and the rating time needed. The signs and symptoms measured by each instrument will be given.

Psychometric qualities will be reviewed using the Centre for Evidence Based Medicine (EBM) Quality Filters and general measures of reliability and validity.

The Centre for Evidence Based Medicine use the following quality filters for reviewing diagnostic studies [<http://cebm.jr2.ox.ac.uk/>, 10-01-01]:

- clearly identified comparison groups, at least one of which is free of the target disorder or derangement;
- an objective diagnostic standard with demonstrably reproducible criteria for any objectively interpreted component;
- interpretation of the test without knowledge of the diagnostic standard result;

- interpretation of the diagnostic standard without knowledge of the test result;
- an analysis consistent with the study design.

Internal consistency, interrater reliability, stability, sensitivity and specificity are used as measures of reliability. Internal consistency refers to the extent to which all items measure the same concept. Interrater reliability is the equivalence of results when different persons use the instrument on the same subject. Stability refers to the results that are obtained on repeated administrations of the instrument (16). Sensitivity and specificity are measures of the diagnostic accuracy of a test.

Content validity, criterion related validity and construct validity are used as measures of validity. Content validity is the content representativeness or the content relevance of the items of an instrument. In criterion related validity the validation of a new instrument comes from comparing the results of the new instrument with results on a related reliable and valid criterion. Measurement of construct validity examines the underlying attribute of a scale. Based on theoretical predictions, relationships are examined (16). Factor analysis and principal component analysis are methods frequently used in testing construct validity.

Results

Twelve instruments that met the inclusion criteria were found in the literature. Table 1 gives an overview of the main characteristics of these instruments.

Four instruments were developed to diagnose delirium: the Clinical Assessment of Confusion A (CAC-A), the Confusion Assessment Method (CAM), the Cognitive Test for Delirium (CTD), the NEECHAM Confusion Scale (NEECHAM). Five instruments were developed to screen severity of symptoms: the Confusion Rating Scale (CRS), the Confusional State Examination (CSE), the Delirium Index (DI), the Delirium Severity Scale (DSS) and the Memorial Delirium Assessment Scale (MDAS). Three instruments were developed for both goals: the Delirium Assessment Scale (DAS), the Delirium Rating Scale (DRS) and the Delirium Symptom Interview (DSI). Nearly all instruments have been developed based on Diagnostic Statistical Manuals (DSM) criteria (1, 17, 18), except for the CAC-A, CRS and NEECHAM. These instruments are based on criteria, identified by nurses as representing acute confusion.

Eight instruments use a combination of observation, interview and/or tests of the patient. The CAC-A and the CRS are rated based on observation and the CTD and the DSS are rated based on tests. Rater expertise varies from research assistants (CTD, DI, DSI, DSS), nurses (CAC-A, CRS, NEECHAM) to physicians (DAS, MDAS). The CAM, CSE, DRS are described for use by different types of raters.

Table 2- Main characteristics of delirium instruments

Developed by
Vermeersch 1990, 1991, 1997 (19-21)
Inouye 1990, 1993 (8, 22)
Williams 1997, 1998, 1991 (15, 23, 24)
Robertsson 1997, 1999 (10, 25)
Hart et al, 1996, 1997 (26, 27)
O'Keeffe 1994 (28)
McCusker 1998 (29)
Trzepacz, 1988, 1995, 1998, 1999 (30-33)
Albert et al, 1992 (34)
Bettin et al 1998 (35)
(36)
Neelon, Champagne et al 1987, 1991, 1996 (37-39)

The number of items of the instruments varies from 4 to 109, and rating time varies from less than five minutes to fifteen minutes. Of five instruments rating time is not described (CRS, CSE, DAS, DRS, NEECHAM).

The two instruments based on testing a patient showed high percentages (CTD 23%, DSS 21%) of patients who were not able to take the test.

In Table 2, an overview is given of the signs and symptoms described in each version of the DSM criteria and of the signs and symptoms rated by each instrument. Most instruments measure the items described in DSM criteria, nevertheless acute onset (core diagnostic feature of all DSM versions) is measured only by the CAM, the CSE and the DRS. Fluctuating nature, also a core diagnostic feature of all DSM versions, is measured by the same instruments and by the DAS and the DSI. The CAC-A, the CRS and the DRS do not measure changes in consciousness or changes in attention-concentration.

Table 2 Items measured by DSM criteria (1, 17, 18) and by delirium instruments

	Acute onset	Fluctuating nature	Physical disorder	Consciousness	Attention/Concentration	Thinking	Disorientation	Memory	Perception	Psychomotor	Sleep/wake	Mood	Other
DSM-III	X	X	X	X	X	X ^{1, 2}	X	X	X ²	X ²	X ²		
DSM-III-R	X	X		X ²	X	X	X	X ²	X ²	X ²	X ²		
DSM-IV	X	X	X	X	X	X ³	X ³	X ³	X				
CAC-A							X	X	X	X			X ⁴
CAM	X	X	X	X	X	X	X	X	X	X	X		
CRS						X	X		X	X			
CSE	X	X		X	X	X	X	X	X	X	X	X	X ⁵
CTD				X	X		X	X					
DAS		X			X	X	X	X		X			
DI				X	X	X	X	X	X	X			
DRS	X	X	X			X	X	X	X	X	X	X	
DSI		X		X	X	X	X	X	X	X	X	X	
DSS					X	X		X					
MDAS				X	X	X	X	X		X			
NEECHAM					X	X	X	X					X ⁶

¹ described by 'incoherent speech', ² at least two of the following, ³ described by 'changed cognition', ⁴ general behavior, ⁵ intensity of current episode, ⁶ vital functions, oxygen saturation, urinary incontinence

Psychometric qualities

The studies reviewed per instrument are described in Table 3 with regard to sample size, diagnostic standard, concordance with EBM Quality Filters, tests and results of reliability, validity, sensitivity and specificity. A total of 22 studies were reviewed. One study (11) tested two instruments in the same population (CAC-A and CAM). The CAM, the DRS and the NEECHAM are tested in at least three different studies.

The studies were merely done in samples of elderly patients. Samples sizes vary from 19 patients to 791 patients; five studies used sample sizes smaller than 50 patients, seven studies used sample sizes between 51 and 100 patients, four studies used sample sizes between 101 and 150 patients and six studies used sample sizes of over 150 patients. The CAM and the DRS are studied in over 900 patients. The percentage of delirious patients in these studies varies from 5.8% (39) to 100% (25, 29, 39). In three samples the number of delirious patients has not been described (19, 24, 37).

EBM Quality Filters

The EBM Quality Filters give insight into the procedures used in diagnostic studies. For instruments developed to diagnose, clearly identified comparison groups, of which at least one group is not diagnosed delirious, are important. The studies of instruments developed for delirium diagnosis, all meet this criterion. The diagnostic standard used in most studies is the diagnosis of an expert physician using the DSM criteria. This 'gold standard' has limited reliability, reproducibility is questionable and the interpretation of the standard is prone to subjectivity. In some studies therefore consensus diagnoses were used (CAM (40), DRS (41), DSI (34)). Two studies used the CAM as diagnostic standard (DI (29), DSS (35)). This suggests a higher reproducibility although it should be noted that the CAM is known to have a 10% false-positive rate (8). Two studies used a general measure of cognitive decline as diagnostic standard (CAC-A (19), CRS (47)). This diagnostic standard for a test of delirium is of limited value.

The criterion, interpretation of the test without knowledge of the diagnostic standard result and vice versa, was met in three studies (CAM (8), DAS (48), DRS (41)). In most studies either the diagnosis was made with knowledge of the test results, or the test was applied with knowledge of the diagnosis.

The value of tests in which the same person applies the diagnostic standard and the test is limited. This has been the case in studies of the CAC-A (19), the CRS (24) and the NEECHAM (38, 39). This aspect of data collection is not described in tests of the CTD(26, 27) and two studies of the NEECHAM (37, 46). The criterion of analysis consistent with design is met in all studies and will be described further in the section reliability and validity.

Study

Vermeersch,

1990

(19)

Pompei, 1995

(11)

Inouye, 1990 (8)

Rockwood, 1994

(42)

Pompei, 1995

(11)

Zou, 1998

(40)

Rolfson, 1999

(43)

Williams, 1988,

1991 (15, 24)

Robertsson,

1997 (25)

Study

Hart et al 1996,
1997(26, 27)

O'Keeffe
1994(28)

McCusker
1985(29)

Trzepacz 1988,
1995(30, 31)

Rosen 1994(41)

Rockwoord
1996(44)

Albert 1992(34)

Bettin 1998(35)

Study

Breitbart

1997(36)

Lawlor 2000(45)

Champagne

1987(37)

Nellon 1992(38)

Neelon 1996(39)

Csokasy 1999

(46)

Reliability

Internal consistency was examined in half of the delirium instruments (CSE, CTD, DRS, DSI, MDAS, NEECHAM). The CTD and the NEECHAM have very good internal consistency, as measured by a Cronbach's alpha > 0.80 . The internal consistency of the NEECHAM is consistent in different samples. The DRS and the MDAS show both excellent in one study (Cr. $\alpha \geq 0.90$), and considerable less reliability in another study (Cr. α 0.40 and 0.78). The CSE and the DSI show moderate to good (Cr. $\alpha \geq 0.40$ and ≤ 0.80) internal consistency.

Interrater reliability was tested in all but one of the instruments, the CTD. In most cases two nurses, physicians or research assistants assessed a patient at the same time. During this assessment, one of them did the interview and test and the other observed. Interrater reliability was measured by agreement in a percentage or a reliability coefficient (Cohen's Kappa, Intra Class Correlation Coefficient or correlation coefficient). In the study of the DSS (35) interrater reliability was tested with the use of videotapes. A research assistant rated the patient. An expert then rated the videotape of the assessment. The ratings of the research assistant and of the expert were compared to determine interrater reliability. Interrater reliability was good to excellent (K, ICC or $r > 0.80$) in almost all instruments. Interrater reliability of the CSE, the DAS and the DI was rated on the level of symptoms and showed a range of from modest to good interrater reliability.

The stability of an instrument for delirium is not testable with delirious patients, since a main characteristic of delirium is the fluctuation of symptoms. Ratings will be influenced by change over time. For this reason, in most cases researchers did not test for stability. Two instruments, the CTD and NEECHAM, were tested on two occasions in stable demented patients. In these patients, test-retest reliability was excellent (0.90 - 0.98).

Sensitivity/Specificity

For instruments developed to diagnose, discriminating between patients with or without the disease or abnormality is vital. In most studies the sensitivity and specificity of delirium instruments was based on the diagnosis of a physician (DSM criteria). As mentioned earlier, this gold standard is not an objective measure. The number of delirious patients used to calculate sensitivity and specificity varied from 10 to 71. The sensitivity of the CAC-A is low. The sensitivity of the CAM shows a broad range (0.13 -1.0). A sensitivity of 0.13 (43), 0.68 (42) and 0.89 (40) were found in studies where the CAM was used by nurses. A sensitivity of 0.46 was found in a study where the CAM was used by research assistants (11). Sensitivity in studies with physicians was excellent (8, 43). The sensitivity of the CTD is excellent (95-100%), of the DAS a little less (83-88%), of the DSI (90%) and of the NEECHAM (95%) very good. For most diagnostic instruments the specificity is good (> 0.80). Excellent rates (> 0.95) are less frequent, perhaps influenced by ratings of demented patients, who suffer from

disturbances of cognition, which overlap with symptoms of delirium. Differentiation between these two diagnoses is the most complex task of an instrument designed to diagnose.

For instruments developed to follow the course of delirium sensitivity to change is important. The CRS, the CSE, the DI and the DSS were used as repeated measures. The CRS was correlated with ratings on the Short Portable Mental Status Questionnaire (SPMSQ) (49), on three days. The correlation between the two measures on the three testing times was low (0.27 - 0.51) (24). The CSE had a sensitivity to change of 0.75 compared with the Clinical Global Impression Scale of Improvement. The DSS had a sensitivity for change of - 0.44 compared with expert ratings.

Validity

Establishing content validity, by the use of experts, was explicitly described in studies of five instruments (CAM, CSE, DRS, NEECHAM). The content validity of the DRS was not tested in the initial study but in a later study (44).

Criterion related validity was tested in all delirium instruments, except the DSI. Four categories of related criteria were used: expert diagnoses of delirium, cognitive tests (in most cases the Mini Mental State Evaluation (MMSE) (50) or the SPMSQ, other delirium instruments and tests developed for specific symptoms.

Limited testing (one criterion) for concurrent validity was done in the CAC-A, the CRS and DSS. Other instruments have been tested by at least two related criteria. Criterion related validity generally shows acceptable correlation (≥ 0.60) but in none of the instruments excellent correlation (≥ 0.90). The absence of a correlation between the CTD and the nurses DRS (- 0.02) is explained by differences in the method and domain of measurement of the instruments (26).

The Mini-Mental State Examination (MMSE) (50) has been used with most instruments to test criterion validity (CAM, CSE, CTD, DAS, DI, DRS, MDAS and NEECHAM). The MMSE is a valid and reliable measure to test cognitive functioning in general. Questions are grouped in seven categories: orientation to time, orientation to place, registration of three words, attention and calculation, recall of three words, language and visual construction. Good correlation between a measurement for delirium and the MMSE is expected since symptoms measured overlap. One third of the delirium instruments (CAM, DAS, DI and DRS) are rated based on a clinical interview combined with the MMSE. The MMSE scores, however, are known to be influenced by age, education and cultural background (51). Correlation between the MMSE and delirium instruments range between 0.50 (NEECHAM (38)) to - 0.91 (MDAS (36)). In three instruments (DRS, MDAS and NEECHAM) the correlation was tested in different samples. All three showed a considerable difference in the two samples (DRS: - 0.43 - - 0.78; MDAS 0.55 - - 0.91; NEECHAM 0.50 - 0.81).

Construct validity was tested in half of the instruments. The relation between delirium and other patient characteristics was studied in the development of three instruments. In all three cases (DI, DRS, NEECHAM) the Barthel Index (52), a valid and reliable measure of Activities of Daily Living (ADL) performance, gave comparable outcomes (- 0.60, - 0.63, 0.70). The Older American Resources and Services (OARS) (53) contains a measure of ADL performance in the elderly and was used in two studies. Correlation between the DI and the OARS and the NEECHAM and the OARS were moderate (- 0.42, 0.47- 0.64). Scales developed for cognitive decline (the Informant Questionnaire on Cognitive Decline in the Elderly (54)) or dementia (Blessed Dementia Scale (55)) showed low correlation with the DI and the DRS. This was not expected since dementia is a high-risk factor for delirium, and the two diagnoses are related in a large group of patients. In five instruments (CAC-A, CTD, DRS, MDAS and NEECHAM), the underlying attribute was confirmed with factor analysis.

Discussion

Development of a measurement tool is a challenging and difficult task. This is particularly difficult when the object of the measurement is a syndrome with a fluctuating nature and a different presentation of symptoms. To eliminate errors of measurement, much emphasis should be given to the reliability and validity of an instrument. In this review, delirium instruments were described and reviewed with regard to psychometric qualities.

The choice of an instrument depends on the goal of use and the nature of the setting. Delirium instruments vary on content, goal, type of ratings, rater expertise and rating time.

Of the twelve instruments reviewed seven (CRS, CSE, CTD, DAS, DI, DSI, DSS) were tested in a single study. The development of the CAC-A, the CRS, the DAS, the DSI and the DSS was limited by testing one aspect of reliability and one aspect of validity. More comprehensively studied were the CAM, the CSE, the CTD, the DI, the DRS, the MDAS and the NEECHAM.

Reliability of delirium instruments shows good to excellent results. Validity of the delirium instruments is overall fair to good. Differences exist, however, in the degree to which reliability and validity were tested and the quality of the testing procedures. With regard to reliability, interrater reliability was established as single measure for reliability in half of the instruments (CAC-A, CAM, CRS, DAS, DI, DSS). The internal consistency of these instruments was not taken into account. With regard to validity content validity was established in five instruments and construct validity in six instruments. The DRS, the MDAS and the NEECHAM were studied for all aspects of validity.

Overall results of the EBM Quality Filters show excellent testing procedures of the CAM, the DAS, the MDAS and the DRS. The procedures used in testing the CAC-A, the CRS, the

CSE, the CTD, the DI, the DSI, the DSS and the NEECHAM Confusion Scale show some drawbacks.

Conclusion

Conclusion of this review is that most delirium instruments show promising results, however, they need further testing. Testing should be done in different samples and on a broader range of aspects with regard to reliability and validity. Much emphasis should be given to the procedures used in future studies.

Only a few instruments reviewed can be considered as 'ready to use' instruments, since they are well tested in more than one sample with good results. Of these instruments some psychometric aspects could be further explored. The CAM is found valid and reliable in several studies when used by physicians. For use by nurses or research assistants further study is recommended. Zou et al. (40) concluded that four factors are important for the rater's CAM diagnosis: 1) training for the CAM, 2) thoroughness of clinical observations, 3) spending enough time with the patient and 4) talking to family members and staff involved with the patient. Testing of internal consistency and construct validity of the CAM is needed, to confirm its' psychometric quality. The DRS is found valid and reliable in several studies, and is more extensively studied than the CAM. Results, however, on internal consistency and criterion related validity range from modest to good. The fact that the DRS does not measure alterations in consciousness nor in attention, needs to be considered. In the revised version that is currently under study, these symptoms will be measured (32). The MDAS is a new instrument, developed for use by physicians in oncology and palliative care. The instrument has been well tested and shows good results. Results on criterion related validity (MMSE) and sensitivity show differences that need to be further explored.

The NEECHAM is the only well tested instrument for use by non-physicians showing good results. A major drawback of the NEECHAM is, however, the question whether it measures delirium or confusion in general. The NEECHAM score should be seen as indication of the presence of delirium, it is not a diagnostic tool. Further development of this scale in which clear procedures of delirium diagnosis are followed, is recommended.

Acknowledgements

This review was made possible by a grant from the National Center on Nursing and Care (LCVV). The authors wish to thank Dr. Rianne de Wit and Elisabeth Tornquist for their suggestions and critical reading.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
2. Trzepacz P, Breitbart W, Franklin J, Levenson J, Martini DR, Wang P. Practice Guideline for the treatment of patients with delirium. *The American Journal of Psychiatry* 1999;156(5 supplement):1-20.
3. Rudberg MA, Pompei P, Foreman MD, Ross RE, Cassel CK. The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing* 1997;26:169-174.
4. Sandberg O, Gustafson Y, Brännström B, Bucht G. Clinical profile of delirium in older patients. *Journal of the American Geriatric Society* 1999;47:1300-1306.
5. Lipowski ZJ. *Delirium: acute confusional states*: Oxford University Press; 1990.
6. Johnson JC, Gottlieb GL, Sullivan E, Wanich C, Kinosiosian B, Forciea MA, et al. Using the DSM-III criteria to diagnose delirium in elderly general medical patients. *Journal of Gerontology: Medical Sciences* 1990;45(3):M113-119.
7. Williams MA, Campbell EB, Raynor SM, Mlynarczyk SM, Ward SE. Reducing acute confusional states in elderly patients with hipfractures. *Research in Nursing and Health* 1985;8:329-337.
8. Inouye SK, Dyck CHv, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the Confusion Assessment Method. *Annals of Internal Medicine* 1990;113(12):941-948.
9. Levkoff S, Liptzin B, Cleary P, Reilly CH, Evans D. Review of research instruments and techniques used to detect delirium. *International Psychogeriatrics* 1991;3(2):253-271.
10. Robertsson B. Assessment scales in delirium. *Dementia and Geriatric Cognitive Disorders* 1999;10:368-379.
11. Pompei P, Foreman M, Cassel CK, Alessi C, Cox D. Detecting delirium among hospitalized older patients. *Archives of Internal Medicine* 1995;155(3):301-7.
12. Rapp CG, Wakefield B, Kundrat M, Montes J, Tripp-Reimer T, Culp K, et al. Acute confusion assessment instruments: clinical versus research usability. *Applied Nursing Research* 2000;13(1):37-45.
13. Musselman DL, Hawthorne CN, Stoudemire A. Screening for delirium: a means to improved outcome in hospitalized elderly patients. *Reviews in Clinical Gerontology* 1997;7:235-256.
14. Smith MJ, Breitbart WS, Platt MM. A critique of instruments and methods to detect, diagnose and rate delirium. *Journal of Pain and Symptom Management* 1995;10(1):35-77.
15. Williams MA. Delirium/Acute Confusional States: evaluation devices in nursing. *International Psychogeriatrics* 1991;3(2):301-308.
16. Polit DF, Hungler BP. *Nursing Research, principles and methods*. 6th ed. New York: J.B. Lippincott Company; 1999.
17. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington: American Psychiatric Association; 1980.
18. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington D.C.: American Psychiatric Association; 1987.

19. Vermeersch PEH. The Clinical Assessment of Confusion-A. *Applied Nursing Research* 1990;3(3):128-133.
20. Vermeersch PEH. Response to 'the cognitive and behavioral nature of acute confusional states'. *Scholarly Inquiry for Nursing Practice: an International Journal* 1991;5(1):17-20.
21. Vermeersch P, Henly SJ. Validation of the structure for the "Clinical Assessment of Confusion-A". *Nursing Research* 1997;46(4):208-213.
22. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of Internal Medicine* 1993;119(6):474-480.
23. Williams MA, Hooloway JR, Winn MC, Wolanin MO, Lawler ML, Westwick CR, et al. Nursing activities and acute confusional states in hospitalized elderly patients. *Research in Nursing and Health* 1979;28(1):25-35.
24. Williams MA, Ward SE, Campbell EB. Confusion: testing versus observation. *Journal of Gerontological Nursing* 1988;14(1):25-30.
25. Robertsson B, Karlsson I, Styrod E, Gottfries CG. Confusional State Evaluation (CSE): an instrument for measuring severity of delirium in the elderly. *British Journal of Psychiatry* 1997;170:565-570.
26. Hart RP, Levenson JL, Sessler NS, Best AM, Schwartz SM, Rutherford LE. Validation of a cognitive test for delirium in medical ICU patients. *Psychosomatics* 1996;37(6):533-546.
27. Hart RP, Best AM, Sessler CN, Levenson JL. Abbreviated cognitive test for delirium. *Journal of Psychosomatic Research* 1997;43(4):417-423.
28. O'Keeffe ST. Rating the severity of delirium: the Delirium Assessment Scale. *International Journal of Geriatric Psychiatry* 1994;9:551-556.
29. McCusker J, Cole M, Bellavance F, Primeau F. Delirium. Reliability and validity of a new measure of severity of delirium. *International Psychogeriatrics* 1998;10(4):421-433.
30. Trzepacz PT, Baker RW, Greenhouse J. A symptom rating scale for delirium. *Psychiatry Research* 1988;23:89-97.
31. Trzepacz PT, Dew MA. Further analyses of the Delirium Rating Scale. *General Hospital Psychiatry* 1995;17:75-79.
32. Trzepacz PT. The Delirium Rating Scale. Its use in consultation-liaison research. *Psychosomatics* 1999;40(3):193-204.
33. Trzepacz PT, Mulsant BH, Dew MA, Pasternak R, Sweet RA, Zubenko GS. Is delirium different when it occurs in dementia? A study using the Delirium Rating Scale. *Journal of Neuropsychiatry* 1998;10(2):199-204.
34. Albert MS, Levkoff SE, Reilly C, Liptzin B, Pilgrim D, Cleary PD, et al. The Delirium Symptom Interview: an interview for the detection of delirium symptoms in hospitalized patients. *Journal of Geriatric Psychiatry and Neurology* 1992;5(1):14-21.
35. Bettin KM, Maletta GJ, Dysken MW, Jilk KM, Weldon DT, Kuskowski M, et al. Measuring delirium severity in older general hospital inpatients without dementia. *The American Journal of Geriatric Psychiatry* 1998;6(4):296-307.

36. Breitbart W, Rosenfeld B, Roth A, Smith MJ, Cohen K, Passik S. The Memorial Delirium Assessment Scale. *Journal of Pain and Symptom Management* 1997;13(3):128-137.
37. Champagne MT, Neelon VJ, McConnel ES, Funk S. The NEECHAM Confusion Scale: assessing acute confusion in hospitalized and nursing home elderly. *The Gerontologist* 1987;27(suppl):4A.
38. Neelon VJ, Champagne MT, McConnel E, Carlson J, Funk SG. Use of the NEECHAM Confusion Scale to assess acute confusional states of hospitalized older patients. In: *Key aspects of elder care; managing falls, incontinence and cognitive impairment*; 1992. p. 278-289.
39. Neelon VJ, Champagne MT, Carlson JR, Funk SG. The NEECHAM Confusion Scale: construction, validation, and clinical testing. *Nursing Research* 1996;45(6):324-330.
40. Zou Y, Cole MG, Primeau FJ, McCusker J, Bellavance F, Laplante J. Detection and diagnosis of delirium in the elderly: psychiatrist diagnosis, Confusion Assessment Method, or consensus diagnosis? *International Psychogeriatrics* 1998;10(3):303-308.
41. Rosen J, Sweet RA, Mulsant BH, Rifai AH, Pasternak R, Zubenko GS. The Delirium Rating Scale in a psychogeriatric inpatient setting. *Journal of Neuropsychiatry* 1994;6(1):30-35.
42. Rockwood K, Cosway S, Stolee P, Kydd D, Carver D, Jarret P, et al. Increasing the recognition of delirium in elderly patients. *Journal of the American Geriatric Society* 1994;42:252-256.
43. Rolfson DB, McElhaney JE, Jhangri GS, Rockwood K. Validity of the Confusion Assessment Method in detecting postoperative delirium in the elderly. *International Psychogeriatrics* 1999;11(4):431-438.
44. Rockwood K, Goodman J, Flynn M, Stolee P. Cross-validation of the Delirium Rating Scale in older patients. *Journal of the American Geriatric Society* 1996;44:839-842.
45. Lawlor PG, Nekolaichuk C, Gagnon B, Mancini IL, Pereira JL, Bruera ED. Clinical utility, factor analysis, and further validation of the Memorial Delirium Assessment Scale in patients with advanced cancer. *Cancer* 2000;88(12):2859-2867.
46. Csokasy J. Assessment of acute confusion: use of the NEECHAM Confusion Scale. *Applied Nursing Research* 1999;12(1):51-55.
47. Williams MA, Campbell EB, Raynor WJ, Musholt MA, Mlynarczyck SM, Crane LF. Predictors of acute confusional states in hospitalized elderly patients. *Research in Nursing and Health* 1985;8:31-40.
48. O'Keeffe ST, Ní Chonchubhair Á. Postoperative delirium in the elderly. *British Journal of Anaesthesia* 1994;73:673-687.
49. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society* 1975;23:433-441.
50. Folstein MG, Folstein SE, McHugh PR. "Mini-Mental State" a practical guide for grading the cognitive state of patients for clinicians. *Journal of Psychiatric Research* 1975;12:189-198.
51. Tombaugh T, McIntyre N. The Mini-Mental State Examination: a comprehensive review. *JAGS* 1992;40:922-925.

52. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Maryland State Medical Journal 1965;14:61-61.
53. Duke University Center for the Study of Aging and Human Development. Multidimensional functional assessment: the OARS methodology. Durham; 1978.
54. Jorm AF, Jacomb PA. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. Psychological Medicine 1989;19:1015-1022.
55. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. British Journal of Psychiatry 1968;114:797-811.

chapter 3

The Delirium Observation Screening (DOS) Scale: A screening instrument for delirium

Marieke J. Schuurmans
Lillie M. Shortridge-Baggett
Sijmen A. Duursma

Submitted

Abstract

This paper describes the development and testing of the Delirium Observation Screening (DOS) Scale, a 25 item scale developed to study early recognition of delirium based on nurses' observations during regular care. The scale was tested for content validity by a group of seven experts in the field of delirium. Internal consistency, predictive validity, concurrent and construct validity were tested in two prospective studies with high risk groups of patients; geriatric patients and elderly hip fracture patients. In a group of patients admitted to a geriatric department (n= 82), 4 patients became delirious, in a group of elderly hip fracture patients (n=92), 18 patients became delirious. The DOS Scale was determined to be content valid and showed high internal consistency in both studies. Predictive validity against the Diagnostic Statistical Manual-IV diagnosis of delirium made by a geriatrician was good in both studies. Concurrent validity testing with the Mini Mental State Examination (MMSE) showed good correlation in the hip fracture patients and moderate correlation in the geriatric patients. Concurrent validity with the research nurse's rated DOS Scale and Confusion Assessment Method (CAM), as tested in the group of hip fracture patients, showed fair to moderate. The construct validity of the DOS Scale was tested against the Informant Questionnaire of Cognitive Decline in Elderly (IQCODE), a pre-existing psychiatric diagnosis and the Barthel Index. Correlation with the IQCODE was good in the study with the hip fracture patients and low in the study with the geriatric patients. A pre-existing psychiatric diagnosis correlated fair in both studies. Correlation with the Barthel Index was low in the geriatric patients and moderate in the hip fracture patients.

The overall conclusion of these studies is that the DOS Scale shows satisfactory validity and reliability, to study early recognition of delirium by nurses' observation.

Introduction

Delirium is a transient organic mental syndrome, common in elderly patients after hospital admission. The syndrome is characterized by a disturbance in consciousness and cognitive functioning or perception that develops in hours to days and fluctuates during the course of the day (1, 2). Presentation of symptoms differs from patient to patient (3-5). The diagnosis of delirium is often overlooked or mistaken for dementia or depression (1, 6-9). Nurses, however, have frequent contacts with patients and are in a position to observe behavioral changes at an early stage (6, 10, 11). To improve recognition of delirium, we intended to study the possibility to come to the diagnosis of delirium with a limited number of nurses' observations.

For successful implementation of delirium screening, nurses need instruments that are based on observation, that allow bedside use during regular care, repetitively and without

respondent burden (12-14). The last fifteen years several instruments have been developed to diagnose delirium or to rate symptoms of delirium (15-20). Most of the instruments were developed for research or to confirm the objectivity of medical diagnoses. Ratings of instruments are based on interviews, tests and observation of patients. In three instruments the ratings are based on observation of patients; the Clinical Assessment of Confusion A (21-23), the Confusion Rating Scale (12, 24, 25) and the NEECHAM Confusion Scale (13, 14, 26). Unfortunately, these instruments share some methodological problems. They are developed to measure acute confusion. The description of acute confusion used is different for each of the instruments. None of them is based on the criteria for delirium as described in the Diagnostic Statistical Manual (DSM) of the American Psychiatric Association (2). The CAC-A showed a low sensitivity for the diagnosis of delirium (0.36) made by a geriatrician according to the DSM-III-R criteria (17). The CRS was never tested against a DSM diagnosis and should be seen as a premature scale according to the developers (25). Ratings of the NEECHAM Confusion Scale showed a correlation of 0.54 to 0.70 (14) and of 0.68 (27) with the DSM-III-R criteria for delirium. In these studies, however, the scales and the DSM criteria were applied by the same research nurse.

In order to study early recognition of delirium by a limited number of nurses' observations the CAC-A, the CRS and the NEECHAM Confusion Scale were not found suitable. Therefore the decision was made to develop a new instrument, the Delirium Observation Screening (DOS) Scale. The items of the scale are behavioral descriptions based on the DSM-IV criteria for delirium (2). Aim of the studies reported here was to test the reliability and validity of DOS Scale and its items.

Methods

Design and sample

The reliability and validity of the DOS Scale were tested in two prospective studies of populations at high risk for delirium. Both studies were accepted by the hospital ethics committee. All patients in both studies were asked to give informed consent to participate; in case of cognitive impairment, surrogate consent was requested from a close relative.

In Study 1, the sample included patients of 70 years and older, admitted to the department of geriatric medicine of the University Medical Center, Utrecht. Patients were excluded if they were diagnosed delirious on admission or in case of discharge or transfer within the first week of admission. In Study 2, the sample included patients with a hip fracture, who were at least 70 years of age, and after surgery were admitted to the surgical department of the Diaconessenhuis Utrecht. Patients were excluded when suspected for delirium on admission

or when it was known that they would be discharged or transferred in less than 5 days after surgery.

Procedure

Data collection started immediately after admission to the geriatric or surgical department, with nurses rating the DOS Scale at the end of every shift, based on their observations during that shift. The DOS Scales were collected in a closed box, to ensure that ratings would not be influenced by ratings on previous shifts. In Study 1 (geriatric patients) data were collected during the first 7 days after admission, in Study 2 (hip-fracture patients) data were collected during the first 6 days after admission. Age, gender and data on pre-hospital condition (Informant Questionnaire of Cognitive Decline in the Elderly (IQCODE)(28-30), psychiatric comorbidity and the Barthel Index (31, 32)) of the patients were collected from an interview with the patient and his or her family, and from medical and nursing records. As part of regular care a geriatrician tested patients with the Mini Mental State Examination (MMSE) (33) in Study 1. In Study 2 the research nurse tested the patients with the MMSE. The diagnosis of delirium was made by a geriatrician based on the DSM-IV criteria (2). Each positive diagnosis was discussed in a group of geriatricians. In both studies the geriatrician was blinded to the DOS Scale ratings of the nurses. In Study 1 the geriatrician saw all patients on a daily basis. In Study 2 a research nurse collected the nurses' DOS Scales and visited the patients on a daily basis. The research nurse, who was experienced in geriatric nursing, rated the DOS Scale, the Confusion Assessment Scale (CAM) (34) and the MMSE during these visits. The visits took place between 14.00 and 17.00 hours and had a duration of approximately 15 minutes. In case of alterations in behavior, a geriatrician of the University Medical Center visited the patient.

Measures

DOS Scale

The DOS Scale has been designed to measure early symptoms of delirium that nurses can assess during regular nursing care. The DOS Scale has been developed based on the DSM-IV criteria (2) for delirium, literature review and clinical experience.

The DOS Scale contains 25 behavioral items which involve eight symptoms (diagnostic and related): disturbances of consciousness (item 1-3), attention and concentration (item 4-6), thinking (item 7-11), memory/orientation (item 12-14), psychomotor activity (item 15-18), sleep/wake pattern (item 19-21), mood (item 22-23) and perception (item 24-25). Items are rated on a 5- point Likert scale based on the frequency of occurrence of the behavioral disturbances (1= never; 2= sometimes; 3= fluctuating; 4= usually; 5= always). A score of 1 is defined as 'normal behavior', meaning absence of behavioral alterations. In Study 2, a 4-

point Likert scale (excluding 'fluctuating') was used. Four items (1, 5, 13 and 14) are reverse scored, i.e., 'normal behavior' is rated as 'always'. An 'unable' score is given in case an item can not be rated due to patient characteristics (patient does not communicate, patient is asleep) or due to nurse characteristics (knowledge deficit). Criteria for rating the item are included in the instrument. The DOS Scale was designed to rate behavior that can be observed by nurses while providing daily care, and involves behavior that is observable by every nurse with a basic knowledge of geriatrics. It can be used also with patients who have a cognitive disturbance. Completion of the instrument requires less than 5 minutes (35).

Content validity of the DOS Scale was determined by experts in two stages as recommended by Lynn (36). Seven members of the Dutch multidisciplinary workgroup on delirium (two psychiatrists who worked with the elderly, one consultation psychiatrist, two geriatricians, two psychiatric consultation nurse specialists) rated the content relevance of the items on the DOS Scale on a 4-point Likert scale (4= very relevant; 3=relevant, but needs a slight change in formulation; 2= only relevant if the formulation of the item is totally changed; 1= not relevant at all). An item was considered acceptable if six of the experts gave it a score of 3 or 4. The experts were also asked to make suggestions on the items and to add items that in their opinion were missing.

The experts rated 21 out of 26 original items content valid; 6 of these 21 items needed small changes in formulation. Five items were rejected and 4 new items were added. In a second round, all the experts rated the final 25-item instrument valid.

CAM

The Confusion Assessment Method (34) is frequently used and has established validity and reliability as a delirium diagnostic instrument. It is a structured 9-item instrument that allows standardized recording of the symptoms of delirium based on the DSM-III-R criteria (37). Although the CAM was developed for use by non-psychiatric clinicians, the training of the rater influences the sensitivity and specificity of the instrument (7, 37-39).

MMSE

The MMSE is a widely used valid and reliable test of cognitive functioning. The MMSE measures orientation, memory, attention, ability to follow verbal and written commands, ability to write a sentence and ability to copy a figure. Patients need to be able to hear, read and write in order to be tested. Scores range between 0 and 30 and ratings below 24 indicate cognitive dysfunction. The MMSE is known to be influenced by age, educational level and cultural background (40, 41).

IQCODE

The IQCODE is a valid and reliable family member rating of cognitive impairment (28-30). Twenty-six items are rated on the degree of change over a 10-year period on a 5-point scale from 1, much improved through 3, not much change to 5, much worse. The average score

on all items gives a final score between 1 and 5. The influences of education, pre-morbid intelligence and cultural differences are discounted in the assessment. There is no influence of admission stress or early symptoms of delirium (28-30).

Barthel Index

The Barthel Index is a valid and reliable measure of Activities of Daily Living (ADL) (32, 42). The original index was scored in 5-point increments, giving a final score between 0 and 100 (31). In 1988, Collin et al. (32) rescored the Index in 1-point increments, increasing the accuracy of the Index. The latter score was used in the here reported studies. A score of 20 means totally independent in ADL and a score of 0 means totally dependent in ADL.

Statistical analysis

All data were processed with SPSS. Since the DOS Scale was developed for further study of nurses' observations all analyses were done at item level (each item describes an observation), and at sum score level (total of 25 items, items 1,5,13, and 14 were recoded before summing).

Reliability

Reliability of the DOS Scale was measured by internal consistency of the scale using Cronbach's alpha ($Cr \alpha$). To control for influences of ratings of 'normal behavior' the internal consistency was measured separately for the ratings of the delirious patients. Due to the fluctuating nature of delirium, test-retest reliability is not testable in delirious patients. Interrater reliability of the DOS Scale has been tested in three other studies and results were satisfactory (43).

Validity

The predictive validity of the DOS Scale was tested against the DSM-IV diagnosis made by a geriatrician. Both in Study 1 and Study 2, a Mann-Whitney U test was used to compare mean DOS Scale ratings of patients later diagnosed as delirious with mean ratings of non-delirious patients.

Concurrent validity of the DOS Scale was tested in Study 2 by comparing nurses' DOS Scale ratings based on their observation during an eight hour day shift, with the research nurse's DOS Scale ratings and the research nurse's CAM ratings. Concurrent validity was also tested by correlating the DOS Scale ratings with the MMSE.

Construct validity of the DOS Scale was tested in both studies by examining the correlations between mean DOS Scale ratings with other cognitive (IQCODE and pre existing psychiatric diagnosis) and clinical health status (Barthel Index) variables related to the occurrence of delirium. All correlation coefficients were calculated by a Spearman rank correlation coefficient (R_s).

Results

During Study 1, 94 patients were admitted to the geriatric department and 82 of them met the inclusion criteria. These patients had a mean age of 83 years (sd 6.17, range 70-96); 29.4% of them were male and 70.6% were female. Only three patients developed delirium during the first week after admission and one additional patient developed delirium on day 8 after admission. In the analyses the patient with delirium diagnosed on day eight, was included in the delirious group of patients.

During Study 2, 101 patients were admitted to the surgical department and 92 of them met the inclusion criteria. These patients had a mean age of 82.3 years (sd 6.65, range 70-98); 13% of them were male and 87% were female. Eighteen patients developed delirium during the first week of admission.

Reliability

Internal consistency of the DOS Scale was calculated per shift (Study 1: 21 shifts, Study 2: 18 shifts). In Study 1 the Cronbach's alpha ranged from 0.74 to 0.93; 12 out of 21 calculated alphas were over 0.90. Cronbach's alpha on the mean ratings was 0.96. In Study 2 Cronbach's alpha ranged from 0.78 to 0.98; 13 out of 18 alphas were over 0.90. Cronbach's alpha on the mean ratings was 0.97. In both studies, analysis of the items using the 'alpha if item deleted' procedure, showed that removal of items did not change the overall alpha more than 0.01. The Cronbach's alpha was also calculated on the mean ratings of delirious patients; in the delirious geriatric patients (n=4) alpha was 0.93, in the delirious hip fracture patients (n=18) alpha was 0.96. Table 1 shows the Cronbach's alpha of the DOS Scale on the level of symptoms.

Validity

The predictive validity of the DOS Scale using the DSM-IV diagnosis of delirium is shown in Table 2. In Study 1 there were significant differences on 16 items between patients' ratings before they were diagnosed as delirious and the ratings of non-delirious patients. All items with regard to alterations in consciousness, attention and concentration showed clear differences between the two groups. In Study 2 only one item (*'has vivid and frightening dreams'*), was not significant different for the ratings of the patients before they were diagnosed as delirious and the ratings of the non-delirious patients. Comparison of DOS Scale sum scores of the patients diagnosed as delirious and the non-delirious patients showed significant differences in both studies (Study 1: sum score 36.3 (sd 10.5) versus 61.6 (sd: 20.2), $p \leq 0.01$; Study 2: sum score 31.4 (sd: 8.7) versus 44.3 (sd: 8.2), $p \leq 0.001$).

Table 1 Cronbach's alpha of the DOS Scale on the level of symptoms in Study 1 (82 geriatric patients) and Study 2 (92 hip fracture patients)

Observations		Geriatric patients		Hip fracture patients	
		Cr α	Cr α if item deleted	Cr α	Cr α if item deleted
	<i>Consciousness</i>	.93		.91	
1	Reacts normally to addressing		.89		.83
2	Dozes during conversation or activities		.86		.89
3	Stares into space		.93		.86
	<i>Attention/concentration</i>	.91		.85	
4	Is easy distracted by stimuli from the environment		.92		.82
5	Maintains attention to conversation or action		.86		.76
6	Does not finish question or answer		.82		.80
	<i>Thinking</i>	.90		.93	
7	Gives answers which do not fit the question		.86		.90
8	Talks slowly or answers slowly		.87		.91
9	Reacts slowly to instructions		.85		.90
10	Speaks incoherently		.85		.90
11	Is suspicious		.93		.93
	<i>Memory/orientation</i>	.96		.94	
12	Thinks to be somewhere else		.96		.96
13	Knows which part of the day it is		.91		.90
14	Remembers recent event		.93		.87
	<i>Psychomotor activity</i>	.81		.86	
15	Is picking, disorderly, restless		.71		.79
16	Hardly moves		.87		.88
17	Pulls IV tubes, feeding tubes, catheters etc.		.69		.78
18	Has unnatural position in bed		.75		.81
	<i>Sleep/wake cycle</i>	.42		.54	
19	Is sleepy/drowsy during the day		.38		.64
20	(night shift) Has vivid and frightening dreams during the night		.40		.48
21	(night shift) Is awake/wakes up restlessly		.16		.20
	<i>Mood</i>	.21		.65	
22	Is easy or sudden emotional (frightened, angry, irritated)		*		*
23	Is apathetic/weary		*		*
	<i>Perception</i>	.81		.89	
24	Sees persons/things as somebody/something else		*		*
25	Sees/hears things which are not there		*		*
* items for which Cr α can not be calculated if one item is removed					

Table 2 Comparison of mean DOS Scale ratings of geriatric patients without delirium (n= 78) and before diagnosed delirious (n = 4) and post-operative hip fracture patients without delirium (n =74) and post-operative hip fracture patients before diagnosed delirious (n = 18)

Observations		Geriatric patients ¹			Hip fracture patients ²		
		Mean non delirious	Mean delirious	p-value	Mean non delirious	Mean delirious	p-value
1	Reacts normally to addressing*	4.45 (.59)	3.05 (1.26)	.03	3.69 (.46)	3.19 (.66)	≤ .001
2	Dozes during conversation or activities	1.30 (.43)	2.38 (.85)	.009	1.18 (.28)	1.68 (.57)	≤ .001
3	Stares into space	1.45 (.49)	2.55 (.92)	.01	1.25 (.38)	1.66 (.51)	≤ .001
4	Is easy distracted by stimuli from the environment	1.62 (.66)	2.36 (.85)	.05	1.34 (.54)	2.14 (.54)	≤ .001
5	Maintains attention to conversation or action*	4.21 (.74)	2.86 (1.28)	.05	3.41 (.69)	2.74 (.69)	≤ .001
6	Does not finish question or answer	1.56 (.65)	2.60 (.96)	.03	1.37 (.44)	1.89 (.39)	≤ .001
7	Gives answers which do not fit the question	1.49 (.66)	2.38 (.82)	.03	1.28 (.41)	1.85 (.47)	≤ .001
8	Talks slowly or answers slowly	1.68 (.74)	2.52 (.99)	ns	1.25 (.40)	1.58 (.52)	.01
9	Reacts slowly to instructions	1.74 (.77)	2.88 (1.02)	.03	1.37 (.49)	1.89 (.56)	≤ .001
10	Speaks incoherently	1.47 (.71)	2.83 (1.10)	.004	1.25 (.39)	1.86 (.66)	≤ .001
11	Is suspicious	1.24 (.45)	1.38 (.75)	ns	1.13 (.29)	1.63 (.63)	≤ .001
12	Thinks to be somewhere else	1.60 (.85)	2.71 (1.80)	ns	1.29 (.55)	2.18 (.49)	≤ .001
13	Knows which part of the day it is*	4.11 (.94)	2.24 (1.75)	.04	3.38 (.79)	2.45 (.86)	≤ .001
14	Remembers recent event*	3.88 (1.10)	2.15 (1.79)	ns	3.38 (.82)	2.29 (.67)	≤ .001
15	Is picking, disorderly, restless	1.32 (.56)	3.20 (1.41)	≤ .001	1.25 (.43)	2.04 (.56)	≤ .001
16	Hardly moves	1.62 (.63)	1.89 (.54)	ns	1.70 (.51)	2.20 (.54)	≤ .001
17	Pulls IV tubes, feeding tubes, catheters etc.	1.16 (.39)	2.78 (1.26)	.006	1.14 (.33)	1.59 (.54)	≤ .001
18	Has unnatural position in bed	1.25 (.33)	2.07 (.84)	≤ .001	1.22 (.28)	1.58 (.34)	≤ .001
19	Is sleepy/drowsy during the day	1.72 (.74)	2.18 (.74)	ns	1.48 (.48)	1.95 (.61)	≤ .001
20	(night shift) Has vivid and frightening dreams during the night	1.15 (.39)	1.00 (0)	ns	1.10 (.23)	1.17 (.32)	ns
21	(night shift) Is awake/wakes up restlessly	1.34 (.62)	2.92 (1.77)	.02	1.22 (.38)	1.60 (.61)	.005
22	Is easy or sudden emotional (frightened, angry, irritated)	1.38 (.50)	1.13 (.25)	ns	1.14 (.33)	1.90 (.81)	≤ .001
23	Is apathetic/weary	1.50 (.65)	1.93 (.71)	ns	1.21 (.34)	1.50 (.43)	.001
24	Sees persons/things as somebody/something else	1.20 (.40)	2.16 (1.28)	.04	1.11 (.24)	1.31 (.41)	.025
25	Sees/hears things which are not there	1.13 (.28)	2.06 (.84)	.01	1.07 (.16)	1.32 (.37)	≤ .001

¹ ratings 1= never, 2= sometimes, 3= fluctuating, 4=usually, 5= always, ² 1= never, 2= sometimes, 3=usually, 4= always
* items are reverse rated
Tested with a Mann-Whitney U test

Table 3 shows the concurrent validity of the DOS Scale as tested in Study 2. Most ratings indicated 'normal behavior' (nurses: range 50% - 89.7%, mean 73.6% sd 11.3; research nurse: range 55.9% - 93.1%, mean 80.3% sd 9.2). On 18 items the research nurse rated more behavior as 'normal' ; and on 6 items the difference was over 10%. The correlation between nurses' DOS Scale ratings (based on observation during an eight hour shift) and the research nurse's DOS Scale ratings (based on observation and interview during 15 minutes) were merely fair (between 0.21 - 0.40) to moderate (0.41- 0.60) on the level of items. Correlation between the nurses' DOS Scale sum score and the research nurse's DOS Scale sum score was R_s 0.54 ($p \leq 0.001$). The correlations with the CAM were comparable. Correlation of the DOS Scale and the CAM, both rated by the research nurse, show merely moderate (0.41- 0.60) to good (0.61 - 0.80) on the level of items, and 0.63 ($p \leq 0.001$) on the level of sum score. Correlation on the items with regard to memory and orientation show the best results in all comparisons. The correlations of the DOS Scale ratings with the MMSE (concurrent validity), the IQCODE, a pre-existing psychiatric diagnosis and the Barthel Index (construct validity) are shown in Table 4 for both studies. In Study 1 MMSE ratings were available of 28 patients (34%) and in Study 2 of 68 patients (74%). Strong correlations, on the level of items, with the MMSE ($R_s > 0.60$) were found in Study 2, in Study 1 these correlations were weaker (R_s range from - 0.09 to 0.74). Correlation between the MMSE and the DOS Scale sum score showed R_s - 0.66 ($p \leq 0.001$) in Study 1 and R_s - 0.79 ($p \leq 0.001$) in Study 2. In Study 1 IQCODE ratings were available of 53 patients (65%) and in Study 2 of 57 patients (62%). Strong correlations, on the level of items, ($R_s > 0.60$) were seen with the IQCODE in Study 2 , in Study 1 these correlations were weaker (R_s range from 0.13 to 0.39). Correlation between the IQCODE and the DOS Scale sum score showed R_s 0.33 ($p \leq 0.05$) in Study 1 and R_s 0.74 ($p \leq 0.001$) in Study 2.

The correlations with a pre-existing psychiatric diagnosis, on the level of items, were weak but significant in both studies; varying from 0.16 to 0.41 in Study 1, and from 0.25 to - 0.53 in Study 2 (the negative direction of the correlation can be explained by the reverse scoring of item 14). Correlation between the pre-existing psychiatric diagnosis and the DOS Scale sum score showed R_s 0.42 ($p \leq 0.001$) in Study 1 and R_s 0.43 ($p \leq 0.001$) in Study 2.

The correlations, on the level of items, with the Barthel Index were merely fair in Study 1 (R_s between 0.21 and 0.40) and moderate in Study 2 (R_s between 0.41 and 0.60). Correlation between the Barthel Index and the DOS Scale sum score showed R_s - 0.26 ($p \leq 0.05$) in Study 1 and R_s - 0.55 ($p \leq 0.001$) in Study 2.

Discussion

Two studies were done to test the psychometric properties of the DOS Scale and its items. The reliability of the DOS Scale, as measured by internal consistency, was high in both studies. The items of the DOS Scale, which were developed to study early recognition of delirium by nurses' observation, showed good predictive validity against the DSM-IV diagnosis of a geriatrician. In Study 2 on all, but one item (*'has vivid and frightening dreams'*) there were significant differences between the mean ratings of patients later diagnosed as delirious and the mean ratings of non-delirious patients. In Study 1, 16 items showed a significant difference. The DOS Scale sum score significantly differentiated the delirious and non-delirious patients in both studies. The DOS Scale sum score of the delirious geriatric patients were higher than of the delirious hip fracture patients.

The correlations between DOS Scale ratings of nurses and DOS Scale ratings of the research nurse in Study 2 were fair to moderate, on the level of items. The correlation on sum score was moderate. The research nurse rated higher percentages of 'normal behavior'. This indicates that nurses who observe patients over 8 hours see more alterations in behavior than the research nurse who sees the patient only during a 15 minute visit. The correlations between the nurses' DOS Scale ratings with the research nurse's CAM ratings in Study 2 showed comparable results. The correlations between the DOS Scale items and the CAM both rated by the research nurse in Study 2, were good. The correlations between the ratings of DOS Scale items by nurses and the MMSE were good in Study 2, and a little less in Study 1. On the level of DOS Scale sum score the correlation was good in both studies. These findings show a good concurrent validity of the DOS Scale.

Construct validity of the DOS Scale, measured in both studies by correlations with the IQCODE, a pre-existing psychiatric diagnosis and the Barthel Index, was better in Study 2 than in Study 1. In Study 2 all correlation coefficients were significant, in Study 1, nearly all were. Most correlations with the IQCODE, on the level of items, were over 0.60 in Study 2, indicating good construct validity. Correlations with psychiatric comorbidity and the Barthel Index were less strong. Differences found between Study 1 and Study 2 might be explained by the differences in characteristics of the populations and by the difference in delirium percentage.

Limitations can be found in both studies in the relatively small sample sizes, the low number of patients who developed delirium and in the number of missing data on the MMSE, IQCODE and the research nurse's ratings. The missing data on the MMSE and the other research nurse's ratings in Study 2 can be explained by logistic reasons (patients unavailable during the research nurse's visit due to nursing interventions, physiotherapy, visits, X-rays, patient asleep, patient not willing to be tested although informed consent was given etc.).

The use of the IQCODE, which in itself has several advantages when compared to other instruments for mental testing, showed to be difficult in both studies. Many elderly patients who live by themselves do not have close relatives who are able to answer questions that are experienced as threatening for the person reviewed. Finally the use of a physicians' diagnosis as 'gold standard' can be discussed. This 'gold standard' has limited reliability, reproducibility is questionable and the interpretation of the standard is prone to subjectivity. On the other hand there is no reliable and valid alternative. The use of other delirium instruments developed for diagnosis suggest a higher reproducibility, but it should be noted that none of the instruments known has a 100% sensitivity and a 100% specificity. In the developmental phase of these instruments it's sensitivity and specificity is also measured against a physician's diagnosis. In order to enhance reliability in this study all patients diagnosed delirious were discussed in a group of geriatricians. The diagnosis was never doubted.

The overall conclusion is that the DOS Scale showed to be a reliable and valid instrument to study nurses' observations of delirium during regular care. The strength of the individual items of the scale show some differences. Items related to consciousness, attention and thinking show the best results, items related to psychomotor activity, sleep/wake cycle, mood and perception show some limitations, mainly in concurrent and construct validity.

This conclusion is based on two prospective studies of different samples of patients, with relatively modest percentages of patients with delirium, therefore these results should be even stronger in larger groups of (delirious) patients. Which has to be tested. The main application of the DOS Scale is research, for use in clinical practice further analysis of items and development of an algorithm with a cut-off point is needed.

Acknowledgements

The development and testing of the DOS Scale was made possible by a grant of the National Center for Nursing and Care (LCVV). The authors would like to thank all patients and nurses who were involved in these studies. We also like to thank dr. G.J. Clevers, drs. Nijnke de Goeijen, drs. Yke Kooistra, Ruth Pel-Littel and the geriatricians of the department of Geriatric Medicine of the University Medical Center, Utrecht for their participation in these studies.

References

1. Lipowski ZJ. Delirium: acute confusional states: Oxford University Press; 1990.
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
3. Francis J, Martin D, Kapoor WN. A prospective study of delirium in hospitalized elderly. *JAMA* 1990;263(8):1097-1101.
4. Rudberg MA, Pompei P, Foreman MD, Ross RE, Cassel CK. The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing* 1997;26:169-174.
5. Sandberg O, Gustafson Y, Brännström B, Bucht G. Clinical profile of delirium in older patients. *Journal of the American Geriatric Society* 1999;47:1300-1306.
6. Foreman MD. The cognitive and behavioral nature of acute confusional states. *Scholarly Inquiry for Nursing Practice: an International Journal* 1990;5(1):3-20.
7. Rockwood K, Cosway S, Stolee P, Kydd D, Carver D, Jarret P, et al. Increasing the recognition of delirium in elderly patients. *Journal of the American Geriatric Society* 1994;42:252-256.
8. Lewis LM, Miller DK, Morley JE, Nork MJ, Lasater LC. Unrecognized delirium in ED geriatric patients. *American Journal of Emergency Medicine* 1995;13(2):142-145.
9. Farrel K, Ganzini L. Misdiagnosing delirium as depression in medically ill elderly patients. *Archives of Internal Medicine* 1995;155:2459-2464.
10. Yeaw EMJ, Abbate JH. Identification of confusion among the elderly in an acute care setting. *Clinical Nurse Specialist* 1993;7(4):192-197.
11. Morency CR, Levkoff SE, Dick KL. Research considerations delirium in hospitalized elders. *Journal of Gerontological Nursing* 1994;20(8):24-30.
12. Williams MA, Ward SE, Campbell EB. Confusion: testing versus observation. *Journal of Gerontological Nursing* 1988;14(1):25-30.
13. Neelon VJ, Champagne MT, McConnel E, Carlson J, Funk SG. Use of the NEECHAM Confusion Scale to assess acute confusional states of hospitalized older patients. In: *Key aspects of elder care; managing falls, incontinence and cognitive impairment*; 1992. p. 278-289.
14. Neelon VJ, Champagne MT, Carlson JR, Funk SG. The NEECHAM Confusion Scale: construction, validation, and clinical testing. *Nursing Research* 1996;45(6):324-330.
15. Levkoff S, Liptzin B, Cleary P, Reilly CH, Evans D. Review of research instruments and techniques used to detect delirium. *International Psychogeriatrics* 1991;3(2):253-271.
16. Smith MJ, Breitbart WS, Platt MM. A critique of instruments and methods to detect, diagnose and rate delirium. *Journal of Pain and Symptom Management* 1995;10(1):35-77.
17. Pompei P, Foreman M, Cassel CK, Alessi C, Cox D. Detecting delirium among hospitalized older patients. *Archives of Internal Medicine* 1995;155(3):301-7.
18. Musselman DL, Hawthorne CN, Stoudemire A. Screening for delirium: a means to improved outcome in hospitalized elderly patients. *Reviews in Clinical Gerontology* 1997;7:235-256.

19. Robertsson B. Assessment scales in delirium. *Dementia and Geriatric Cognitive Disorders* 1999;10:368-379.
20. Rapp CG, Wakefield B, Kundrat M, Menten J, Tripp-Reimer T, Culp K, et al. Acute confusion assessment instruments: clinical versus research usability. *Applied Nursing Research* 2000;13(1):37-45.
21. Vermeersch PEH. The Clinical Assessment of Confusion-A. *Applied Nursing Research* 1990;3(3):128-133.
22. Vermeersch PEH. Response to 'the cognitive and behavioral nature of acute confusional states'. *Scholarly Inquiry for Nursing Practice: an International Journal* 1991;5(1):17-20.
23. Vermeersch P, Henly SJ. Validation of the structure for the "Clinical Assessment of Confusion-A". *Nursing Research* 1997;46(4):208-213.
24. Williams MA, Hooloway JR, Winn MC, Wolanin MO, Lawler ML, Westwick CR, et al. Nursing activities and acute confusional states in hospitalized elderly patients. *Research in Nursing and Health* 1979;28(1):25-35.
25. Williams MA. Delirium/Acute Confusional States: evaluation devices in nursing. *International Psychogeriatrics* 1991;3(2):301-308.
26. Champagne MT, Neelon VJ, McConnel ES, Funk S. The NEECHAM Confusion Scale: assessing acute confusion in hospitalized and nursing home elderly. *The Gerontologist* 1987;27(suppl):4A.
27. Csokasy J. Assessment of acute confusion: use of the NEECHAM Confusion Scale. *Applied Nursing Research* 1999;12(1):51-55.
28. Jorm AF, Jacomb PA. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine* 1989;19:1015-1022.
29. Jorm AF, Scott R, Cullen JS, MacKinnon AJ. Performance of the Informant Questionnaire on cognitive decline in the Elderly (IQCODE) as a screening test for dementia. *Psychological Medicine* 1991;21:785-790.
30. Jorm AF, Christensen H, Henderson AS, Jacomb PA, Korten AE, MacKinnon A. Informant ratings of cognitive decline of elderly people: relationship to longitudinal change on cognitive tests. *Age and Ageing* 1996;25:125-129.
31. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Medical Journal* 1965;14:61-61.
32. Collin C, Wade DT, Davies S, Horne V. The Barthel Index: a reliability study. *Int Disabil Studies* 1988;10:61-63.
33. Folstein MG, Folstein SE, McHugh PR. "Mini-Mental State" a practical guide for grading the cognitive state of patients for clinicians. *Journal of Psychiatric Research* 1975;12:189-198.
34. Inouye SK, Dyck CHV, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the Confusion Assessment Method. *Annals of Internal Medicine* 1990;113(12):941-948.

35. Schuurmans MJ. Vroegtijdige signalering van het delier door verpleegkundigen. In: Onderzoek in de verpleging en verzorging in relatie tot de praktijk; Maastricht: De Tijdstroom/LCVV; 1996. p. 55-60.
36. Lynn MR. Determination and quantification of content validity. *Nursing Research* 1986;35(6):382-385.
37. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. Washington D.C.: American Psychiatric Association; 1987.
38. Zou Y, Cole MG, Primeau FJ, McCusker J, Bellavance F, Laplante J. Detection and diagnosis of delirium in the elderly: psychiatrist diagnosis, Confusion Assessment Method, or consensus diagnosis? *International Psychogeriatrics* 1998;10(3):303-308.
39. Rolfson DB, McElhaney JE, Jhangri GS, Rockwood K. Validity of the Confusion Assessment Method in detecting postoperative delirium in the elderly. *International Psychogeriatrics* 1999;11(4):431-438.
40. Anthony JC, LeResche L, Niaz U, Von K rff MR, Folstein MF. Limits of the 'Mini-Mental-State' as a screening test for dementia and delirium among hospital patients. *Psychological Medicine* 1982;12:397-408.
41. Tombaugh T, McIntyre N. The Mini-Mental State Examination: a comprehensive review. *JAGS* 1992;40:922-925.
42. Haan Rd, Limburg M, Schuling J, Broeshart J, Jonkers L, Zuylen Pv. Klinimetrische evaluatie van de Barthel-Index, een maat voor beperkingen in het dagelijks functioneren. *Nederlands Tijdschrift voor Geneeskunde* 1993;137(18):917-921.
43. Schuurmans MJ, Shortridge-Baggett LM, Duursma SA. The troublesome issue of interrater reliability: A study on nurses' observation of the symptoms of delirium. submitted: *Journal of Nursing Measurement* 2001.

chapter 4

The troublesome issue of interrater reliability: A study on nurses' observation of the symptoms of delirium

Marieke J. Schuurmans
Lillie M. Shortridge-Baggett
Sijmen A. Duursma

Submitted

Abstract

Delirium is an acute problem in elderly patients characterized by disturbances in consciousness, cognition or perception. The Delirium Observation Screening (DOS) Scale is an instrument developed for use in nursing practice based on observation by staff. Three different studies were done to test the DOS Scale for interrater reliability. In two studies there were simultaneous observations by pairs of nurses. In one study a videotape was used for observation by different groups of nurses. Agreement was expressed in different measures (percentage, Cohen's Kappa, Spearman rank correlation, Mann-Whitney U test). On the level of items Kappa scores were low. Other measures showed satisfactory interrater reliability in two of the three studies. Influences on interrater reliability of observations in clinical practice and the use of different measures are discussed.

Introduction

Delirium is an acute problem in elderly patients associated with high morbidity and mortality (1-8). The major symptoms are a disturbance of consciousness and cognition or perception(9); these symptoms fluctuate. Delirium develops in the course of a few hours to days (10). Early recognition of delirium symptoms will lead to better outcomes and nurses are in a position to observe changes in patient behavior (11, 12,13).

The Delirium Observation Screening (DOS) Scale was developed to rate the observed behavioral changes by nurses during regular care (14). The DOS Scale contains 25 behavioral items which involve eight symptoms (diagnostic and related): disturbance of consciousness (item 1-3), attention and concentration (item 4-6), memory and orientation (item 12-14), psychomotor activity (item 15-18), sleep/wake pattern (item 19-21), mood (item 22-23) and perception (item 24-25). Items are rated by nurses at the end of their shift, on a 5-point Likert scale based on the frequency of occurrence of the behavioral disturbance (1=never; 2=sometimes, 3= fluctuating; 4=usually; 5=always). A score of 1 is defined as 'normal behavior', meaning absence of behavioral alterations. Four items (1, 5, 13, 14) are reverse scored, i.e., 'normal behavior' is rated as 'always'. A score for 'unable' is given in cases where the nurse is not able to rate a behavior because of patient characteristics (for instance, no verbal contact possible or coma) or nurse characteristics (knowledge deficit or inexperience with the type of behavior). The DOS Scale has been developed based on the DSM-IV criteria for delirium (9), literature review, and clinical experience. Content validity was established by a group of experts using the method described by Lynn (15). The DOS Scale has been developed for nurses with basic knowledge of geriatric care, to be used during regular care and to be rated in a few minutes.

Interrater reliability reflects the extent to which an instrument gives equivalent results when used by different persons regarding the same subject (16). Tests for interrater reliability involve at least two raters and several subjects with different levels of the variable being measured. In most studies on delirium instruments interrater reliability is measured based on simultaneous assessment of two nurses, physicians or research assistants (17, 18). Tests of interrater reliability often show a similar design. The first phase is the training of research personnel in using the instrument. The second phase is to observe or test a few cases and rate them together with discussion on differences in ratings afterwards. The third phase is to rate a few cases at the same time, blinded for each other's ratings and to compute a Cohen's Kappa score of agreement. In some studies written vignettes (19, 20) or videotaped patients (21-24) are used. Most delirium instruments include testing of patients' capacities. Interrater reliability coefficients of delirium instruments are in these instruments over 0.90 (17, 18).

Rating patients' responses to a test is less subject to differences in interpretation of each rater than rating behavioral observations. The Confusion Rating Scale (CRS) (19, 25, 26) and the NEECHAM Confusion Scale (NEECHAM) (27-29) are observational delirium instruments. The CRS has an interrater reliability of 86.5% based on paper and pencil testing using vignettes (19). The interrater reliability of the NEECHAM Confusion Scale has been tested in several samples. Interrater reliability between two researchers was $r = 0.96$ (30), between trained staff nurses was Pearson's $r = 0.91$ and between staff nurses and researchers 91% (no r described), Kappa 0.65 (28). The studies of the CRS and the NEECHAM Confusion Scale do not describe the procedures of interrater reliability in detail nor the agreement on the level of items. The Confusional State Evaluation (CSE) (31) is rated based on an interview and observations. The description of interrater reliability of the CSE is more thoroughly. The interrater reliability was tested by two trained experienced nurses, rating the observations during an interview with elderly delirious patients. Agreement was described on item level as well as on total scores using different measures. Weighted Kappa ranged from 0.34 to 0.88 and Spearman rank correlation coefficients from 0.43 to 1.00 and the 'exact' agreement (accepting difference of one rating point) was 82-100%. Interrater reliability of rating observed behavior that fluctuates, is not further discussed in studies of the CRS, the NEECHAM or the CSE.

This article describes three studies conducted to test the interrater reliability of the DOS Scale. Since the DOS Scale has been developed for use by non-research personnel and for observation of more than one event, practical and methodological problems were faced, which may negatively influence the interrater reliability.

Methods

Study 1

Pairs of nurses made observations during the day shift as they cared for six patients referred for psychiatric consultation in a university hospital. Three of the patients were diagnosed delirious, the other three had psychiatric diagnoses. The nurses did not make all their observations together, and they rated patients' behavior at the end of their shift blinded for each others ratings and for the diagnosis of the patient. They had no special training and no earlier experience in rating the DOS Scale. The nurses did not know that the DOS Scale was developed for delirium observation; they were informed that the scale was developed for observation of psychiatric symptoms in general. After the ratings, the results were compared by the researcher and discussed with the nurses.

Study 2

Pairs of nurses observed for 4 hours (between 18.00 hours and 22.00 hours) thirty-six patients admitted to a department of geriatric medicine in a university hospital. One nurse was providing regular care to the patients; the other nurse did not participate in care for the patients but observed all care situations. Five nurses participated in this study. They were trained in nursing in geriatric medicine and had a minimum of four years of experience in this field. The nurses all had experience with the DOS Scale. All patients were selected by a geriatrician and had one or more symptoms of delirium. Patients and their peers were informed about the goal of the observations and were asked for oral consent. At 22.00 the nurses rated the behavior observed blinded for each other's observations.

Study 3

Four patients with mild to severe symptoms of delirium were videotaped. Each patient showed less severe symptoms during the process of taping. None of the tapes showed the combination of symptoms that clearly met the DSM-IV criteria for the diagnosis delirium.

In addition, five care situations of a nurse and a delirious patient during an evening shift were acted and videotaped. The actor was well experienced in playing a confused elderly patient. The nurse was experienced in nursing in geriatric medicine. The tape was made with a script, but to create the most natural situation the exact interactions, words and movements were not discussed in advance. Geriatricians reviewed the videotape for the DSM-IV criteria of delirium. Patients' behavior was rated on the DOS Scale based on the observation of the tape by 44 nurses and 5 clinical experts. All nurses were working in a university hospital (surgical and medical departments) and had no experience with the DOS Scale. Their age ranged from 22 to 56 years (mean: 33.8 sd: 8.2) and they had 1 to 35 (mean: 12.3 sd: 9.4)

years of experience in nursing. Almost 50% had experience in nursing elderly patients, mainly in hospital setting and 25.6% in a nursing home. None of them had a specific training in elderly care. The experts were all clinical nurse specialists in geriatric medicine or psychiatric consultation. Their age ranged from 35 – 53 (mean: 44.8 sd: 4.4) and they had 10 to 30 (mean: 22.6 sd: 8.0) years of experience in nursing. They all had special training in elderly care and many years of experience with geriatric patients in hospital settings.

All nurses and experts individually observed the tape in a private place where they were not distracted by other stimuli. They rated patient's behaviors immediately after observing the tape. This version of the DOS Scale differed slightly from the one used in Studies 1 and 2. Ratings were on a 4-point Likert scale and description of the symptoms (disturbance in consciousness, attention/concentration, thinking, etcetera) on which the items were developed were not printed on the scale (in the earlier version the related symptoms were printed in between the items). These changes were made based on the results in Study 1 and Study 2.

Statistical analysis

Data were analyzed using SPSS and a separate program for Cohen's Kappa measurements. Analyses were done at item level, giving 25 outcomes for each analysis, and at sum score level (total of 25 items, items 1, 5, 13 and 14 were recoded before summing). Agreement between the raters was described in percentage agreement; 'true' agreement giving the percentage of 'true' (100%) agreement (unable ratings excluded) and 'exact' agreement (31) giving the percentage of agreement accepting one rating point difference (unable ratings excluded). The percentage of unable ratings and ratings indicating 'normal behavior' were described. The DOS Scale item ratings of the two raters in Study 1 and Study 2 were compared using Cohen's Kappa coefficient (K) and the Spearman rank order correlation coefficient (R_s). The DOS Scale item ratings of nurses and experts in Study 3 were compared by a Mann-Whitney U test. DOS Scale sum scores were described in range, mean and standard deviation, and correlated (Study 1 and Study 2) with a Spearman rank correlation test and compared (Study 3) by a Mann-Whitney U test.

Results

Table 1 shows the various measures of agreement on the level of items in Study 1 (n= 6 patients) and Study 2 (n= 36 patients). Unable ratings are found limited in Study 1 and are found in nearly all items in Study 2. In Study 2 all memory/orientation related items show 50% or more unable ratings.

High percentages of ratings that indicate 'normal behavior' are seen in nearly all items in both studies. Differences between raters in the percentage of ratings that indicate 'normal behavior' are considerable; a difference of more than 20% in 7 items in Study 1 and in 5 items in Study 2. In Study 2 the second observer rates in 16 items a higher percentage of 'normal behavior' ratings.

In both studies there is a considerable difference in 'true' agreement and 'exact' agreement. 'true' agreement varies from 33% to 100%; mean 54.4%, sd 19.7% (Study 1) and from 27% to 77%; mean 49.4%, sd 12.7%. (Study 2). 'Exact' agreement varies from 66% - 100% , mean 89.6 sd 13.2% (Study 1) and 59% to 97%; mean 77.4%, sd 13.2% (Study 2).

Kappa scores vary in Study 1 from fair (0.21 - 0.41) to very good (> 0.81), with almost half of all items showing poor (< 0.20) Kappa scores. In Study 2 Kappa scores are poor to fair. Correlation measured by Spearman's rho show agreement rates in the range $r = 0.00$ to $r = 1.00$ in Study 1 and 0.05 to 0.63 in Study 2.

Table 2 shows the various measures of agreement at item level in Study 3 (n= 44 nurses and 5 experts). Nurses show more unable ratings than experts. High unable ratings are found in observation 17 'pulls at IV tubes, feeding tubes, catheters etc.' (nurses: 65.1%, experts: 40%). The videotaped patient had no such equipment but was tossing and turning his sheets nearly all the time.

Experts' ratings of 'normal behavior' in this patient were higher on most items than the nurses' ratings (experts versus nurses: 13 items show higher percentages of 'normal behavior'). Percentages of 'true' agreement and 'exact' agreement show considerable differences in Study 3. The 'true' agreement varies from 37.8% to 86%; mean 54.7% sd 12.7% (nurses) and from 40% to 100%; mean 63.5%, sd 14.9%. (experts). The 'exact' agreement varies from 53.3% - 100% , mean 82.1 sd 12.3% (nurses) and 50% to 100%; mean 88.6%, sd 16.2% (experts).

The troublesome issue of interrater reliability:
A study on nurses' observation of the symptoms of delirium

The agreement in DOS Scale sum scores is shown in Table 3. In Study 1 and Study 2 the sum score of the first rater is 8 points higher than the sum score of the second rater. The correlation of the sum score, calculated by a Spearman rank correlation coefficient, of both raters is excellent in Study 1 (R_s : 0.89, $p < 0.05$) and moderate in Study 2 (R_s : 0.42, $p < 0.01$). In Study 3 all nurses and experts observed the same patient, however, in both groups there is a considerable range of the sum scores. Comparison of the mean sum score of nurses and experts, with a Mann-Whitney U test shows no significant difference.

Table 3 Comparison of the DOS Scale sum score by two (groups of) raters in 3 studies

	Range ¹ Rater 1 Nurse (S3)	Range ¹ Rater 2 Expert (S3)	Mean (sd) Rater 1 Nurse (S3)	Mean (sd) Rater 2 Expert (S3)	R_s ²
Study 1	30 – 72	24 - 55	46.7(14.4)	38.3 (11.3)	.89*
Study 2	18 – 73	19 – 61	45.7 (17.3)	37.7 (12.7)	.42**
Study 3 ³	33 - 53	34 - 47	43.8 (4.6)	42.4 (5.2)	-
¹ Range of sum score: Study 1, Study 2: min 23 – max 115; Study 3: min 23 – max 92 ² Correlation between two raters, rating 6 (Study 1) and 36 (Study 2) patients, calculated by a Spearman rank correlation test (* $p \leq .05$, ** $p \leq .01$), ³ Correlation of the ratings of 44 nurses- 5 experts was not calculated since they were rating 1 patient					

Discussion

Three different studies are presented on interrater reliability of the DOS Scale. In Study 1 nurses had no specific experience in behavioral observations and no experience with the DOS Scale. Half of the patients ($n=3$) were delirious, the other had psychiatric symptoms. The rating was done at the end of a day shift in which two nurses cared for the one patient, however, not all observations were made the same moment. In Study 2 nurses were trained and experienced in behavioral observation and had experience with the DOS Scale. All patients had at least one symptom of delirium. Rating was done after four hours of simultaneous observations in the evening shift. In Study 3 none of the nurses had specific training, the experts did. The one patient observed had a diagnosis of delirium. All nurses observed the same patient since a videotape of the patient was used.

In most studies interrater reliability is measured by Cohen's Kappa (K). The advantage of K coefficients above simple percentages of agreement is the correction for agreement that can be expected by chance alone. The disadvantage of K is found in the effect of symmetrical imbalance in the margin totals. Large symmetrical imbalance in the marginal totals can convert reasonable percentage agreements in low values of K (32, 33). The high percentage

of ratings that indicate 'normal behavior' in Study 1 and Study 2 gave a serious imbalance of marginal totals, influencing K. The use of K should not be overestimated in these cases (34). Percentage agreement in the three studies, however, showed differences between 'true' agreement and 'exact' agreement. 'Exact' agreement was introduced in a study on the development of the Confusional State Evaluation (CSE) (31). Defined as the acceptance of a difference of one scale step. In Study 1 and 2 disagreement on 'sometimes' or 'fluctuating' was more frequent (Study 1: 4.8% and Study 2: 12.3%) than rating both 'sometimes' (Study 1: 3.0% and Study 2: 3.3%) or both 'fluctuating' (Study 1: 2.3% and Study 2: 4.9%). The same phenomenon was seen in ratings of 'fluctuating' and 'usually'. This means that the differentiation between 'sometimes' and 'fluctuating', and between 'fluctuating' and 'usually' is not clear. For this reason the 5-point Likert scale was changed in a 4-point Likert scale, excluding 'fluctuating', in Study 3.

Some nurses, however, might tend to rate more positive (error of leniency) and others might tend to rate more negative (error of severity). 'Exact' agreement as a measure for interrater reliability leads to good levels of percentage agreement of the DOS Scale in the three studies.

Apart from measures, agreement on observing behavior is a difficult matter. The DOS Scale was developed for nurses with a basic knowledge of geriatric medicine to observe their patients' behavior during the provision of daily care. Structured observations by non-research personnel may have the same problems as self-report scales (16). This means that observational biases can influence the rating of the patients' behavior. In Study 1 and 2 some observers might already have known the patient, which may have lead to identification of observations with previous ones (assimilatory bias). The fact that in both studies the sum score of the first rater (in Study 2 this was the primary care giver) is higher than the sum score of the second rater, might be a result of nurses knowing the patient. In a study of Rasmussen and Creason (35) nurses stressed the importance of knowing a patient in order to be able to recognize small changes in behavior.

Another factor influencing nurses' observation is the clinical setting itself. Interruptions and uncontrollable stimuli can create mitigating circumstances that threat precise observation and measurement. To avoid this influence videotapes can be used. On the other hand, videotapes must be made in a way that the entire phenomena is taped (23). For these reasons we decided to make a videotape with an actor instead of a patient. Geriatricians validated the tape. But even with valid and reliable tapes rating behavior raises difficulties. In a study of the interrater reliability of the Clinical Dementia Rating (CDR) (24) in which nurses and physicians rated four videotaped CDR interviews a moderate Kappa (overall 0.62) was found. In patients without dementia ($K = 0.62$) or with severe dementia ($K = 0.88$), Kappa was higher than in patients with questionable dementia ($K = 0.33$). The interrater reliability

of the Glasgow Coma Scale, a scale for standardized observation of patients with impaired consciousness, was also tested with the use of videotapes (21). A mean overall Kappa of 0.47 for nurses and 0.39 for physicians was found. Ratings on verbal responses and eye opening showed higher agreement coefficients. Ratings of motor response, however, showed moderate Kappa's (0.47 - 0.49). That rating observation of motor response by different raters can lead to problems, was demonstrated by a study in ergonomics (36). In a real work place study of postures required for work tasks in a stamping plant, two experienced and trained raters found agreement of $K = 0.07 - 0.43$ and percentage agreement of 39 - 54%. When data were dichotomized to presence or absence, Kappa changed to 0.05 - 0.55 and percentage agreement to 53% to 99%. Conclusion of this study is the need for more appropriate statistical methods, that lead to greater insight into sources of variability in reliability. Interrater reliability of the Nurses' Observation Scale for Geriatric Patients (NOSGER) was studied under the most realistic conditions (37). Each subject (113 elderly patients with dementia or depression) was rated by an individual pair of raters (nurses and other caregivers). Raters were not trained but only given brief verbal instructions. Assessment of memory and Activities of Daily Living (ADL) by two different raters was more objective ($r > 0.85$) than assessment of mood ($r = 0.76$), social behavior ($r = 0.68$) and disturbing behavior ($r = 0.70$).

A study on interrater reliability of delusions and bizarre delusions by psychiatrists with fair to moderate reliability, blamed different definitions and different levels of experience (20). In our studies various groups of nurses (untrained, trained, experts) showed some differences in ratings of 'normal behavior' in patients. In Study 1 (untrained nurses) in only eight items both nurses of a couple rated at least 50% of the observed behavior as not 'normal behavior'. In all other items both, or one of them, rated mainly 'normal behavior'. These were all patients referred to a consultation psychiatrist due to behavioral alterations. In Study 2 (trained experienced nurses) the nurse who observed and did not participate in the care, in 16 items rated a higher percentage of 'normal behavior' than the nurse who took care of the patient. Indicating that the nurse who was involved with the patient saw more alterations in behavior. In Study 3 the experts in 13 items rated a higher percentage of 'normal behavior' than the nurses; indicating that the nurses rated the observed behavior (which was videotaped and thus in all cases the same) more often as altered than the experts. This last indication is supported by the mean values of items of nurses, which were higher in 13 items. From the work of Benner (38) it is generally accepted that nurses and experienced nurses have different ways of processing information. Experienced nurses start to generate a hypothesis very early in their assessment. Every piece of information is tested for increase or decrease of the likelihood of the hypothesis. Nurses without specific experience in a certain area are much more open in their view. For rating the DOS Scale this might mean that more

experienced nurses do not look at the items, but at the symptoms or diagnosis represented by these items. For this reason we decided to remove the description of the symptoms on which the items were developed from the scale.

Conclusion

The DOS Scale sum score showed good interrater reliability in Study 1, moderate interrater reliability in Study 2 and good interrater reliability in Study 3. On the level of items there were considerable differences, however, most items showed good percentage agreement and good correlation between the two raters in Study 1. In Study 2 the results were less, items showed difficulties varying from high percentages of unable scores to low agreement between the raters. In Study 3 most items showed good interrater reliability. Nurses and experts had problems with a few items since both groups rated low agreement on these items. Interrater reliability of the items 20 and 21 was not tested in these studies.

Limitations can be found in the studies of the interrater reliability of the DOS Scale. Small sample size in Study 1, large proportions of 'normal behavior' in the three studies and the use of a single videotape in Study 3. Ratings on the DOS Scale are based on observations of nurses during a whole shift. Patients are observed during eight hours in different care situations (bathing, feeding, talking, wound care, etc.). The observed behavior is known to fluctuate. Procedures regularly used to test instruments for interrater reliability were not suitable for testing the DOS Scale. Rating observations of patients during regular care, and rating one video-taped patient, however, may have influenced the findings with regard to interrater reliability.

The DOS Scale showed satisfactory interrater reliability on sum score and on most items as well. Instruments with good interrater reliability when tested by researchers do not guarantee good interrater reliability when used by nurses in clinical practice (23). Specific emphasize to this aspect should be given when instruments developed for use in clinical practice are presented. Interrater reliability in observation based instruments need further discussion in nursing science, since observation is one of the most important tools in nursing assessment.

Acknowledgement

Study 2 and 3 were made possible by a grant from the National Center for Nursing and Care (LCVV). The authors especially thank Gabriël Roodbol who collected data in Nijmegen (Study 1), Thomas Borggreve (actor), Ruth Pel-Littel (nurse), Jan van der Woude (advisor) and the geriatricians of the University Medical Center in Utrecht (validation) for their part in developing the video-tape, all patients and nurses who participated in the studies and prof.dr. C.J.E. Schuurmans for his advice in statistics.

References

1. Lipowski ZJ. Delirium in the elderly patient. *New England Journal of Medicine* 1989;320(2):578-582.
2. Brännström B, Gustafson Y, Norberg A, Winblad B. ADL performance and dependency on nursing care in patients with hip fractures and acute confusion in a task allocation care system. *Scandinavian Journal of Caring Sciences* 1991;5(1):3-11.
3. Foreman MD, Grabowski R. Diagnostic dilemma: cognitive impairment in the elderly. *Journal of Gerontological Nursing* 1992;18(9):5-12.
4. Foreman MD. Acute confusion in the elderly. *Annual Review of Nursing Research* 1993;11:3-30.
5. Inouye SK. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *The American Journal of Medicine* 1994;97:278-288.
6. Pompei P, Foreman M, Rudberg MA, Inouye SK, Braund V, Cassel CK. Delirium in hospitalized older persons: outcomes and predictors. *Journal of the American Geriatric Society* 1994;42:809-815.
7. O'Keeffe S, Lavan L. The prognostic significance of delirium in older hospital patients. *Journal of the American Geriatric Society* 1997;45:174-178.
8. Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P. Does delirium contribute to poor hospital outcomes. *Journal of General Internal Medicine* 1998;13:234-242.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
10. Lipowski ZJ. *Delirium: acute confusional states*: Oxford University Press; 1990.
11. Foreman MD. The cognitive and behavioral nature of acute confusional states. *Scholarly Inquiry for Nursing Practice: an International Journal* 1990;5(1):3-20.
12. Yeaw EMJ, Abbate JH. Identification of confusion among the elderly in an acute care setting. *Clinical Nurse Specialist* 1993;7(4):192-197.
13. Morency CR, Levkoff SE, Dick KL. Research considerations delirium in hospitalized elders. *Journal of Gerontological Nursing* 1994;20(8):24-30.
14. Schuurmans MJ. Vroegtijdige signalering van het delier door verpleegkundigen. In: *Onderzoek in de verpleging en verzorging in relatie tot de praktijk*; 1996; Maastricht: De Tijdstroom/LCVV; 1996. p. 55-60.
15. Lynn MR. Determination and quantification of content validity. *Nursing Research* 1986;35(6):382-385.
16. Polit DF, Hungler BP. *Nursing Research, principles and methods*. 6th ed. New York: J.B. Lippincott Company; 1999.
17. Smith MJ, Breitbart WS, Platt MM. A critique of instruments and methods to detect, diagnose and rate delirium. *Journal of Pain and Symptom Management* 1995;10(1):35-77.

18. Rapp CG, Wakefield B, Kundrat M, Montes J, Tripp-Reimer T, Culp K, et al. Acute confusion assessment instruments: clinical versus research usability. *Applied Nursing Research* 2000;13(1):37-45.
19. Williams MA. Delirium/Acute Confusional States: evaluation devices in nursing. *International Psychogeriatrics* 1991;3(2):301-308.
20. Mojtabai R, Nicholson R. Interrater reliability of ratings of delusions and bizarre delusions. *American Journal of Psychiatry* 1995;152(12):1804-1806.
21. Juarez VJ, Lyons M. Interrater reliability of the Glasgow Coma Scale. *Journal of Neuroscience Nursing* 1995;27(5):283-286.
22. Bettin KM, Maletta GJ, Dysken MW, Jilk KM, Weldon DT, Kuskowski M, et al. Measuring delirium severity in older general hospital inpatients without dementia. *The American Journal of Geriatric Psychiatry* 1998;6(4):296-307.
23. Whall AL. The measurement of need-driven dementia-compromised behavior. *Journal of Gerontological Nursing* 1999;25(9):33-37, 50-51.
24. Rockwood K, Strang D, MacKnight C, Downer R, Morris JC. Interrater reliability of the Clinical Dementia Rating in a multicenter trial. *JAGS* 2000;48:558-559.
25. Williams MA, Campbell EB, Raynor SM, Mlynarczyk SM, Ward SE. Reducing acute confusional states in elderly patients with hipfractures. *Research in Nursing and Health* 1985;8:329-337.
26. Williams MA, Ward SE, Campbell EB. Confusion: testing versus observation. *Journal of Gerontological Nursing* 1988;14(1):25-30.
27. Neelon VJ, Champagne MT, McConnel E, Carlson J, Funk SG. Use of the NEECHAM Confusion Scale to assess acute confusional states of hospitalized older patients. In: *Key aspects of elder care; managing falls, incontinence and cognitive impairment*; 1992. p. 278-289.
28. Neelon VJ, Champagne MT, Carlson JR, Funk SG. The NEECHAM Confusion Scale: construction, validation, and clinical testing. *Nursing Research* 1996;45(6):324-330.
29. Miller J, Neelon V, Champagne M, Baily D, Ng'andu N, Belyea M, et al. The assessment of acute confusion as part of nursing care. *Applied Nursing Research* 1997;10(3):143-151.
30. Champagne MT, Neelon VJ, McConnel ES, Funk S. The NEECHAM Confusion Scale: assessing acute confusion in hospitalized and nursing home elderly. *The Gerontologist* 1987;27(suppl):4A.
31. Robertsson B, Karlsson I, Styrod E, Gottfries CG. Confusional State Evaluation (CSE): an instrument for measuring severity of delirium in the elderly. *British Journal of Psychiatry* 1997;170:565-570.
32. Feinstein AR, Cicchetti DV. High agreement but low kappa: I the problems of two paradoxes. *Journal of Clinical Epidemiology* 1990;43(6):543-549.
33. Cicchetti DV, Feinstein AR. High agreement but low kappa: II resolving the paradoxes. *Journal of Clinical Epidemiology* 1990;43(6):551-558.

34. Houwelingen JCv, Stijnen T, Strik Rv. Inleiding tot de medische statistiek. Bunge: Wetenschappelijke uitgeverij; 1995.
35. Rasmussen BH, Creason NS. Nurses' perception of the phenomenon confusion in elderly hospitalized patients. *Vard i Norden* 1991;21(11):5-12.
36. Burt S, Punnet L. Evaluation of interrater reliability for posture observations in a field study. *Applied Ergonomics* 1999;30:121-135.
37. Wahle M, Haller S, Spiegel R. Validation of the NOSGER (Nurses' Observation Scale for Geriatric Patients): reliability and validity of a caregiver rating instrument. *International Psychogeriatrics* 1996;8(4):525-547.
38. Benner P. From novice to expert. Menlo Park, CA: Addison Wesley; 1984.

chapter 5

Recognition of delirium by standard nursing observation

Gabriël Roodbol

Marieke J. Schuurmans

Claudia J. Gamel

Carla Frederiks

Lillie M. Shortridge-Baggett

Submitted

Abstract

This study examined the validity, reliability and ease of use of the Delirium Observation Screening (DOS) Scale, a scale of 25 items developed to assess the symptoms of delirium during regular nursing care. When tested with 57 patients referred for psychiatric consultation in a university hospital, 17 items of the DOS Scale differentiated significantly between patients with delirium (n=33) and those without delirium (n=24). On 8 of these 17 items, the mean difference between the groups was more than 1 on a 5-point scale. The DOS Scale sum score differentiated significantly between the delirious and non-delirious patients. The internal consistency of the DOS Scale was very good with this sample ($\alpha = 0.92$). Six patients were rated by two nurses, percentage agreement was used to examine interrater reliability. Agreement varied from poor to excellent on the level of items, on 13 items 'exact agreement' was 100%. The DOS Scale sum score showed high correlation between the two raters ($R_s = 0.89$). The ease of use of the DOS Scale was evaluated to be high. These results suggest that the DOS Scale is suitable for the observation of delirium symptoms by nurses. Further development of the DOS Scale in prospective studies in different populations is recommended.

Introduction

Delirium is a psychiatric syndrome resulting primarily from the disturbed somatic condition of the patient. Its major symptoms are disturbances in consciousness and cognitive functioning that develop in hours to days and fluctuate during the course of the day. The diagnostic criteria for delirium given in the Diagnostic and Statistical Manual IV criteria of the American Psychiatric Association (1) are shown in Table 1.

In addition to these symptoms, the DSM-IV describes three related symptoms of delirium: disturbed mood, disturbed sleep-wake pattern and hyperactive or hypo-active psychomotor activity (1, 2).

Delirious patients may have different symptom presentations and because of these different presentations and the fluctuating nature of symptoms the diagnosis is difficult to make. Nurses have frequent contacts with patients and are in a position to observe changes in cognitive functioning. Yet several studies have shown that nurses often do not recognize delirium (3-6). Lack of knowledge and lack of awareness of the importance of the syndrome may explain this finding.

The Delirium Observation Screening (DOS) Scale is a recently developed instrument designed to identify early symptoms of delirium (7). In the study reported here, the DOS Scale was evaluated for validity, reliability and ease of use in general hospital patients with psychiatric symptoms.

Table 1 **Diagnostic criteria of DSM-IV for delirium diagnosis**

A	Disturbance of consciousness with reduced ability to focus, sustain or shift attention
B	Changed cognition or the development of a perceptual disturbance
C	Disturbance develops in a short period of time and fluctuates over the course of the day
D	There is evidence from history, physical examination or laboratory findings that the disturbance is: <ol style="list-style-type: none"> 1 Physiological consequence of general condition 2 Caused by intoxication 3 Caused by medication 4 Caused by more than one etiology

Methods

Design and sample

This study was approved by of the hospital medical ethics committee. Subjects were patients admitted to the Nijmegen University Hospital and referred to the Psychiatric Consultation Services of the hospital. Patients who were referred to a mental hospital after Emergency Room consultation, patients who were admitted for attempted suicide, patients who were seen by a clinical nurse specialist in psychiatry and not evaluated by a psychiatrist, and patients who were admitted in the weekend and started treatment immediately were excluded from the study. Nurses who cared for the patients participated in the study. All nurses in the hospital are educated at the bachelor's level or working toward the baccalaureate. Over a 3-month period, 117 patients were referred for psychiatric consultation but 45 of these were excluded from the study for one of the reasons given above. A total of 68 DOS Scales were rated by nurses (94% of the eligible patients). Eleven patients were rated on the DOS Scale twice: five due to re-referral for psychiatric consultation and six to evaluate interrater reliability. For each double case, one rating sheet was deleted, blind. Thus data on 57 patients were analyzed.

Procedure

The attending physician referred patients for psychiatric consultation. A consultant psychiatrist examined the patients according to the regular procedures for psychiatric consultation. Diagnoses were made using the DSM-IV criteria. The nurse caring for the

patient at the time of the psychiatric consultation was asked to complete the DOS Scale at the end of his or her shift. The nurse was blinded to the diagnosis made by the psychiatrist. The nurse was asked to evaluate the DOS Scale after rating the patient. The nurses who participated in the six paired ratings were interviewed afterwards to gain more insight in their experiences with the scale.

Measures

The DOS Scale

The DOS Scale is designed to measure early symptoms of delirium that nurses can assess during regular nursing care. The DOS Scale is developed based on the DSM-IV criteria (1) for delirium, literature review and clinical experience. Content validity was established by a group of experts using the method as set forth by Lynn (8).

The DOS Scale contains 25 behavioral items which describe eight symptoms (diagnostic and related): disturbances of consciousness (item 1-3), attention and concentration (item 4-6), thinking (item 7-11), memory/orientation (item 12-14), psychomotor activity (item 15-18), sleep/wake pattern (item 19-21), mood (item 22-23) and perception (item 24-25). Items are rated on a 5-point Likert scale based on the frequency of occurrence of the behavioral disturbances (1= never; 2= sometimes; 3= fluctuating; 4= usually; 5= always). A score of 1 is defined as 'normal behavior', meaning absence of behavioral alterations. Four items (1, 5, 13 and 14) are reverse scored, i.e., 'normal behavior' is rated as 'always'. Criteria for rating the item are included in the instrument. The DOS Scale was designed to rate behavior that can be observed by nurses while providing daily care, and involves behavior that is observable by every nurse with a basic knowledge of geriatrics. It can also be used with patients who are already cognitively disturbed. Completion of the instrument requires less than 5 minutes (7). For this study the instrument was called "Nursing Observation of Psychiatric Symptoms" to prevent bias towards rating patient behavior in the direction of delirium; also information about delirium was removed from the manual.

Statistical analysis

Data were processed using SPSS. Since the study was done in dayshift, two items (20, 21) that could be rated only in the night shift were excluded from the analysis. All analysis were done at item level and at sum score level (total of 25 items, item 1, 5, 13 and 14 were recoded before summing).

The reliability of the DOS Scale was measured by internal consistency and the interrater reliability in this population. Internal consistency was calculated by Cronbach's alpha (α). Interrater reliability was tested with six patients who were rated by two nurses, and is reported in the percentage of agreement and 'exact agreement', defined as the agreement

within one scale step (9). Agreement on DOS Scale sum score is described by a Spearman rank correlation coefficient.

Concurrent validity was tested by comparing the DOS Scores to the diagnosis of the psychiatrist using the Mann-Whitney U test. Ratings of patients diagnosed as delirious were compared to ratings of patients who received other psychiatric diagnosis.

Ease of use was evaluated by a standard set of questions on the language used to describe the behavioral observations and the knowledge and time needed to rate the DOS Scale. Responses were given on a 4-point Likert scale.

Results

Thirty-three of the 57 patients were diagnosed as delirious. Psychiatric diagnoses in the group of non-delirious patients ($n = 24$) included major depression, adjustment disorders, post-traumatic stress disorder and somatoform disorders. Delirious patients were significantly older (mean 63 years, sd 14.9, range 24-82 years) than the non-delirious (mean 47 years, sd 15.1, range 25-75 years), X^2 : $p < 0.001$, which is consistent with the view that age is a risk factor for delirium (2). There were almost four times as many delirious males as delirious females (26 males, 7 females, X^2 : $p < 0.00$). Williams-Russo (10) found male gender to be a risk factor for delirium, but there is no scientific explanation for this.

Reliability

Internal consistency was rated using Chronbach's alpha ($Cr \alpha$). Since this study was done on the day shift two items (20 and 21) relevant only to night shift nurses were excluded from the analysis. Cases with missing values were deleted ($n=34$). For the remaining cases ($n= 23$) the DOS Scale showed excellent internal consistency, $Cr \alpha = 0.92$. Analysis of the items using the 'alpha if item deleted' procedure showed that removal of items did not have major consequence for the overall alpha. Internal consistency was also calculated on the level of symptoms, as shown in Table 1. All symptoms, except sleep-wake cycle and mood, showed good to excellent internal consistency.

Interrater reliability of the DOS Scale was tested in six patients. Both nurses rated a considerable percentage of 'normal behavior', however, nurse 1 rated less 'normal behavior' compared with nurse 2 (nurse 1, mean 37.7 % sd 24.2; nurse 2, mean 54.4% sd 23.7). 'True agreement' (complete agreement) ranged from 33% to 100% (mean 54.4%, sd 19.7%) and 'exact agreement' (agreement within 1 point on the 5-point rating scale) ranged from 66% to 100% (mean 89.6%, sd 13.2%). On 10 items the raters differed more than one point on the scale. 'Exact agreement' on these items was: item 5 (67%), item 7 (84%), item 8 (67%), item 9 (83%), item 10 (83%), item 11 (67%), item 14 (80%), item 16 (80%), item 18 (84%) and

item 22 (66%). The DOS Scale sum score by nurse 1 ranged from 30-72, mean 46.7 (sd 11.4) and by nurse 2 from 24-55, mean 38.3 (sd 11.3). The sum scores had good correlation: $R_s = 0.89$, $p = 0.02$.

Table 1 Cronbach's alpha of the DOS Scale on the level of symptoms

Observations		Cr α	Cr α if item deleted
	<i>Consciousness</i>	.72	
1	Reacts normally to addressing		.77
2	Dozes during conversation or activities		.43
3	Stares into space		.70
	<i>Concentration/Attention</i>	.65	
4	Is easily distracted by stimuli from the environment		.62
5	Maintains attention to conversation or action		.66
6	Does not finish question or answer		.27
	<i>Thinking</i>	.80	
7	Gives answers which do not fit the question		.79
8	Talks slowly or answers slowly		.76
9	Reacts slowly to instructions		.69
10	Speaks incoherently		.73
11	Is suspicious		.83
	<i>Memory/orientation</i>	.79	
12	Thinks to be somewhere else		.85
13	Knows which part of the day it is		.59
14	Remembers recent event		.64
	<i>Psychomotor activity</i>	.66	
15	Is picking, disorderly, restless		.64
16	Hardly moves		.64
17	Pulls IV tubes, feeding tubes, catheters etc.		.50
18	Has unnatural position in bed		.61
	<i>Sleep/wake cycle</i>	-	
19	Is sleepy/drowsy during the day		-
20	(night shift) Has vivid and frightening dreams during the night		-
21	(night shift) Is awake/wakes up restlessly		
	<i>Mood</i>	-.62	
22	Is easily or suddenly emotional (frightened, angry, irritated)		*
23	Is apathetic/weary		*
	<i>Perception</i>	.97	
24	Sees persons/things as somebody/something else		*
25	Sees/hears things which are not there		*
Cr α could not be calculated on the symptom disturbed sleep-wake cycle since only 1 item (19) was measured			
* items for which Cr α can not be calculated if one item is removed			

Validity

The significance of the differences in DOS Scale ratings of delirious and non-delirious patients was tested using the Mann-Whitney U test. Results are given in Table 2.

Seventeen of the 23 behavioral items included in the analysis showed significant differences between delirious and non-delirious patients. On 8 items there was a mean difference between the two groups of one scale step or more. Seven of these items reflected the symptoms of disturbed thinking and memory/orientation. Comparison of DOS Scale sum scores of delirious (mean 50.4, sd 13.4) and non-delirious (mean 35.8, sd 7.2) patients showed significant differences ($p \leq 0.001$).

Table 2 Comparison of mean DOS Scale ratings of patients with psychiatric symptoms not diagnosed delirious (n= 24) and of patients with psychiatric symptoms diagnosed delirious (n= 33)

Observations		Mean ¹ non delirious	Mean ¹ delirious	p-value
1	Reacts normally to addressing*	4.37 (1.21)	3.42 (1.20)	$\leq .001$
2	Dozes during conversation or activities	1.26 (.54)	2.06 (1.09)	.002
3	Stares into space	1.61 (.72)	2.16 (1.11)	ns
4	Is easy distracted by stimuli from the environment	1.74 (.86)	2.40 (1.13)	.02
5	Maintains attention to conversation or action *	3.87 (1.18)	3.16 (1.02)	.03
6	Does not finish question or answer	1.77 (.97)	2.62 (1.29)	.010
7	Gives answers which do not fit the question	1.35 (.93)	2.54 (1.07)	$\leq .001$
8	Talks slowly or answers slowly	1.83 (1.15)	2.83 (1.39)	.007
9	Reacts slowly to instructions	1.83 (1.03)	3.03 (1.21)	$\leq .001$
10	Speaks incoherently	1.22 (.42)	2.52 (1.15)	$\leq .001$
11	Is suspicious	1.50 (.96)	2.15 (1.16)	.02
12	Thinks to be somewhere else	1.13 (.45)	2.32 (1.25)	$\leq .001$
13	Knows which part of the day it is *	4.57 (1.16)	3.00 (1.36)	$\leq .001$
14	Remembers recent event *	4.42 (1.02)	2.81 (1.36)	$\leq .001$
15	Is picking, disorderly, restless	1.46 (.83)	2.91 (1.25)	$\leq .001$
16	Hardly moves	1.96 (1.30)	2.43 (1.33)	ns
17	Pulls IV tubes, feeding tubes, catheters etc.	1.00 (.00)	1.88 (1.24)	$\leq .001$
18	Has unnatural position in bed	1.32 (.57)	2.00 (1.48)	ns
19	Is sleepy/drowsy during the day	2.57 (1.27)	2.39 (.92)	ns
20	(night shift) Has vivid and frightening dreams during the night	-	-	-
21	(night shift) Is awake/wakes up restlessly	-	-	-
22	Is easy or sudden emotional (frightened, angry, irritated)	1.78 (1.09)	2.69 (1.55)	ns
23	Is apathetic/weary	2.58 (1.32)	2.69 (1.34)	ns
24	Sees persons/things as somebody/something else	1.04 (.21)	2.00 (1.20)	$\leq .001$
25	Sees/hears things which are not there	1.04 (.21)	1.88 (1.11)	$\leq .001$

¹ 1=never, 2=sometimes, 3=fluctuating, 4= usually, 5=always, * items are reverse rated
 Items 20 and 21 were not rated since rating took place in day shift
 Tested with a Mann-Whitney U test

Ease of use

The nurses (n=68) evaluated the ease of use of the DOS Scale very highly. The terminology used was clear to them (93%) and the descriptions of the behaviors on the instrument corresponded well with language used in practice (93%). In general, the nurses believed they had sufficient knowledge and experience to rate the items on the DOS Scale (85%). Eighty-five percent agreed that the items were listed in a neutral fashion. They thought the 5 points of the Likert scale differentiated clearly the observed frequency of behavior (82%), and rating the scale was easy (59%). Of the nurses, however, 31% did not answer the question on ease of rating, and almost half of the nurses (47%) felt that some items could be interpreted in different ways. Most nurses felt that the instructions were useful (80%) and they did not need to ask a colleague for help in rating the DOS Scale (85%). The majority believed that the DOS Scale was helpful in detecting psychiatric symptoms (79%) and could offer additional value to their clinical practice (66%).

Discussion

This study examined the ability of the DOS Scale to discriminate delirious patients from non-delirious patients with psychiatric symptoms based on standard nurses' observations. Of the 25 standard observations described in the items of the DOS Scale, 17 showed significant discriminative value. Eight items were rated considerably higher (> 1 scale step difference) in delirious patients than in the non-delirious patients. The DOS Scale sum score significantly differentiated the delirious from the non-delirious patients. The internal consistency of the DOS Scale is very good (Cr α = 0.92). There were, however, several cases with missing values since nurses were not able to rate all items for every patient. An explanation for this came from the pairs of nurses involved in the interrater reliability measurement. They said that they were unable to rate all the behaviors listed when they could not communicate verbally with the patients. Thus the DOS Scale might be less useful in patients who cannot communicate verbally. Interrater reliability on DOS Scale sum score level was good.

Main limitation of this study is found in the method and small sample size used for testing interrater reliability. The observations were made during regular nursing care, but symptoms of delirium fluctuate, thus it is not clear whether the two nurses were able to observe the same behavior. The variability of delirium might then explain the differences in rating the items. Also, in interviews with the nurses made clear that their ratings were influenced by their explanation of the patients' behavior. For instance, a patient was claiming that his children had rejected him since their visits had become less frequent. One nurse rated the patient as '*suspicious*', while the other sympathized with the patient and did not rate him as

suspicious. The phenomenon of nurses normalizing their patients' behavior in giving explanations when asked to rate behavior also is found in a study of Morency et al. (6).

The ease of use of the DOS Scale was evaluated to be high. A few items gave difficulty in rating. Of the respondents, 47% stated that some items could be interpreted in several ways, though 49% did not report any difficulty. From the pair wise ratings and the interviews, it became clear that the items with which the nurses had most problems were the less valid and reliable observations. For example, two items on disturbed thinking (*talks and answers slowly* and *is suspicious*) were very difficult to rate. Ratings were influenced by the nurse's approach to the patient and by the fact that the observation of the patient's behavior could be interpreted in different ways. In the interviews the nurses noted that the observation '*responds and answers slowly*' had to be compared with previous experiences with the patient.

In this study the good face validity of the DOS Scale was confirmed, it supports clinical observations and a minimum of specific knowledge is needed.

The results of this study show that the DOS Scale is suitable for observation of delirium symptoms by nurses. Further refinement of some DOS Scale items is recommended. Prospective studies with different groups of patients could help to further develop the DOS Scale to a valid and reliable instrument for early recognition of delirium by nurses.

Acknowledgements

The authors would like to thank all patients and nurses who participated in the study as well as the psychiatrists of the Psychiatric Consultation Service of the Nijmegen University Hospital for their contribution to this study.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th edition ed. Washington D.C.: American Psychiatric Association; 1994.
2. Lipowski ZJ. Delirium: acute confusional states: Oxford University Press; 1990.
3. Palmateer LM, McCartney JR. Do nurses know when patients have cognitive deficits? *Journal of Gerontological Nursing* 1985;11(2):6-16.
4. Foreman MD. Acute confusional states in hospitalized elderly: a research dilemma. *Nursing Research* 1986;35(1):34-38.
5. Brady PF. Labeling of confusion in the elderly. *Journal of Gerontological Nursing* 1987;13(6):29-32.
6. Morency CR, Levkoff SE, Dick KL. Research considerations delirium in hospitalized elders. *Journal of Gerontological Nursing* 1994;20(8):24-30.
7. Schuurmans MJ. Vroegtijdige signalering van het delier door verpleegkundigen. In: *Onderzoek in de verpleging en verzorging in relatie tot de praktijk*; Maastricht: De Tijdstroom/LCVV; 1996. p. 55-60.
8. Lynn MR. Determination and quantification of content validity. *Nursing Research* 1986;35(6):382-385.
9. Robertsson B, Karlsson I, Styrod E, Gottfries CG. Confusional State Evaluation (CSE): an instrument for measuring severity of delirium in the elderly. *British Journal of Psychiatry* 1997;170:565-570.
10. Williams-Russo P, Urquhart BL, Sharrock NE, Charlson ME. Post-operative delirium: predictors and prognosis in elderly orthopedic patients. *Journal of the American Geriatric Society* 1992;40:759-767.

chapter 6

Early symptoms of delirium observed by nurses in patients with a hip fracture

Marieke J. Schuurmans

A. Rogier T. Donders

Lillie M. Shortridge-Baggett

Sijmen A. Duursma

Submitted

Abstract

This prospective study in elderly patients with a hip fracture was designed to investigate the possibility to come to the diagnosis of delirium with a limited number of standard observations by nurses. Six days from admission, nurses rated their observations of 92 patients, per shift on the Delirium Observation Screening (DOS) Scale, a 25 item observational scale developed for early recognition of delirium during regular care. Eighteen patients developed delirium as diagnosed by a geriatrician using the Diagnostic Statistical Manual IV criteria. The nurses' ratings of patients before diagnosed as delirious were compared with the ratings of non-delirious patients. Significant differences were found on nearly all items in the three shifts. Ratings of patients known with psychiatric comorbidity showed some significant differences with those of patients before diagnosed as delirious. Based on these findings a reduction of items took place and an algorithm was developed. The selected items were all related to diagnostic symptoms of delirium. With the final 13 items a sensitivity of 100% and a specificity of 68% was found after 2 days of rating.

Introduction

Delirium is one of the most frequent forms of psychopathology in elderly patients. The diagnosis is often overlooked or mistaken for dementia or depression (1-6). The diagnostic criteria are clearly described in the Diagnostic Statistical Manual (DSM) of the American Psychiatric Association (7), however, presentation of symptoms differs from patient to patient (7-10). Symptoms of delirium fluctuate and tend to be stronger in the evening and at night. The diagnosis of delirium is a primarily clinical, based on careful observation and awareness of the diagnostic criteria (11). Effective case finding should be the first step in the improvement of care of delirious patients (12). Due to frequent, around the clock, contact with patients nurses are in a strategic position to observe onset of delirium in hospitalized elderly patients (13-15). Studies in which trained (research) nurses used standardized observations showed promising results with regard to early recognition of delirium (16-18). The aim of this study is to investigate the possibility to come to the diagnosis of delirium with a limited number of standard observations by nurses. Nurses' observations during regular care in day shift, evening and night shift will be compared to study the usefulness of more than one assessment per day. Since delirium is often confused with psychiatric problems, ratings of patients before diagnosed as delirious will be compared with ratings of patients with psychiatric problems.

Methods

Design and sample

In order to study observation of symptoms preceding a diagnosis of delirium a prospective design in a high risk population was chosen. Patients with a hip fracture are known to be at risk for delirium (19-24). During 20 month data were collected on patients over 70 years of age, admitted with a hip fracture in a medium sized general hospital in the center of the Netherlands. Patients were excluded from the study when they were suspected to have delirium on admission, when transferred to the ICU after surgery or when discharged less than 5 days after surgery. The study was approved by the hospital ethics committee. All patients gave informed consent; in case of cognitive impairment, surrogate consent was obtained from a close relative.

Procedure

Data collection started at the moment of admission and continued for a maximum of 6 days. During regular care nurses observed patients for early symptoms of delirium three times a day. The nurses recorded their observations at the end of each shift (day, evening and night shift) on the Delirium Observation Screening (DOS) Scale (25). A research nurse collected the nurses' observations and visited the patients on a daily base. In case of alterations of behavior a geriatrician from the University Medical Center Utrecht visited the patient. The diagnosis of delirium was made by a geriatrician based on the DSM-IV criteria (7). Each diagnosis was discussed in a group of geriatricians. Demographic data (age, gender, living conditions) and data with regard of the pre-hospital condition, sensory deficits, Activities of Daily Living (ADLs) measured by Barthel Index (26, 27), cognitive status by the Informant Questionnaire of Cognitive Decline in the Elderly (IQCODE) (28-30), a pre-existing psychiatric diagnosis (including dementia), comorbidity and medication use of the patients were collected from an interview with the patient and his or her family and medical and nursing records. Time between the fall and hospital admission and between hospital admission and surgery were recorded.

Measures

The DOS Scale has been developed based on the DSM-IV criteria for delirium, literature review and clinical experience. Experts reviewing the DOS Scale using the method described by Lynn (31), reported good content validity. The scale contains 25 observational items related to the symptoms (diagnostic and related) described in the DSM-IV criteria of delirium: disturbances of consciousness (item 1-3); disturbances of attention/concentration (item 4-6); disturbances of thinking (item 7-11); disturbances of memory or orientation (item 12-14);

disturbances of psychomotor activity (item 15-18); disturbances of sleep-wake pattern (item 19-21); disturbances of mood (item 22-23); and disturbances of perception (item 24-25). Items are rated on a 4-point Likert scale based on the frequency of occurrence of the behavioral disturbances (1= never, 2=sometimes, 3=usually, 4=always). A score of 1 is defined as 'normal behavior', meaning absence of behavioral alterations. Four items (1, 5, 13, 14) are reverse scored, i.e., 'normal behavior' is rated as 'always'. A score for 'unable' is given in cases where the nurse is not able to rate a behavior because of patient characteristics (for instance, no verbal contact possible, patient is asleep) or nurse characteristics (knowledge deficit or inexperience with type of behavior). The DOS Scale is designed for nurses with basic knowledge of geriatric care, to observe the alterations of behavior during regular care and to be rated in a few minutes. The scale has been tested for reliability and validity in different samples of patients (32-34). The DOS Scale showed high internal consistency (Cronbach's $\alpha > 0.90$) in all studies done. Determination of interrater reliability of this scale was not easy since ratings are based on nurses' observations during regular care and symptoms of delirium fluctuate. Different studies were done, one with use of a video tape, and interrater reliability was satisfactory. Predictive validity, against a DSM-IV diagnosis by a geriatrician or psychiatrist, showed significant differences between delirious and non-delirious patients in all samples of patients. Correlation with the Mini Mental State Examination (MMSE) was good ($R_s = 0.66$ and $R_s = 0.79$ ($p < 0.001$)). Correlation of the DOS Scale with the pre-hospital condition of the patient (cognitive impairment measured by the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE), psychiatric diagnosis and ADLs measured by a Barthel Index) was moderate to good.

Statistical analysis

Data were analyzed using SPSS. Measures used to get insight in nurses' ratings of patients without delirium and nurses' ratings of patients later diagnosed as delirious were:

- mean of the item ratings before delirium was diagnosed
- variances of the item ratings before delirium was diagnosed
- proportion of ratings that indicate altered behavior: the number of ratings with a score higher than 1 (and in items 1, 5, 13 and 14 a score less than 4) was divided by the total number of ratings per patient. A score higher than 1 indicates that the symptom is observed (2=sometimes, 3=usually, 4=always). Dividing the number of ratings that indicate presence of symptoms by the total number of ratings gave a score between 0 and 1 representing the proportion of ratings indicating alteration of behavior.

These measures were calculated on all ratings and on the ratings per shift (day, evening, night). Comparison of these measures was done using the Mann-Whitney U test, since measures were not normally distributed. To examine the possibility of misclassification due to

psychiatric diseases, like dementia, ratings were compared with the Mann-Whitney U test between patients with psychiatric comorbidity without delirium and all patients later diagnosed as delirious.

In addition, because the planned analysis demonstrated that the DOS Scale differentiated significantly between patients later diagnosed as delirious and non-delirious patients, we limited the number of items and calculated sensitivity and specificity coefficients. In order to calculate sensitivity and specificity of the scale an algorithm was developed. Items were selected using the following criteria:

- items show a significant difference of > 0.50 in mean value and/or
- items show a significant difference of > 0.50 in mean variance and/or
- items show a proportion of altered behavior in which the non-delirious patients rate between 0.00 and 0.10 or the later diagnosed as delirious patients rate > 0.70 and/or
- items show a significant difference between patients later diagnosed as delirious and patients with other psychiatric symptoms.

To be selected, items had to score at least on two criteria and in more than one shift. In the selection process the clinical relevance of items was taken in account. Of the selected items the discriminative value of mean ratings and the range of ratings over 3 shifts (day, evening and night) was studied. For the three shifts per day the mean and the range of ratings on each item were calculated. An algorithm was developed based on three ratings per day. Items with a mean rating higher than 1 (or lower than 4 in the reverse rated items) or a range with more than one scale step, were given one point. Summing these points enabled to calculate sensitivity and specificity given a cut off score. If a patients' score was equal to or larger than the cut off score, the patient was labeled 'diagnosed as delirious by the DOS Scale'. To simulate a sequential application of the test, scores were calculated for each day, only for patients who were not labeled 'diagnosed as delirious by the DOS Scale' in the days before. The cumulative total of patients labeled 'diagnosed as delirious by the DOS Scale' on for example day 2, is the sum of numbers of patients labeled 'diagnosed as delirious by the DOS Scale' on day 1 and day 2. To calculate sensitivity and specificity against the DSM-IV diagnosis of the geriatrician, we used either the patients on day 1 labeled 'diagnosed as delirious by the DOS Scale', or the cumulative total of patients labeled 'diagnosed as delirious by the DOS Scale' on a number of days.

Results

Data were collected on 92 patients. Seventy eight percent of the patients were female, their mean age was 82.3 years (range from 70 to 98, sd 6.65) and most patients were living in the community single (42.2% in the community single; 22.8% in the community with partner;

23.9% residential home; 7.6% nursing home; 3.3% other). Most patients were relatively ADL independent before their hip fracture (Barthel Index score range 6-20, mean 16, sd 4.66), the mean number of comorbid diseases was 2.8 (range 0-7, sd 1.57) and the mean number of medications used before admission 3.7 (range 0-15, sd 3.37). Almost 70% of the patients were admitted within 8 hours after a fall and nearly 75% had surgery in the first 24 hours after admission.

During a maximum of 18 shifts nurses rated the DOS Scale for each patient. Most patients were admitted during evening shift (62%) and day shift (22%). Missing values (not rated and 'unable' ratings) were found in all shifts. In day shift a mean of 14.4% was not rated, a mean of 5% 'unable because of patients characteristics' and a mean of 1.4% 'unable because of nurse characteristics'. In evening shift this was 24.4%, 7% and 1.5% and in night shift 24.4%, 31.4% and 2.7%. Missing ratings were due to several reasons: temporary employees, high workload or simply forgotten. There was no pattern of missing values except for the fact that most missing values were reported in evening shift and night shift. By use of mean values, variances of means and percentage of ratings that indicate altered behavior per patient, influence of missing values in the calculations was limited.

During the first week of admission 74 patients (80.4%) did not develop delirium and 18 patients (19.6%) did develop delirium (range 2-6 days, mean 3 days, sd 1.3). The DOS Scale ratings were compared between the patients without delirium and the patients later diagnosed as delirious.

The means of mean ratings and variances of ratings of patients who did not develop delirium, were compared with the means of mean ratings and variances of ratings of patients who were later diagnosed as delirious. Results are shown in Table 1.

Before being diagnosed as delirious, the group of patients who developed delirium for all items of the DOS Scale showed significant more symptoms than the other patients, except for the item *'has vivid and frightening dreams during the night'*. In the non-delirious patients the symptoms of delirium were hardly ever seen. Most means of mean ratings in the group of non-delirious patients stayed close to rating 1 (never observed during this shift), or, for the reverse-rated items (1, 5, 13, and 14) close to 4 (always observed during this shift). The means of mean ratings of patients later diagnosed as delirious show a tendency towards 2 (sometimes), but some behavior was rarely observed. The differences in mean variances were in most items significant. Ratings of patients later diagnosed as delirious showed significantly higher variances, suggesting that behavior differed more from shift to shift. Two items related to memory and orientation (*'knows which part of the day it is'*, *'remembers recent event'*), one item related to psychomotor activity (*'hardly moves'*), two items related to the sleep-wake pattern (*'is sleepy/drowsy during the day'*, *'has vivid and frightening dreams*

during the night') and the item '*seeing/hearing persons/things as something else*' did not differ significantly in variance.

To gain insight into the usefulness of the observations done during the three shifts, measures were calculated for each shift. Significance of differences between non-delirious and patients later diagnosed as delirious are shown for each shift separately in Table 2, Table 3 and Table 4.

Table 1 Comparison of mean and variance of mean DOS Scale ratings of post-operative hip fracture patients without delirium (n =74) and post-operative hip fracture patients before being diagnosed as delirious (n = 18)

Observations		Mean ratings ¹			Variance of ratings		
		Non delirious	Delirious	p-value	Non delirious	Delirious	p-value
1	Reacts normally to addressing*	3.69 (.46)	3.19 (.66)	≤ .001	.26 (.32)	.60 (.56)	.005
2	Dozes during conversation or activities	1.18 (.28)	1.68 (.57)	≤ .001	.13 (.23)	.68 (.68)	≤ .001
3	Stares into space	1.25 (.38)	1.66 (.51)	≤ .001	.16 (.24)	.55 (.65)	≤ .001
4	Is easy distracted by stimuli from the environment	1.34 (.54)	2.14 (.54)	≤ .001	.27 (.37)	.67 (.60)	≤ .001
5	Maintains attention to conversation or action*	3.41 (.69)	2.74 (.69)	≤ .001	.40 (.48)	.71 (.64)	.03
6	Does not finish question or answer	1.37 (.44)	1.89 (.39)	≤ .001	.43 (.56)	.98 (1.05)	.006
7	Gives answers which do not fit the question	1.28 (.41)	1.85 (.47)	≤ .001	.20 (.29)	.80 (.74)	≤ .001
8	Talks slowly or answers slowly	1.25 (.40)	1.58 (.52)	.01	.17 (.31)	.42 (.50)	.02
9	Reacts slowly to instructions	1.37 (.49)	1.89 (.56)	≤ .001	.22 (.30)	.68 (.64)	≤ .001
10	Speaks incoherently	1.25 (.39)	1.86 (.66)	≤ .001	.21 (.32)	.68 (.77)	≤ .001
11	Is suspicious	1.13 (.29)	1.63 (.63)	≤ .001	.15 (.36)	.42 (.52)	≤ .001
12	Thinks to be somewhere else	1.29 (.55)	2.18 (.49)	≤ .001	.19 (.36)	1.17 (1.15)	≤ .001
13	Knows which part of the day it is*	3.38 (.79)	2.45 (.86)	≤ .001	.50 (.49)	.80 (.70)	ns
14	Remembers recent event*	3.38 (.82)	2.29 (.67)	≤ .001	.42 (.46)	.76 (.89)	ns
15	Is picking, disorderly, restless	1.25 (.43)	2.04 (.56)	≤ .001	.23 (.36)	.93 (.83)	≤ .001
16	Hardly moves	1.70 (.51)	2.20 (.54)	≤ .001	.61 (.39)	.64 (.39)	ns
17	Pulls IV tubes, feeding tubes, catheters etc.	1.14 (.33)	1.59 (.54)	≤ .001	.13 (.35)	.59 (.72)	≤ .001
18	Has unnatural position in bed	1.22 (.28)	1.58 (.34)	≤ .001	.20 (.25)	.59 (.60)	≤ .001
19	Is sleepy/drowsy during the day	1.48 (.48)	1.95 (.61)	≤ .001	.35 (.32)	.51 (.49)	ns
20	(night shift) Has vivid and frightening dreams during the night	1.10 (.23)	1.17 (.32)	ns	.19 (.52)	.13 (.20)	ns
21	(night shift) Is awake/wakes up restlessly	1.22 (.38)	1.60 (.61)	.005	.23 (.47)	.76 (.78)	.004
22	Is easy or sudden emotional (frightened, angry, irritated)	1.14 (.33)	1.90 (.81)	≤ .001	.00 (.15)	.76 (1.05)	≤ .001
23	Is apathetic/weary	1.21 (.34)	1.50 (.43)	.001	.17 (.28)	.45 (.53)	.005
24	Sees persons/things as somebody/something else	1.11 (.24)	1.31 (.41)	.03	.11 (.24)	.25 (.35)	.04
25	Sees/hears things which are not there	1.07 (.16)	1.32 (.37)	≤ .001	.00 (.15)	.22 (.19)	≤ .001

¹ ratings 1= never, 2= sometimes, 3= usually, 4= always, * items are reverse rated
Tested with a Mann-Whitney U test

In the day shift, 22 items (of the 23 to be rated in day shift) differed significantly in mean values for the non-delirious group and the group later diagnosed as delirious. Eleven items differed significantly in mean variances and 22 items in the proportion of altered behavior. The three measures all showed higher ratings in patients later diagnosed as delirious. In the evening shift, 20 items (of the 23 to be rated in evening shift) showed significant differences in mean values, 14 in mean variances and 20 in the proportion of altered behavior. The three measures showed all higher ratings in patients later diagnosed as delirious. In the night shift, 14 items (of the 24 to be rated in night shift) showed significant differences in mean values, 8 in mean variances and 13 in the proportion of altered behavior. All shifts gave relevant, significant information for early recognition of delirium. The three measures showed, except for the mean variance of ratings in item 5 and 16 all higher ratings in patients later diagnosed as delirious. Item 20 (*'has vivid and frightening dreams during the night'*), which was described especially for rating in night shift, did not show much ratings indicating presence of this observation.

Observations of symptoms of delirium might be confused with observations of patients with a psychiatric comorbidity (like dementia). In this sample of patients, 27 were known with psychiatric comorbidity of which 10 patients developed delirium. The ratings of patients with a psychiatric comorbidity who did not develop delirium (n= 17) were compared with those of all patients who developed delirium (n=18). Differences were found in variance of item 2 *'dozes during conversation or activities'* (p =0.02), mean of item 4 *'is easily distracted by stimuli'* (p=0.02), variance of item 7 *'gives answers that do not fit'* (p= 0.01), mean and variance of item 15 *'is picking, disorderly, restless'* (p= 0.007 and p= 0.02), mean and variance of item 22 *'is easy or sudden emotional'* (p= 0.02 and p= 0.003) and mean and variance of item 25 *'sees/hears things which are not there'* (p= 0.03 and p= 0.05). All significant findings showed a higher rating in the delirious patients than in the patients with psychiatric comorbidity.

Since the DOS Scale differentiated well between patients later diagnosed as delirious and non-delirious patients, selection of items took place based on the criteria described earlier. The items 2, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 17, 22, 24 and 25 were selected and used to develop and test an algorithm in order to calculate sensitivity and specificity, as described in the section statistical analysis. In this process the items 11, 16 and 24 were deleted since their predictive value was low. Of the final 13 items an algorithm was tested with the ratings over one day (3 shifts). One point was given for items with a mean rating higher than 1 (or lower than 4 in the reverse rated items) or a range with more than one scale step difference. Sum of the DOS Scale algorithm has a maximum value of 11 and a minimum value of 0. For the resulting scale (see Appendix 1) a cut off point of 8 showed a 100% sensitivity and

68% specificity after two days of rating. Positive predictive value of the DOS Scale algorithm was 43%, negative predictive value of the DOS Scale algorithm was 100%. Accuracy (correctly classified delirious (18) or non-delirious (50) / all cases (92)) of the DOS Scale algorithm was 75%. The patients classified as false positive were compared with the true positive and the true negative, as is shown in Table 5.

Table 5 Characteristics of patients classified delirious false positive (n= 24) compared with patients correctly classified negative (n= 50) or positive (n= 18) by the DOS Scale algorithm with cut off point 8

	False positive	Correct negative	Correct positive
Mean age (sd)	84.3 (7.3)	81.1 (6.1)	82.6 (6.9)
Mean Barthel Index (sd)	14.4 (4.9)	17.2 (4.3)	14.6 (4.7)
<i>Living condition</i>			
Living alone	17 %	61 %	28 %
Living with partner	25 %	20 %	22 %
Residential home	46 %	10 %	33 %
Nursing home	12 %	4%	11 %
Mean IQCODE (sd) ¹	4.30 (.60)	3.47 (.45)	4.13 (.54)
Psychiatric comorbidity	41.7 %	14.3 %	55.6 %
Hearing impairment	40.9 %	24.5 %	16.7 %
Vision impairment	36.4 %	24.5 %	29.4 %
¹ IQCODE ratings were available of 71% (false positive), 52% (correct negative) and 72% (correct positive) of the patients			

The false positive patients had a high mean age, lived more often in a residential home, showed pre-admission dependency in their ADLs, pre-existing cognitive dysfunction and almost half of them had a hearing impairment.

Discussion

Early recognition of delirium is of great interest for elderly patients and for the health care system. Several studies have shown a high mortality, high morbidity, a prolonged hospital stay and nursing home placement after admission due to delirium (3, 11, 37-43). Because of its fluctuating nature and different presentations, the diagnosis of delirium is difficult to make (2, 44). Nurses and doctors often do not recognize cognitive deficits in elderly patients (4, 45-49).

Most delirium instruments are focussed on the diagnosis by a physician rather than on symptoms recognition by primary care givers. The fluctuating nature of delirium, however, limits the sensitivity of once a day interviews (50). In order to diagnose delirium recent changes in patients' observable behavior and cognitive functioning should be brought to the

attention of the physician (1). Due to their frequent contacts with patients and their spouses nurses may be more likely to recognize less severe cognitive changes (47).

The findings in this study show clearly that nurses can early observe symptoms during regular care. Nearly all symptoms described as diagnostic and related are seen in the days before the diagnosis of delirium was made. Observations done in day, evening and night shift all contributed to the recognition of delirium. This seems in contrast with the findings of Treloir and McDonald (49). In a study on recognition of cognitive impairment (dementia and delirium), they found that night nursing staff were not able to identify patients who were cognitive impaired. Treloir and McDonald tried to explain their finding by night nurses' coping with patients' behavior, high workload and lack of knowledge of the patients.

In this study six items of the DOS Scale showed significant differences between the patients with psychiatric comorbidity and patients later diagnosed as delirious. In a study of patients referred for psychiatric consultation (33) the DOS Scale significantly differentiated the delirious patients and the patients with psychiatric symptoms ($p \leq 0.001$). Seventeen of 23 DOS Scale items rated showed significant difference. The psychiatric diagnoses in the non-delirious group included major depression, adjustment disorders, post-traumatic stress disorders and somatoform disorders. Possible explanation for the different number of differentiating items in the two studies (6 versus 17) can be found in the ratings of (severely) demented patients. Distinguishing observations of (severely) demented patients and of delirious patients is difficult (51).

The study described here could be questioned with regard to the high number of missing values on the nurses' observations. The number of missing values is a result of data collection by nurses during regular care. The busyness of the situation in which nurses work filling out forms can give problems (17). The strength of this method, however, is found in the value of the results for clinical practice. The use of the mean value, variance and proportion of altered behavior limited the influence of missing values on the results of this study. The results of this study show that a limited number of standard observations of staff nurses enables the diagnosis of delirium in an early stage. The transferability of this result to daily practice is expected to be high. The number of delirious patients on which these results are based is modest, however, the results presented are significant. Significant findings have a lower likelihood in small samples; therefore these results are expected to be replicated in a larger study. Replication of this study in samples of medical elderly patients and larger samples of surgical elderly patients is recommended.

Finally a diagnosis made by a physician as 'gold standard' can be questioned with regard to reliability. In delirium research, however, there is no alternative standard. Strict use of the DSM criteria enhances the reliability. Each diagnosis made by the geriatrician in this study was discussed in a group of geriatricians and the diagnosis proposed was never rejected.

This study confirmed the possibility to contribute to the diagnosis of delirium a limited number of standard observations by nurses. The observations were done during regular care and did not require specific training of nurses. The sensitivity of the DOS Scale algorithm based on two days of rating was 100% and the specificity 68%. False positives rates were found in ratings of patients with a high age, pre-existing ADL dependency, pre-existing cognitive and sensory impairment (hearing or vision). A positive predictive value of 43 % means that less than half of the patients 'diagnosed delirious by the DOS Scale' were delirious according to the DSM-IV diagnosis. Not one of the patients who were 'diagnosed not delirious by the DOS Scale', were delirious. The DOS Scale was developed to screen for symptoms of delirium in order to contribute to the diagnosis of delirium. The screening of patients by nurses' ratings of the DOS Scale enables an early diagnosis of delirium. An effective case finding method would initiate an improvement in the quality of care (12).

Acknowledgements

This study made possible by a grant of the National Center for Nursing and Care (LCVV). The authors would like to thank all patients and their families who participated in the study and express our gratitude to all nurses who rated their observations as well as to dr. G.J. Clevers, R. Pel-Littel and the geriatricians of the department of Geriatric Medicine of the University Medical Center for their contribute to this study.

References

1. Lipowski ZJ. Delirium: acute confusional states: Oxford University Press; 1990.
2. Foreman MD. The cognitive and behavioral nature of acute confusional states. *Scholarly Inquiry for Nursing Practice: an International Journal* 1990;5(1):3-20.
3. Rockwood K, Cosway S, Stolee P, Kydd D, Carver D, Jarret P, et al. Increasing the recognition of delirium in elderly patients. *Journal of the American Geriatric Society* 1994;42:252-256.
4. Lewis LM, Miller DK, Morley JE, Nork MJ, Lasater LC. Unrecognized delirium in ED geriatric patients. *American Journal of Emergency Medicine* 1995;13(2):142-145.
5. Farrel K, Ganzini L. Misdiagnosing delirium as depression in medically ill elderly patients. *Archives of Internal Medicine* 1995;155:2459-2464.
6. Johnson J. Identifying and recognizing delirium. *Dementia and Geriatric Cognitive Disorders* 1999;10(5):353-358.
7. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
8. Williams MA, Campbell EB, Raynor WJ, Musholt MA, Mlynarczyk SM, Crane LF. Predictors of acute confusional states in hospitalized elderly patients. *Research in Nursing and Health* 1985;8:31-40.
9. Rudberg MA, Pompei P, Foreman MD, Ross RE, Cassel CK. The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing* 1997;26:169-174.
10. Sandberg O, Gustafson Y, Brännström B, Bucht G. Clinical profile of delirium in older patients. *Journal of the American Geriatric Society* 1999;47:1300-1306.
11. Inouye SK. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *The American Journal of Medicine* 1994;97:278-288.
12. Inouye SK, Schlesinger MJ, Lydon TJ. Delirium: a symptom of how hospital care is failing older persons and a window to improve quality of hospital care. *The American Journal of Medicine* 1999;106(May):565-571.
13. Foreman MD. Acute confusional states in hospitalized elderly: a research dilemma. *Nursing Research* 1986;35(1):34-38.
14. Yeaw EMJ, Abbate JH. Identification of confusion among the elderly in an acute care setting. *Clinical Nurse Specialist* 1993;7(4):192-197.
15. Morency CR, Levkoff SE, Dick KL. Research considerations delirium in hospitalized elders. *Journal of Gerontological Nursing* 1994;20(8):24-30.
16. Williams MA, Ward SE, Campbell EB. Confusion: testing versus observation. *Journal of Gerontological Nursing* 1988;14(1):25-30.
17. Neelon VJ, Champagne MT, McConnel E, Carlson J, Funk SG. Use of the NEECHAM Confusion Scale to assess acute confusional states of hospitalized older patients. In: *Key aspects of elder care; managing falls, incontinence and cognitive impairment*; 1992; 1992. p. 278-289.

18. Neelon VJ, Champagne MT, Carlson JR, Funk SG. The NEECHAM Confusion Scale: construction, validation, and clinical testing. *Nursing Research* 1996;45(6):324-330.
19. Williams MA, Campbell EB, Raynor SM, Mlynarczyk SM, Ward SE. Reducing acute confusional states in elderly patients with hipfractures. *Research in Nursing and Health* 1985;8:329-337.
20. Berggren D, Gustafson Y, Eriksson B, Bucht G, Hansson L-I, Reiz S, et al. Postoperative confusion after anesthesia in elderly patients with femoral neck fractures. *Anesthesia and Analgesia* 1987;66:497-504.
21. Gustafson Y, Berggren D, Brännström B, Bucht G, Norberg A, Hansson LI, et al. Acute confusional states in elderly patients treated for femoral neck fracture. *JAGS* 1988;36:525-530.
22. Mullen JO, Mullen NL. Hip fracture mortality. *Clinical Orthopaedics and Related Research* 1992;280:214-222.
23. Milisen K. An intervention study for delirium in elderly hip fracture patients [PhD thesis]. Leuven: Katholic University Leuven; 1999.
24. Marcantonio ER, Flacker JM, Michaels M, Resnick NM. Delirium is independently associated with poor functional recovery after hip fracture. *JAGS* 2000;48:618-624.
25. Schuurmans MJ. Vroegtijdige signalering van het delier door verpleegkundigen. In: *Onderzoek in de verpleging en verzorging in relatie tot de praktijk*; 1996; Maastricht: De Tijdstroom/LCVV; 1996. p. 55-60.
26. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Medical Journal* 1965;14:61-61.
27. Haan Rd, Limburg M, Schuling J, Broeshart J, Jonkers L, Zuylen Pv. Klinimetrische evaluatie van de Barthel-Index, een maat voor beperkingen in het dagelijks functioneren. *Nederlands Tijdschrift voor Geneeskunde* 1993;137(18):917-921.
28. Jorm AF, Jacomb PA. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine* 1989;19:1015-1022.
29. Jorm AF, Scott R, Cullen JS, MacKinnon AJ. Performance of the Informant Questionnaire on cognitive decline in the Elderly (IQCODE) as a screening test for dementia. *Psychological Medicine* 1991;21:785-790.
30. Jorm AF, Christensen H, Henderson AS, Jacomb PA, Korten AE, MacKinnon A. Informant ratings of cognitive decline of elderly people: relationship to longitudinal change on cognitive tests. *Age and Ageing* 1996;25:125-129.
31. Lynn MR. Determination and quantification of content validity. *Nursing Research* 1986;35(6):382-385.
32. Schuurmans MJ, Shortridge-Baggett LM, Duursma SA. The Delirium Observation (DOS) Scale: A screening instrument for delirium. submitted: *Research in Nursing and Health* 2001.
33. Roodbol G, Schuurmans MJ, Gamel CJ, Frederiks C, Shortridge-Baggett LM. Recognition of delirium by standard nursing observation. submitted: *International Journal of Nursing Studies* 2001.

34. Schuurmans MJ, Shortridge-Baggett LM, Duursma SA. The troublesome issue of interrater reliability: A study on nurses' observation of the symptoms of delirium. submitted: *Journal of Nursing Measurement* 2001.
35. Folstein MG, Folstein SE, McHugh PR. "Mini-Mental State" a practical guide for grading the cognitive state of patients for clinicians. *Journal of Psychiatric Research* 1975;12:189-198.
36. Collin C, Wade DT, Davies S, Horne V. The Barthel Index: a reliability study. *Int Disabil Studies* 1988;10:61-63.
37. Lipowski ZJ. Delirium in the elderly patient. *New England Journal of Medicine* 1989;320(2):578-582.
38. Brännström B. Care of the acutely confused elderly hipfracture patient [PhD dissertation]. Umea: University of Umea; 1991.
39. Foreman MD, Grabowski R. Diagnostic dilemma: cognitive impairment in the elderly. *Journal of Gerontological Nursing* 1992;18(9):5-12.
40. Foreman MD. Acute confusion in the elderly. *Annual Review of Nursing Research* 1993;11:3-30.
41. Pompei P, Foreman M, Rudberg MA, Inouye SK, Braund V, Cassel CK. Delirium in hospitalized older persons: outcomes and predictors. *Journal of the American Geriatric Society* 1994;42:809-815.
42. O'Keeffe S, Lavan L. The prognostic significance of delirium in older hospital patients. *Journal of the American Geriatric Society* 1997;45:174-178.
43. Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P. Does delirium contribute to poor hospital outcomes. *Journal of General Internal Medicine* 1998;13:234-242.
44. Rasin JH. Confusion. *Nursing Clinics of North America* 1990;25(4):909-918.
45. Palmateer LM, McCartney JR. Do nurses know when patients have cognitive deficits? *Journal of Gerontological Nursing* 1985;11(2):6-16.
46. Morency CR. Mental status change in the elderly: recognizing and treating delirium. *Journal of Professional Nursing* 1990;6(6):356-365.
47. Francis J. When do physicians and nurses recognize and document delirium? *Journal of the American Geriatric Society* 1992;40 (supplement 10):SA 10.
48. Bowler C, Boyle A, Branford M, Cooper SA, Harper R, Lindesay J. Detection of psychiatric disorders in elderly medical inpatients. *Age and Ageing* 1994;23:307-311.
49. Treloar AJ, Macdonald AJD. Recognition of cognitive impairment by day and night nursing staff among acute geriatric patients. *Journal of the Royal Society of Medicine* 1995;88:196-198.
50. Marcantonio ER, Goldman L, Mangione CM, Ludwig LE, Muraca B, Haslauer CM, et al. A clinical prediction rule for delirium after elective noncardiac surgery. *JAMA* 1994;271(2):134-139.
51. O'Keeffe ST. Rating the severity of delirium: the Delirium Assessment Scale. *International Journal of Geriatric Psychiatry* 1994;9:551-556.

chapter 7

Elderly patients with a hip fracture: The risk for delirium

Marieke J. Schuurmans

Sijmen A. Duursma

Lillie M. Shortridge-Baggett

Geert Jan Clevers

Ruth E. Pel-Littel

Submitted

Abstract

This prospective study is designed to investigate risk factors for delirium in elderly hip fracture patients that can be recognized by nurses. Of 92 elderly patients with a hip fracture data were collected on predisposing and precipitating factors of delirium. Predisposing factors included age, gender, sensory impairments, functional impairment before the hip fracture, residency before admission, pre-existing cognitive impairment, comorbidity and medication use. Precipitating factors included factors related to surgery and to the postoperative period. Factors related to surgery studied were time between admission and surgery, type of surgery, type of anesthesia, duration of surgery and anesthesia and complications during surgery. Factors studied in the postoperative period were slow recovery, malnutrition, dehydration, addition of three or more medications, introduction of bladder catheter, infections, complications and falls, and use of morphine. Eighteen patients developed delirium as diagnosed by a geriatrician using the DSM-IV criteria. Data of delirious patients were compared with those of non-delirious patients.

The findings of this study confirm that elderly hip fracture patients with premorbid ADL dependency, psychiatric comorbidity (including dementia) and a high number of comorbid problems are at risk for the development of delirium. Based on these findings it is recommended that nurses should assess patients' pre-fracture functional and cognitive capacities in an early stage of the hospital stay. Nurses, however, should be alert to postoperative delirium in "healthy elderly" patients. Monitoring of symptoms postoperatively in all elderly patients is advised.

Introduction

Delirium, or acute confusion is a transient organic mental syndrome characterized by reduced level of consciousness, reduced ability to maintain attention, perceptual disturbances and memory impairment. The syndrome develops in a short period of time and symptoms fluctuate (1, 2). Elderly hip fracture patients are vulnerable to developing delirium postoperatively; incidences of delirium in these patients range from 10% to 61% (Table 1). All studies, except two, reported here showed incidences over 20% and the mean incidence in all studies was 25% (2503 patients, 618 diagnosed delirious). Differences can be explained in part by study design (prospective versus retrospective), diagnostic criteria (inclusion of pre-operative delirium or not) and characteristics of the study sample (age, medical and psychological condition).

Table 1 Delirium in hip-fracture patients, review of studies

Study	Sample	Diagnosis	Incidence	Note
Williams et al. 1985 (3)	170 ≥ 60 years (mean 78.8 years) 57 ≥ 60 years (mean age 80.6 years)	Research based criteria	52 % 44 %	Intervention study: specific interpersonal and environmental nursing actions
Berggren et al. 1987(4)	57 > 64 years,	DSM-III-criteria	44 %	No demented patients included
Gustafson et al. 1988(5)	111 > 65 years (mean: 79.3 years)	DSM-III-criteria	61 %	33% delirium pre-operative which continued post-operative, 28% delirium post-operative
Brännström et al. 1991(6)	35 mean 78.2 years	DSM-III-criteria	42 %	Postoperative delirium
Gustafson et al. 1991 (7)	103 age > 65 years (mean: 79.5 years)	DSM-III-criteria	48 %	
Mullen and Mullen 1992 (8)	400 age > 60 years	Research based criteria	25 %	
Bowman 1997(9)	17 mean 80 years	DSM-III	47 %	
Edlund et al. 1999 (10)	54 mean 77.1 years (range 40-98years)	DSM-III-R criteria	28%	19% delirium pre-operative which continued post-operative, 9 % delirium post-operative
Milisen 1999 (11)	60 median age 80 60 median age 82	CAM (DSM-III-R criteria)	23 % 20 %	Intervention study: an evidence-based interdisciplinary intervention program to prevent and reduce delirium
Marcantonio et al. 2000 (12)	126 > 65 years (mean 79 years)	CAM (DSM-III-R criteria)	41 %	
Brauer et al. 2000 (13)	571 > 50 years, median 85 years (range 69-101)	CAM (DSM-III-R criteria)	10%	4.4 % delirium pre-operative 5.1% delirium post-operative
Dolan et al. 2000 (14)	682 > 65 years	CAM by proxi	14 %	Pre-operatively delirium, no demented patients were included

Delirium in hip fracture patients is associated with decline in cognitive function, decline in activities of daily living (ADLs), decline in ambulating, new nursing home placement, more postoperative complications and death (4-6, 8, 10, 12) . Delirious hip fracture patients need intensive nursing care and many experience prolonged hospital stay. Declines in ADLs and mobility remain significant after 6 months (5, 6, 10, 12, 14).

The cause of the delirium in hip fracture patients is not always clear. Brauer et al. (13) found in 54 delirious hip fracture patients a definite cause in 7% and a possible cause in 11%; in addition, 61% of the delirious patients had one or more comorbid conditions. In a study of 400 consecutive hip fracture patients, Mullen and Mullen (8) found that delirium (25% of all patients) in 94% was associated with a medical complication. In 79%, delirium was the presenting symptom of the complication. The main complications were pneumonia,

medication side effects, stroke, sepsis, urinary infection, diabetic coma, anemia, pain and congestive heart failure. These complications are comparable to causes of delirium in studies of elderly hospitalized patients (2, 15-17). Mullen and Mullen conclude that deterioration in mental status post-operatively is primarily a symptom of an organic complication that needs to be diagnosed. This view was supported by Perez (18) who concluded that medical evaluation is imperative to recognize and correct the cause of post-operative delirium in hip fracture patients. Feldt and Griffin (19) found more urinary tract infections and a shorter period of pain management in delirious hip fracture patients compared with non-delirious patients.

Inouye et al. (20, 21) have divided risk factors for delirium in predisposing and precipitating factors. Predisposing factors are present at the time of the hospital admission and represent the baseline vulnerability of the patient. Precipitating factors are noxious insults or hospital related factors that contribute to the development of delirium. Inouye et al. suggest that patients with high baseline vulnerability to delirium may develop delirium with any precipitating factor, even one of mild degree. On the other hand, patients with low vulnerability may be resistant to delirium even with noxious insults.

Identified predisposing factors in hip fracture patients include age (3, 12, 14), level of pre-injury activity (3), dependency in activities of daily living (ADLs) (12, 14), pre-fracture cognitive impairment (10, 12), history of depression (4), high medical comorbidity (12, 14), use of drugs with anticholinergic effects (4, 10) and nursing home residence (12).

Identified precipitating factors include prolonged waiting time for surgery (10), narcotics (3), lack of post-operative mobility (3), pain (3) and problems in pain management (19). Anesthetic technique was not found to have any influence (4). In general medical elderly patients five independent precipitating factors were found to predict the development of delirium: immobility, malnutrition, more than three medications added, use of bladder catheter and any iatrogenic event during hospitalization (21).

Risk factors for delirium in surgical elderly patients in general are the same as for patients with hip fracture (22). Additional factors found are a history of alcohol abuse (23, 24), abnormal serum biochemistry values (23), type of surgery (23) female gender (25), low intraoperative haemoglobeline value (25), intraoperative blood loss, post-operative blood transfusions and lower post-operative haematocrit (24).

Because interventions to reduce or prevent delirium in hip fracture patients are likely to be most effective in the pre- and immediate post-operative periods (26), assessment of risk factors should be incorporated in pre-operative evaluation of patients (12). Therefore this study was designed to answer to following question: Is identification of risk factors useful for early recognition of delirium by nurses? Condition for the study design was that findings could be incorporated into nursing admission history and bedside screening.

Methods

Design and sample

This study had a prospective design. Subjects were patients consecutively admitted with a hip fracture to the Diaconessenhuis Utrecht, a medium-sized general hospital in the center of the Netherlands. Patients were included in the study if they were at least 70 years of age, were admitted with a traumatic hip fracture and were admitted to a surgical ward after surgery. All patients gave informed consent; in case of cognitive impairment, surrogate consent was obtained from a close relative. Patients who were suspected to have delirium on admission and patients who were discharged in less than 5 days after surgery, were excluded. The medical ethics committee of the hospital approved the study.

Procedure

Patients were included from May 1998 to December 1999. Data were collected by a research nurse and by the nurses in the surgical department. Most patients directly went from the emergency room to surgery; they were included in the study immediately after surgery. Nurses in the surgical ward observed patients for the first 6 days after admission during every shift, to look for symptoms of delirium and registered their observations on the Delirium Observation Screening (DOS) Scale (27). The DOS Scale is a 25 item observation rated screening scale developed for early recognition of symptoms. The research nurse collected the nurses' observations and visited the patients on a daily basis during week time. In case of alterations in behavior, a geriatrician from the University Medical Center in Utrecht visited the patient. A diagnosis of delirium was made by the geriatrician based on the Diagnostic Statistical Manual-IV criteria (28). Each diagnosis was discussed in a group of geriatricians.

Data on predisposing and precipitating factors were collected by a research nurse, based on an interview with the patient and/or his relatives, medical and nursing records and, in case of incomplete or unclear data, additional information from nurses and surgeons.

Measures

Data collected from the interview with the patient and or family included age, gender, impairment of speech, hearing or vision, functional impairment before the hip fracture, residency before admission, pre-existing cognitive impairment and time between fall and hospital admission. Functional impairment before the hip fracture was measured by the Barthel Index (29, 30), which measures dependency in ADLs. Pre-existing cognitive impairment was measured by the Informant Questionnaire of Cognitive Decline in the Elderly

(IQCODE) (31-33), a 26 item family member rated instrument that screens for cognitive decline. Data on comorbidities, medications, time between admission and surgery and the surgery itself were collected from medical records and discharge letters.

Data were collected on the following precipitating factors: slow recovery, resulting in diminished regain of ADL and mobility, malnutrition, dehydration, addition of three or more medications, introduction of a bladder catheter, infections, complications and fall and use of morphine in the first 6 days of admission. These data were collected from nursing and medical records.

Statistical analysis

All data were analyzed using SPSS. Presence of risk factors in non-delirious patients and delirious patients was compared by Chi-square tests and Mann-Whitney U tests.

Results

During this 20-month study, 92 patients who met the inclusion criteria were enrolled. Eighteen patients (19.6%) developed delirium within the first six days of admission (range 2-6 days, mean 3 days sd 1.3). Table 2 shows the characteristics and Table 3 the comorbidities and medication use of the delirious and non-delirious patients.

Both groups had a high age (mean 82.2 years, sd 6.8 and 82.6 years, sd 6.9) and were mainly women (86.5% - 86.9%). In the non-delirious group most patients (> 50%) had no sensory problems, were independent in ADLs (Barthel \geq 18), lived in the community and did not suffer from serious cognitive decline (IQCODE $<$ 4). The group of delirious patients was more often dependent in ADLs, lived more frequently in a residential home or nursing home and suffered more frequently from serious cognitive decline. The differences, however, between the delirious and the non-delirious groups were not significant except for some of the ADLs and Barthel Index total score.

A high percentage, however, of missing data on the IQCODE occurred, especially in the group of non-delirious patients. Missing data were more common among patients living alone, who were more likely to have no close relative, able to rate the IQCODE.

Table 2 Characteristics on admission of non-delirious (n =74) and delirious (n =18) patients hip fracture patients

	Non-delirious	Delirious	p-value
<i>Age</i>	70 - 98	71 - 95	
70-80	26 (35.1%)	5 (27.7%)	
80-90	38 (51.4%)	9 (50.0%)	
>90	10 (13.5)	4 (22.2%)	
Mean (sd)	82.2 y (6.6)	82.6 y (6.9)	ns
<i>Gender</i>			
Male	10 (13.5%)	2 (11.1%)	
Female	64 (86.5%)	16 (88.9%)	ns
<i>Sensory problems</i>			
Hearing	22 (29.7%)	3 (16.7%)	ns
Vision	20 (27.0%)	5 (27.8%)	ns
Speech	6 (8.1%)	3 (16.7%)	ns
<i>ADL support needed</i>			
Bathing	29 (39.2%)	12 (66.7%)	.04
Hair/tooth care	11 (14.9%)	4 (22.2%)	ns
Clothing	25 (33.8%)	10 (55.6%)	ns
Toilet	19 (25.7%)	8 (44.4%)	ns
Feeding	11 (14.9%)	3 (16.7%)	ns
Faeces incontinence	6 (8.1%)	5 (27.7%)	.02
Urine incontinence	13 (17.6%)	7 (38.9%)	ns
Transfer	15 (20.3%)	5 (27.7%)	ns
Mobility	14 (18.9%)	4 (22.2%)	ns
Stairs	38 (51.4%)	13 (72.2%)	ns
<i>Barthel Index total</i>			
18 – 20	43 (58.0%)	5 (27.7%)	
0 -17	29 (39.2%)	13 (72.2%)	.02
Unknown	2 (2.7%)		
<i>Living</i>			
Living Alone	34 (45.9%)	5 (27.8%)	
Living with partner	17 (23.0%)	4 (22.2%)	
Residential home	16 (21.6%)	6 (33.3%)	
Nursing home	5 (6.8%)	2 (11.1%)	
Other	2 (2.7%)	1 (5.6%)	ns
<i>IQCODE</i>			
3.0 – 3.49	19 (43.2%)	2 (15.4%)	
3.5 – 3.99	6 (13.6%)	2 (15.0%)	
4.0 – 4.49	11 (25.0%)	5 (38.5%)	
4.5 – 5.0	8 (18.2%)	4 (30.8%)	
Unknown	30 (40.5%)	5 (27.8%)	
Mean (sd)	3.8 (.66)	4.1 (.54)	ns
All tests were Chi-square tests, except for age, living and IQCODE. These were compared by a Mann-Whitney U test			

Table 3 Comorbidity of non-delirious (n=74) and delirious (n=18) hip fracture patients

	Non-delirious	Delirious	p-value
<i>Comorbidity</i>			
Haematological	4 (5.4%)	1 (5.6%)	ns
Neurological	9 (12.2%)	4 (22.2%)	ns
Motion	27 (36.5%)	9 (50.0%)	ns
Cardiovascular	35 (47.3%)	10 (55.5%)	ns
Pulmonal	19 (25.7%)	3 (16.7%)	ns
Gastro-intestinal	26 (35.2%)	6 (33.3%)	ns
Uro-genital	22 (29.7%)	8 (44.4%)	ns
Endocrinological	10 (13.5%)	1 (5.6%)	ns
Skin	4 (5.4%)	3 (26.7%)	ns
Psychiatric	17 (23.0%)	10 (55.5%)	.006
<i>Number of comorbid problems</i>			
0	4 (5.4%)		
1 – 3	55 (74.3%)	11 (61.1%)	
4 – 7	15 (20.3%)	7 (38.9%)	
Mean (sd)	2.6 (1.6)	3.4 (1.4)	.02
<i>Number of medications</i>			
Range	0 –13	0 – 15	
0	14 (18.9%)	2 (11.1%)	
1 – 2	19 (25.6%)	5 (27.8%)	
3 – 5	28 (37.8%)	3 (16.7%)	
6 – 10	10 (13.5%)	4 (22.2%)	
>10	3 (4.1%)	3 (16.7%)	
Mean (sd)	3.4 (3.1)	5.1 (4.4)	ns
Use of psycho-active drugs*	27 (36.5%)	10 (55.6%)	ns
* hypnotics, sedatives, anxiolytics, antipsychotics, lithium, anti-depressants All tests were Chi-square tests, except for number of comorbidities and medications these were compared by a Mann-Whitney U test			

The delirious and non-delirious patients did not show much difference in comorbidity before the fracture, except for psychiatric comorbidity ($p = 0.006$) and the total number of comorbid problems ($p = 0.02$). Although delirious patients used more medications (mean 5.1, sd 4.4 versus 2.6, sd 1.6) as well as more psycho-active drugs (55.6% versus 36.5%) before their hip fracture, the differences were not significant. The number of medications were correlated with the number of comorbidity problems ($R_s: 0.63, p < .01$). The number of medications were correlated with the use of psycho-active medication ($R_s: 0.68, p < .01$). Table 4 shows the time between fall, hospital admission and surgery and aspects related to the surgery.

Table 4 Factors related to the hip fracture and surgery in non-delirious (n=74) and delirious (n=18) hip fracture patients

	Non-delirious	Delirious	p-value
<i>Time between fall and hospital admission</i>			
< 8 hours	53 (71.6%)	12 (66.7%)	
8-16 hours	2 (2.7%)	1 (5.6%)	
16-24 hours	4 (5.4%)	3 (16.7%)	
> 24 hours	10 (13.5%)	1 (5.6%)	ns
unknown	5 (6.8%)	1 (5.6%)	
<i>Time between admission – surgery</i>			
0 - 12 hours	36 (48.6%)	10 (55.6%)	
12 – 24 hours	28 (37.8%)	5 (27.8%)	
24 – 48 hours	5 (6.8%)		
48 – 72 hours	4 (5.4%)	3 (16.7%)	ns
<i>Type of surgery</i>			
Hipprothesis	6 (8.1%)	4 (22.2%)	
Osteosyntheses	4 (5.4%)		
Gammanail	49 (66.2%)	13 (72.2%)	
Dhs	11 (14.9%)	1 (5.3%)	ns
Unknown	4 (5.4%)		
<i>Type of anesthesia</i>			
Local	3 (4.1%)		
Spinal	70 (95.9%)	18 (100%)	
Total			ns
<i>Duration surgery</i>			
< 45 min	18 (24.3%)	1 (5.6%)	
45-90 min	45 (60.8%)	16 (88.9%)	
90-120 min	9 (12.2%)	1 (5.6%)	
> 120 min	2 (2.7%)		ns
<i>Duration anesth.</i>			
41 – 199 min		38 – 94 min	
0 – 60 min	23 (31.1%)	4 (22.2%)	
61 - 90 min	26 (35.1%)	12 (66.7%)	
> 90 min	22 (29.7%)	2 (11.1%)	ns
<i>Bloodpressure drop*</i>			
> 10 minutes	12 (16.2%)	1 (5.6%)	ns
<i>Bloodloss > 300 ml</i>			
	5 (6.8%)		-
* systolic bloodpressure ↓ 20 mmHg All tests were Mann-Whitney U tests, except for type of anesthesia, bloodpressure and blood loss these were compared with a Chi-square test			

Most patients were admitted within 8 hours after their fall and had surgery within one day. Of the patients who had to wait over two days before surgery 43% developed a delirium. This finding was, however, not significant (time between admission and surgery < 48 hours, compared with > 48 hours, X^2 p-value 0.14). Most fractures were repaired with a gamma nail, although in the delirious group one out of five patients received a hip prosthesis. Nearly all patients had surgery under spinal anesthesia. More non-delirious patients had longer

duration of anesthesia and surgery of over 90 minutes or a drop in blood pressure compared to delirious patients, but none of these differences was significant.

In the post-operative period no precipitating factor differentiated significantly between the delirious (before diagnosis) and non-delirious patients, as is shown in Table 5.

Table 5 Post-operative factors in non-delirious (n=74) and delirious (before diagnosed as) (n=18) and hip fracture patients

	Non-delirious	Delirious	p-value
No regain ADL	13 (17.6%)	4 (22.2%)	ns
Immobility	17 (23.0%)	1 (5.6%)	ns
Malnutrition	3 (4.1%)	-	ns
Dehydration	9 (12.2%)	1 (5.6%)	ns
> 3 medications ad.	26 (35.1%)	5 (27.8%)	ns
Bladdercatheter	62 (83.8%)	17 (94.4%)	ns
Morphine use	37 (78.7%)*	13 (93.0%)*	ns
Wound infection	-	-	ns
Airway infection	3 (4.1%)	-	ns
Urinary tract infect	3 (4.1%)	2 (11.1%)	ns
Complications (other)	26 (35.1%)	9 (50.0%)	ns
Falls	2 (2.7%)	2 (11.1%)	ns
* data on morphine use were not collected in the first 30 patients All tests were Chi-square tests			

High percentages of complications other than infections and falls were found in both groups (35.1%-50.0%): they included cardiopulmonary problems (9 patients), low hemoglobin level (11) followed by blood transfusion (8 patients), pressure ulcers (6), gastro-intestinal problems (3) and other problems (6) such as insufficient controlled diabetes mellitus and blood in urine.

Discussion

This study was undertaken to answer the question: Is identification of risk factors useful for early recognition of delirium by nurses? In this prospective study data were collected on 92 hip fracture patients. Compared to other studies on delirium in hip fracture patients, the percentage of delirium is below average, with a high mean age of patients. The predisposing factors studied showed that patients with psychiatric comorbidity and patients with a high number of comorbid problems are more at risk to develop a post-operative delirium. These findings confirm that nurses should be aware of the development of delirium especially in the older old (> 80 years) with premorbid psychiatric problems (including cognitive decline) and a high number of comorbid problems (3, 4, 10, 12, 14). Pre-fracture dependency in ADLs showed significant a difference, confirming that pre-fracture ADL dependency is also a predisposing risk factor for delirium (12, 14).

With regard to the precipitating factors no factors related to the surgery and the post-operative recovery were found to differentiate between the patients later diagnosed delirious and the non-delirious patients. Nevertheless, 43 % of the patients who waited over 48 hours before surgery developed delirium. This suggests that a prolonged waiting time may be a precipitating factor, as found in a study of Edlund et al. (10). In most patients there was a limited time between fall and hospital admission, and between hospital admission and surgery. None of the patients underwent total anesthesia.

In this study, however, two women of over 80 years of age in good shape (Barthel Index score: 20), living in the community, with no previous cognitive decline or psychiatric condition, became delirious. No clear differences were found between the two delirious women and the other "healthy elderly" (n=30). "Healthy elderly" is here defined as: Barthel Index score ≥ 18 , living condition in the community, IQCODE < 4 and/or no psychiatric comorbidity. The two patients had two comorbid conditions and took two or three medications. One of them had premorbid heart failure and the other was found short of breath with a high pulse, when the delirium became clear. This means that even with the "healthy elderly" nurses should be aware of the possible development of post-operative delirium. Close monitoring for symptoms in the post-operative days is recommended.

Limitations of this study can be found in its sample size and procedures of data collection. Although the design is prospective it was not possible to assess all patients pre-operatively. Information collected from patients and relatives with regard to pre-fracture functioning might be biased. The use of the IQCODE showed limitations in this study. Many elderly people who live by themselves do not have relatives that are close enough to be able to answer questions that are experienced as threatening for the person reviewed. The IQCODE, however, has several advantages when compared to other instruments for mental testing. There is no influence of education, pre-morbid intelligence and cultural differences are discounted in the assessment and there is no influence of admission stress, post-operative situation or possible symptoms of delirium. Data from medical and nursing records on precipitating factors, such as malnutrition and dehydration, can be suspected for underreporting. The sample size in this study did not allow regression analysis in order to answer the question studied. Therefore it was not possible to account for associations between factors.

Conclusion

The results of this study permit to come to the following conclusions regarding elderly patients with a hip fracture: identification of risk factors is useful for early recognition of delirium by nurses. Predisposing risk factors confirmed are pre-fracture ADL dependency, psychiatric pre-existing comorbidity and a high number of comorbid problems. Patients with a prolonged waiting time before surgery should be suspected at risk for post-operative delirium. Nursing assessment of patients' pre-fracture functional and cognitive capacities in an early stage after hospital admission is recommended. Delirium, however, can develop in "healthy elderly" with a hip fracture. Nurses' observations of post-operative symptoms are needed for the early recognition of delirium.

Acknowledgements

This study was made possible by a grant of the National Center for Nursing and Care (LCVV). We would like to thank all the patients and their families and the nurses who were involved in this study.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington D.C.: American Psychiatric Association; 1987.
2. Lipowski ZJ. Delirium: acute confusional states: Oxford University Press; 1990.
3. Williams MA, Campbell EB, Raynor SM, Mlynarczyk SM, Ward SE. Reducing acute confusional states in elderly patients with hipfractures. *Research in Nursing and Health* 1985;8:329-337.
4. Berggren D, Gustafson Y, Eriksson B, Bucht G, Hansson L-I, Reiz S, et al. Postoperative confusion after anesthesia in elderly patients with femoral neck fractures. *Anesthesia and Analgesia* 1987;66:497-504.
5. Gustafson Y, Berggren D, Brännström B, Bucht G, Norberg A, Hansson LI, et al. Acute confusional states in elderly patients treated for femoral neck fracture. *JAGS* 1988;36:525-530.
6. Brännström B, Gustafson Y, Norberg A, Winblad B. ADL performance and dependency on nursing care in patients with hip fractures and acute confusion in a task allocation care system. *Scandinavian Journal of Caring Sciences* 1991;5(1):3-11.
7. Gustafson Y, Brännström B, Berggren D, Ragnarsson JI, Sigaard J, Bucht G, et al. A geriatric-anesthesiologic program to reduce acute confusion states in elderly patients treated for femoral neck fractures. *JAGS* 1991;39:655-662.
8. Mullen JO, Mullen NL. Hip fracture mortality. *Clinical Orthopaedics and Related Research* 1992;280:214-222.
9. Bowman AM. Sleep satisfaction, perceived pain and acute confusion in elderly clients undergoing orthopaedic procedures. *Journal of Advanced Nursing* 1997;26:550-564.
10. Edlund A, Lundström M, Lundström G, Hedqvist B, Gustafson Y. Clinical profile of delirium in patients treated for femoral neck fractures. *Dementia and Geriatric Cognitive Disorders* 1999(10):325-329.
11. Milisen K. An intervention study for delirium in elderly hip fracture patients [PhD]. Leuven: Katholic University Leuven; 1999.
12. Marcantonio ER, Flacker JM, Michaels M, Resnick NM. Delirium is independently associated with poor functional recovery after hip fracture. *JAGS* 2000;48:618-624.
13. Brauer C, Morrison RS, Silberzweig SB, Siu AL. The cause of delirium in patients with hip fracture. *Archives of Internal Medicine* 2000;160:1856-1860.
14. Dolan MM, Hawkes WG, Itkin Zimmerman S, Morrison RS, Gruber-Baldini AL, Hebel JR, et al. Delirium on hospital admission in aged hip fracture patients: prediction of mortality and 2-year functional outcomes. *Journal of Gerontology: Medical Sciences* 2000;55A(9):M527-M534.
15. Foreman MD. The cognitive and behavioral nature of acute confusional states. *Scholarly Inquiry for Nursing Practice: an International Journal* 1990;5(1):3-20.
16. George J, Bleasdale S, Singleton SJ. Causes and prognosis of delirium in elderly patients admitted to a district general hospital. *Age and Ageing* 1997;26:423-427.

17. Rudberg MA, Pompei P, Foreman MD, Ross RE, Cassel CK. The natural history of delirium in older hospitalized patients: a syndrome of heterogeneity. *Age and Ageing* 1997;26:169-174.
18. Perez ED. Hip fracture: physicians take more active role in patient care. *Geriatrics* 1994;49(4):31-37.
19. Feldt KS, Griffin PL. Delirium in hip-fractured elders. *Clinical Gerontologist* 1999;20(2):75-78.
20. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of Internal Medicine* 1993;119(6):474-480.
21. Inouye SK, Charpentier PA. Precipitating factors for delirium in hospitalized elderly persons; a predictive model and interrelationship with baseline vulnerability. *JAMA* 1996;275(11):852-857.
22. Bitondo Dyer C, Ashton CM, Teasdale TA. Postoperative delirium. *Archives of Internal Medicine* 1995;155(13):461-465.
23. Marcantonio ER, Goldman L, Mangione CM, Ludwig LE, Muraca B, Haslauer CM, et al. A clinical prediction rule for delirium after elective noncardiac surgery. *JAMA* 1994;271(2):134-139.
24. Marcantonio ER, Goldman L, Orav EJ, Cook EF, Lee TH. The association of intraoperative factors with the development of postoperative delirium. *The American Journal of Medicine* 1998;105:380-384.
25. Hofsté WJ, Linssen CAM, Boezeman EHJF, Hengeveld JS, Leusink JA, Boer Ad. Delirium and cognitive disorders after cardiac operations: relationship to pre- and intraoperative quantitative electroencephalogram. *International Journal of Clinical Monitoring and Computing* 1997(14):29-36.
26. Milisen K, Abraham IL, Broos PLO. Postoperative variation in neurocognitive and functional status in elderly hip fracture patients. *Journal of Advanced Nursing* 1998;27:59-67.
27. Schuurmans MJ. Vroegtijdige signalering van het delier door verpleegkundigen. In: *Onderzoek in de verpleging en verzorging in relatie tot de praktijk*; Maastricht: De Tijdstroom/LCVV; 1996. p. 55-60.
28. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th edition ed. Washington D.C.: American Psychiatric Association; 1994.
29. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Medical Journal* 1965;14:61-61.
30. Haan Rd, Limburg M, Schuling J, Broeshart J, Jonkers L, Zuylen Pv. Klinimetrische evaluatie van de Barthel-Index, een maat voor beperkingen in het dagelijks functioneren. *Nederlands Tijdschrift voor Geneeskunde* 1993;137(18):917-921.
31. Jorm AF, Jacomb PA. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine* 1989;19:1015-1022.

32. Jorm AF, Christensen H, Henderson AS, Jacomb PA, Korten AE, MacKinnon A. Informant ratings of cognitive decline of elderly people: relationship to longitudinal change on cognitive tests. *Age and Ageing* 1996;25:125-129.
33. Jorm AF, Scott R, Cullen JS, MacKinnon AJ. Performance of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) as a screening test for dementia. *Psychological Medicine* 1991;21:785-790.

chapter 8

Low risk for delirium at a geriatric medicine department

Marieke J. Schuurmans

Nienke C. de Goeijen

Yke Y. Kooistra

Jaap J. van der Bijl

Lillie M. Shortridge-Baggett

Sijmen A. Duursma

An abstract of this article was published in Dutch: Nederlands Tijdschrift voor Geneeskunde 2000; 144(6): 302

Submitted (this version)

Abstract

This prospective study was designed to investigate risk factors for delirium in patients admitted to a geriatric department, that can be recognized by nurses. Of 82 elderly patients admitted to a geriatric department data were collected on predisposing and precipitating factors of delirium during the first week of admission. Predisposing factors included age, gender, sensory impairments, functional impairment, pre-existing cognitive impairment and comorbidity. Precipitating factors studied were use of physical restraints, malnutrition, addition of three or more medications, introduction of bladder catheter and any iatrogenic event. During the first week of admission 3 patients developed delirium as diagnosed by a geriatrician using the DSM-IV criteria. This number was too low to meaningfully compare delirious patients and non-delirious patients. The non-delirious patients, however, scored high on the predisposing factors. Their mean age was 82.7 years, more than 50% needed much assistance or were fully dependent in their ADLs, 57% had sensory impairments and 69% had pre-existing cognitive impairment. Nearly all non-delirious patients (90%) had three or more diagnosis. One or more precipitating factors occurred during the first week of admission in 68% of the non-delirious patients. Conclusion of this study is that in a high risk group of geriatric patients, the occurrence of delirium was exceptionally low. This finding may suggest a preventive effect of the structure and approach in a department of geriatric medicine.

Introduction

Delirium is a transient organic mental syndrome, common in older patients after hospital-admission (1). Incidence rates range from 5 to 50% (2-8). The major characteristics of the syndrome are: reduced level of consciousness, reduced ability to maintain attention, perceptual disturbances and memory impairment. The syndrome develops in a short period of time and the symptoms fluctuate (9).

Some patients are more vulnerable to develop a delirium than others. Inouye et al. (2, 10) divided risk factors for delirium into predisposing and precipitating factors. Predisposing factors are factors present at the time of hospital admission and reflect baseline vulnerability. Precipitating factors are noxious insults or hospital related factors that contribute to the development of delirium. Inouye et al. (10) suggest that patients with high baseline vulnerability to delirium may develop delirium with any precipitating factor, even one that is mild. Patients with low vulnerability would be resistant to the development of delirium, even with noxious insults. Several studies have tried to identify predisposing and precipitating factors in elderly patients (2, 7, 10-14). In all, but one study (12), pre-existing cognitive impairment was identified as a predisposing factor for delirium. Other predisposing factors

include severe illness (2, 13, 14), dehydration (2, 13, 14), high age (7, 12, 14), male gender (7, 12), alcohol abuse (11, 12), burden of comorbidity (11, 14), fracture (7), depression (11) and vision impairment (2). Elie et al. (15), who reviewed 27 articles on risk factors for delirium, published between 1966 and 1995, found that dementia, advanced age and medical illness were consistently identified as risk factors.

Inouye et al. (10) developed a model with five precipitating factors: use of physical restraints, malnutrition, more than three medications newly prescribed, use of bladder catheter and any iatrogenic event. The model strongly predicted the occurrence of delirium in the developmental cohort, but less strongly in the validation cohort. Precipitating factors found by Schor et al. (7) were infection and use of neuroleptics or narcotics.

In textbooks environmental factors or hospital-related factors like sleep deprivation, unfamiliar environment, frequent reallocation and lack of windows are also described as precipitating factors, but these factors have not been validated by research (16-18). No significant correlations have been found between delirium and psychosocial factors including stress and lack of social support (16, 17).

This prospective study examined risk factors and the occurrence of delirium in a group of elderly patients. The study was designed in a way that findings could be incorporated into nursing admission history and bedside screening.

Methods

Design and sample

This study had a prospective design. The hospital ethics committee approved the study. All patients admitted to the 24-bed department of geriatric medicine in the University Medical Center in Utrecht, the Netherlands were asked to participate in this study. Patients who were diagnosed delirious on admission were excluded.

Patients are admitted to the department of geriatric medicine if they are of advanced age and have internal or neurological problems, sometimes combined with psychiatric problems, or multiple problems. All patients can be considered as 'acute patients'. One third of the patients were emergencies, who needed admittance immediately or at the same day. The other patients were admitted within some days after the request of and discussion with the general practitioner or via the outpatient clinic for geriatric medicine. No patients were admitted for rehabilitation or longterm care. In the Netherlands rehabilitation and longterm care for the elderly occur in the typical Dutch nursing homes. About 40% of the patients in nursing homes are discharged after some weeks or months to home or a residential house. The approach in the department of geriatric medicine is patient centered, problem oriented and multidisciplinary. The mean length of stay is 23 days and the mortality rate is 8% (19).

Geriatricians staff the department and all nurses are registered nurses specially trained in care in geriatric medicine. The department has 24 beds in two units, each unit has two single rooms, three rooms for two patients and one room for four patients. Each unit is staffed by four nurses and a teamleader on the day shift, two nurses on the evening shift and one nurse on the night shift. In the daytime most patients are encouraged and supported to get out of bed and participate in activities. An occupational therapist organizes activities on individual and group level. Daily interventions by a physiotherapist are part of regular care.

Procedure

Data were collected from September 1996 to February 1997. The responsible physician asked for consent from patients who met the inclusion criteria. Surrogate consent was requested in case of incompetence. Within 2 hours after admission the physician and the primary nurse had an admission interview with the patient and relatives and a physical examination followed the interview. Nurses observed patients every shift for symptoms of delirium and registered their observations on the Delirium Observation Screening Scale (20), a 25-item observational rating scale. A geriatrician saw the patients daily and diagnosed delirium based on the Diagnostic Statistical Manual-IV criteria (9). To check the reliability of the diagnosis 10% of the patients were also seen by a psychiatrist. Agreement between the diagnoses of the geriatrician and the psychiatrist for the patients who were compared was 100%.

Data for the study were collected from the admission interview, charts, and additional information was obtained from the family, physician and nurses.

Measures

Immediately after enrollment in the study, information on predisposing factors was collected, including age, gender, functional impairment, speech impairment, hearing or vision impairments, pre-existing cognitive impairment, severity of illness and use of medication. Functional impairment was measured with the Barthel Index (21, 22), concerning ten activities of daily living (ADL). Impairment of speech, hearing and vision were rated as none, mild or severe. Pre-existing cognitive impairment was measured with the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (23-25), a 26 item family member rated instrument that screens for cognitive decline, rating scores between 1 and 5. The influence of education, premorbid intelligence and cultural differences are discounted in the assessment; admission stress or early symptoms of delirium do not affect the ratings. Severity of illness was measured by counting the total number of medical diagnoses according to the ICD-9.

After a diagnosis of delirium by an independent geriatrician, or 7 days after admission, data on precipitating factors as described by Inouye et al. (10) were collected. The information was obtained from the patient's record, with day to day notes from nurses and physicians. The discharge letter was used as a second source of information.

Statistical analysis

Data were analyzed using SPSS. Presence of risk factors was calculated in percentages.

Results

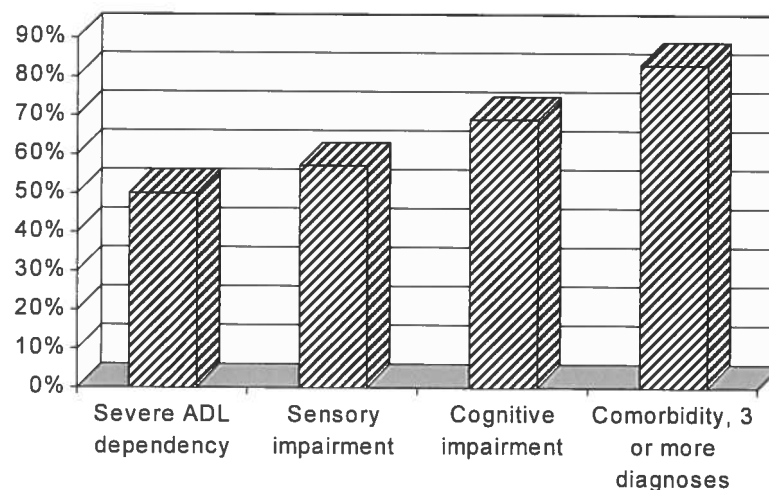
During the period of study 94 patients were admitted to the department of geriatric medicine. Ten were excluded from the study for the following reasons: delirium on admission (8 patients), transfer to intensive care unit (1 patient), hospital stay of less than 7 days (1 patient) and non-responding identification numbers (2 patients). In this sample the main diagnoses were cardiovascular (40% of all patients), gastric-intestinal (38.8%), psychiatric (38.8%), mobility problems (36.4%), urogenital (43.2%) and pulmonary (27%).

Of the 82 patients studied, 3 patients (3.7 %) developed delirium within the first week after admission. One patient was diagnosed as delirious on the eighth day of admission and one patient was diagnosed as delirious one month after admission. The group of delirious patients was too small for statistical analyses and therefore is not described further here.

Predisposing factors

The geriatric non-delirious patients (n =78) had a mean age of 82.7 (range between 70 and 96, sd 6.1) and a gender distribution of 28% males and 72% females. Figure 1 shows the presence of predisposing factors in the non-delirious patients. Among these patients, a functional impairment score on the Bartel Index ranged from 0 (fully dependent) to 20 (fully independent). More than 50% of the patients had a score of 11 or less, meaning they needed much assistance or were fully dependent. Most patients needed help in bathing (80%), dressing (75%) and using the bathroom (65%). Only half of the patients was able to feed themselves independently. Fewer than 35% were independent in mobility, and fewer than 20% were independent in all their activities of daily living. Over half of the patients (57%) had one or more sensory disorder(s), mainly disorders of hearing or vision. Fifteen percent of the patients had both hearing and vision disorders.

Figure 1 Predisposing factors for delirium in 78 non-delirious geriatric patients



Scores on the IQCODE, measure of pre-existing cognitive decline were available for 69% of the patients. In this group 41% showed beginning to mild cognitive decline (score 3.5 – 4.49) and 28% showed severe cognitive decline (score 4.5-5).

The number of ICD-9 diagnoses (co-morbidities) was used as a measure of severity of illness. Eighty-three percent of the group had three or more diagnoses (range 1 to 7, mean 3.7 with a standard deviation of 1.4).

Precipitating factors

One or more precipitating factors occurred in 68% of the patients. The most frequent precipitating factor was the addition of three or more medications (in 35% of all patients). In 24% of the patients a urinary catheter was introduced. Malnutrition was seen in 18% and infections occurred in 13%. Physical restraints were used in 6% of the patients and iatrogenic events occurred in 4.7% of the patients.

Discussion

Only three patients (3.7%) developed delirium during this 4.5 month prospective study of patients admitted to a department of geriatric medicine. This incidence is exceptionally low. A high baseline vulnerability to delirium, as indicated by high age, functional impairment, prior cognitive impairment and severe illness (comorbidity), was present in many patients in this

group. It is widely assumed that patients with high baseline vulnerability to delirium may develop delirium with any precipitating factor, even one of mild degree. Precipitating factors occurred for 68% of the patients during the first week of their hospital stay. Yet none of these patients did develop delirium during this period. A preventive effect might have been produced by the structure and approach of the specialized department in geriatric medicine (assignment of patients, problem-oriented approach and a highly educated multidisciplinary team). A study by Inouye et al.(26) supports this view. In a large controlled clinical trial they studied the effect of a delirium risk modification program implemented by a highly trained health care team. The program targeted six risk factors for intervention: cognitive impairment, sleep deprivation, immobility, visual impairment, hearing impairment and dehydration. The intervention strategy resulted in a significant reduction in the number and duration of episodes of delirium in hospitalized older patients. This was the first controlled clinical trial to provide evidence of the effectiveness of hospital-based interventions to reduce the incidence of delirium among at-risk older hospitalized patients. Earlier studies of multidisciplinary team interventions for delirium did not provide such strong support (27-29). Management of risk factors for delirium as prescribed in the protocols of Inouye et al. (26) is comparable to regular care in the department of geriatric medicine in Utrecht. The preventive effects of a specialized department of geriatric medicine need to be further studied. Research effects and costs of such a facility are important.

The study described here had a small sample size and a large part of the data were collected during regular admission assessments by nurses and physicians, which might limit reliability and validity. The use of the number of diagnoses as a measure of severity of illness is also perhaps questionable, though comorbidity itself is a risk factor (11, 14).

Clearly, to prevent delirium, knowledge of risk factors and early recognition of symptoms are important. Further evaluation of predictive models is recommended.

Acknowledgement

This study was supported by a grant of the National Center for Nursing and Care (LCVV). The authors would like to thank the participating patients and their families, the nurses and physicians of the department of Geriatric Medicine, and F. Balk, consultant psychiatrist, of the University Medical Center in Utrecht.

References

1. Lipowski ZJ. Delirium: acute confusional states: Oxford University Press; 1990.
2. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of Internal Medicine* 1993;119(6):474-480.
3. Johnson JC, Gottlieb GL, Sullivan E, Wanich C, Kinosisian B, Forciea MA, et al. Using the DSM-III criteria to diagnose delirium in elderly general medical patients. *Journal of Gerontology: Medical Sciences* 1990;45(3):M113-119.
4. Rockwood K. The occurrence and duration of symptoms in elderly patients with delirium. *Journal of Gerontology: Medical Sciences* 1993;48(4):M162-166.
5. Francis J, Martin D, Kapoor WN. A prospective study of delirium in hospitalized elderly. *JAMA* 1990;263(8):1097-1101.
6. Levkoff SE, Evans DA, Liptzin B, Cleary PD, Lipsitz LA, Wetle TT, et al. Delirium: the occurrence and persistence of symptoms among elderly hospitalized patients. *Archives of Internal Medicine* 1992;152(2):334-340.
7. Schor JD, Levkoff SE, Lipsitz LA, Reilly CH, Cleary PD, Rowe JW, et al. Risk factors for delirium in hospitalized elderly. *JAMA* 1992;267(6):827-831.
8. Williams MA, Campbell EB, Raynor SM, Mlynarczyk SM, Ward SE. Reducing acute confusional states in elderly patients with hipfractures. *Research in Nursing and Health* 1985;8:329-337.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
10. Inouye SK, Charpentier PA. Precipitating factors for delirium in hospitalized elderly persons; a predictive model and interrelationship with baseline vulnerability. *JAMA* 1996;275(11):852-857.
11. Pompei P, Foreman M, Rudberg MA, Inouye SK, Braund V, Cassel CK. Delirium in hospitalized older persons: outcomes and predictors. *Journal of the American Geriatric Society* 1994;42:809-815.
12. Williams-Russo P, Urquhart BL, Sharrock NE, Charlson ME. Post-operative delirium: predictors and prognosis in elderly orthopedic patients. *Journal of the American Geriatric Society* 1992;40:759-767.
13. O'Keeffe ST. Predicting delirium in elderly patients: development and validation of a risk-stratification model. *Age and Ageing* 1996;25:317-321.
14. Eden BM, Foreman MD, Sisk R. Delirium: comparison of four predictive models in hospitalized critically ill elderly patients. *Applied Nursing Research* 1998;11(1):27-35.
15. Elie M, Cole MG, Primeau FJ, Bellavance F. Delirium risk factors in elderly hospitalized patients. *Journal General Internal Medicine* 1998;13:204-212.
16. Inouye SK. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *The American Journal of Medicine* 1994;97:278-288.

17. Foreman MD. Acute confusional states in hospitalized elderly: a research dilemma. *Nursing Research* 1986;35(1):34-38.
18. Rajacich D, Faux S. The relationship between relocation and alterations in mental status among elderly hospitalized patients. *The Canadian Journal of Nursing Research* 1988;20(4):31-42.
19. Jaarverslag afdeling geriatrie. Utrecht: AZU; 1996.
20. Schuurmans MJ. Vroegtijdige signalering van het delier door verpleegkundigen. In: *Onderzoek in de verpleging en verzorging in relatie tot de praktijk*; Maastricht: De Tijdstroom/LCVV; 1996. p. 55-60.
21. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Medical Journal* 1965;14:61-61.
22. Haan Rd, Limburg M, Schuling J, Broeshart J, Jonkers L, Zuylen Pv. Klinimetrische evaluatie van de Barthel-Index, een maat voor beperkingen in het dagelijks functioneren. *Nederlands Tijdschrift voor Geneeskunde* 1993;137(18):917-921.
23. Jorm AF, Jacomb PA. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. *Psychological Medicine* 1989;19:1015-1022.
24. Jorm AF, Scott R, Cullen JS, MacKinnon AJ. Performance of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) as a screening test for dementia. *Psychological Medicine* 1991;21:785-790.
25. Jorm AF, Christensen H, Henderson AS, Jacomb PA, Korten AE, MacKinnon A. Informant ratings of cognitive decline of elderly people: relationship to longitudinal change on cognitive tests. *Age and Ageing* 1996;25:125-129.
26. Inouye SK, Bogardus ST, Charpentier PA, Leo-Summers L, Acampora D, Holford TR, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. *The New England Journal of Medicine* 1999;340(9):669-676.
27. Britton A, Russell R. Multidisciplinary team interventions for delirium in patients with chronic cognitive impairment. Oxford: The Cochrane Library; 1999.
28. Epling J, Taylor H. Preventing delirium in hospitalized older patients. *The Journal of Family Practice* 1999;48(6):417-418.
29. Young LJ, George J. Do guidelines influence the process and outcomes of care in delirium? In: *British Geriatrics Society Spring Meeting* 1999. p. 23.

chapter 9

General Discussion and Conclusions

Introduction

Delirium is one of the most frequent forms of psychopathology in elderly hospitalized patients. Despite the high incidence, delirium is often not recognized or misdiagnosed for dementia or depression. Not diagnosing delirium can have far-reaching consequences for the prognosis of elderly patients.

The fluctuating nature of the syndrome and the onset in evening and night should facilitate early recognition of symptoms by nurses. The studies presented in this thesis were designed to study the possibility to come to the diagnosis of delirium by a limited number of standard nurses' observations. Six studies were conducted of which four studies in patients: geriatric patients (2 samples), hip fracture patients and hospital patients referred for psychiatric consultation. Two methodological studies were done with experts and nurses. In order to study nurses' observations of delirium the Delirium Observation Screening (DOS) Scale was developed and tested for reliability and validity. An additional question was studied with regard to early recognition of delirium, focussing on identification of risk factors.

In this chapter the general aspects of the studies are considered in a broader perspective. The methodological issues are discussed, the major findings are further outlined and implications for clinical practice and recommendations for future research are made.

Methodological issues

Design and samples

In order to study early recognition of delirium a prospective design in two high-risk groups of patients was chosen: patients admitted to a department of geriatric medicine and elderly patients admitted to a surgical department. All patients of the surgical department had a hip fracture. The expectation, based on other studies, was an incidence of delirium of 20%- 25% in geriatric patients (1) and of 25% in hip fracture patients (2, 3). The study of 82 consecutive admissions to the department of geriatric medicine showed a percentage of 3.7% (Chapter 3, 8). This percentage is exceptionally low in a population rated high on risk factors for delirium. Several explanations can be given for this finding. Theories about the influence of risk factors can be questioned, however, factors as high age, pre-existing cognitive impairment, ADL impairment and a high number of comorbid conditions have been identified as risk factors in several studies in different groups of patients (4-8). A more likely explanation for the low number of delirious patients might be the preventive influence of the structure and approach of the specialized department of geriatric medicine. Standard care in this department is comparable with the preventive interventions for delirium in a large clinical trial by Inouye et al. (9). The intensive monitoring of symptoms as part of the research protocol may have strengthened this effect. This explanation is tentative. Further research

is needed with regard to the preventive effect of the specialized care in the department of geriatric medicine and the possible preventive effect of intensive monitoring.

The study of 92 hip fracture patients, admitted to the department of surgery, showed 19.6 % delirium (Chapter 3, 7, 8). This percentage is lower than in some other studies. An explanation can be the use of the DSM-IV criteria, while studies of delirium in hip fracture patients with incidences over 40% (10-14) were based on the DSM-III criteria. The DSM-III criteria are known to identify more patients as delirious than the DSM-III-R criteria (15). The latter are near to the DSM-IV criteria.

The total number of delirious patients, on which the results in our studies are based, is modest. With this modest number of delirious patients, however, significant results are found. Significant findings are less expected in small numbers of patients or in samples with a low incidence of the problem studied. The expectation is that the results presented in this thesis can be replicated in larger studies or in samples with a higher percentage of delirious patients.

Procedures

As a result of the nature of the study nurses were asked to collect the data during regular care. They observed the patients and rated delirium symptoms. In the study of the hip fracture patients this procedure lead to a number of missing data (Chapter 6). Temporary employees, high workload, simply forgotten and other reasons were given by nurses for not rating a patient's behavior. Nurses' reports as assessment tool showed similar problems in a study of Neelon et al. (16). Analysis of the missing values showed that most missing of data occurred in evening shifts and night shifts. In these shifts nurses have less frequent contacts with patients and during the night shifts patients are often asleep. By the use of mean values, variances of means and percentages of ratings that indicate altered behavior, the influence of missing values was limited.

Measures

Diagnosis of delirium

In this study the Diagnostic Statistical Manual-IV criteria of the American Psychiatric Association for delirium (17) were used as standard for diagnosis. These criteria are a research based revision (15, 18, 19) of earlier versions of the DSM (20, 21). These earlier versions were mainly based on clinical discussions. The changes made in the last version are thought to make the application of the criteria easier, especially in elderly patients with pre-existing cognitive deficits (18, 19, 22).

In the studies described in this thesis a geriatrician and/or a psychiatrist applied the DSM-IV criteria. A diagnosis by a single physician used as 'gold standard' can be discussed with

regard to reliability, reproducibility and interpretation of the standard. On the other hand there is no alternative standard in delirium research. Possible improvement of this standard could be the use of instruments developed to diagnose delirium or an independent diagnosis made by two clinicians. The use of instruments developed to diagnose delirium, such as the Confusion Assessment Method (23), suggests a higher reproducibility, however, none of the instruments described (Chapter 2) shows a 100% sensitivity and a 100% specificity. The sensitivity and specificity rates of these instruments are also calculated against a physician's diagnosis. With regard to the diagnosis made by two trained clinicians (geriatricians or psychiatrists), in the study in geriatric patients (Chapter 3, 8), all patients were diagnosed by a geriatrician and 10% of all patients were seen by a psychiatrist. Full agreement was found on the acceptance or rejection of the diagnosis delirium. In the study in hip fracture patients (Chapter 3, 6, 7) all patients seen by a geriatrician were discussed in a group of geriatricians. The diagnoses proposed were never questioned or rejected. Limitation in this last study was that a geriatrician did not see all patients. The research nurse selected which patients had to be seen based on alterations of behavior. This can also raise the question whether the ratings indicating altered behavior before the diagnosis was made, indicate early recognition of symptoms of delirium or, reflect a delay in the diagnosis of delirium by the geriatrician. Based on the procedures and the findings of other studies and experience in clinical practice, this last possibility is not very plausible. Visits of the geriatrician were in all cases on the same day as the report of alteration of behavior by the research nurse. The development of delirium ranged from day 2 to 6, mean day 3, which is comparable with a study of Milisen et al. (24). As a result of the active case finding by the research nurse the diagnosis of delirium can expected to be made in an early stage compared to regular practice.

The DOS Scale

In the last fifteen years several instruments have been developed to diagnose delirium or to screen patients for symptoms of delirium (Chapter 2). None of the instruments studied was suitable to study early recognition of delirium by nurses' observations during regular care. Therefore the Delirium Observation Screening (DOS) Scale was developed (Chapter 3). In the validation process experts participated (content validity) as well as three groups of patients; geriatric patients, hip fracture patients and patients referred for psychiatric consultation. Reliability was tested in the same groups of patients, in an additional sample of geriatric patients and in a sample of nurses and nursing experts rating a video-taped case. The usability in clinical practice was examined in a sample of nurses. The 5-point Likert scale (never, sometimes, fluctuating, usually, always) was used in the studies with the geriatric patients and the hospitalized patients referred for psychiatric consultation. The results of these studies showed difficulties with the rating of 'fluctuating'. Therefore in the study with the

hip fracture patients and the video taped patient a 4-point Likert scale was used (excluding 'fluctuating').

The strength (of the development) of the DOS Scale can be found in the extent to which this scale is tested for its psychometric properties (Chapter 3, 4, 5). Several aspects of reliability and validity were tested in more than one sample. Results were fairly consistent. Questions can be raised to the choice to analyze the DOS Scale on the sum score and on the level of items, giving 25 outcomes for each measure. This gives tables with many figures and results with a large range of outcomes. This choice was made since the DOS Scale was developed to study nurses' observations with regard to early recognition of delirium. It is known that nurses are not able to observe the symptoms of delirium as such (25, 26). A concept as 'disturbance of consciousness' is known to be vague (27) and hard to rate in a reliable manner. Therefore diagnostic and related symptoms of delirium were operationalized in items that describe observable behavior. Analysis on the level of items gave insight in the value of each item and allowed a limitation in the number of items. Overall the items of the DOS Scale showed satisfactory to good reliability and validity and the scale was evaluated as positive for use in clinical care. A difficulty was found in establishing interrater reliability. Testing of interrater reliability in an instrument developed for observation during regular care showed some major drawbacks. The methods (data collection during regular care), samples (small, large proportion of 'normal behavior') and analyses (percentage agreement and Spearman rank correlation coefficients) used in these studies can be criticized (Chapter 4). Training of raters can enhance the interrater reliability (28, 29). Nevertheless, instruments with good interrater reliability, when tested by researchers do not guarantee good interrater reliability when used by nurses in clinical practice (30).

The results in the hip fracture patients allowed to reduce the number of items of the DOS Scale and to develop an algorithm. This final scale should be seen as a preliminary scale since it is based on data of 18 delirious hip fracture patients. The results found on a sample of this size are, however, promising. Further study in different samples should confirm the strength of the shortened scale in early recognition.

Predisposing and precipitating factors

The predisposing and precipitating factors studied were all known from earlier studies (Chapter 1, 7, 8). This limits the results to confirmation of earlier findings. The choice of factors studied was guided by the possibility of assessment of these factors by nurses. Major limitation of this part of the data collection is, however, the use of data from medical and nursing reports. This may have lead to misclassification of comorbid factors such as dementia and to underreport of factors such as malnutrition and dehydration.

Major findings

Major findings with regard to the questions studied, can be described as follows:

- *Is it possible to come to the diagnosis of delirium with a limited number of standard observations by nurses?*

Yes, to come to the diagnosis of delirium with a limited number of standard observations by nurses is possible. Ratings of observations on the DOS Scale in elderly hip fracture patients, showed that preceding a diagnosis of delirium nurses observed behavioral alterations in thirteen DOS Scale items related to diagnostic symptoms of delirium. These alterations fluctuate and observations in all three shifts (day, evening and night) gave significant results to early recognition. Based on these findings an algorithm of the differentiating items of the DOS Scale showed with 48 hours of rating a sensitivity of 100% and a specificity of 68%.

- *Is the Delirium Observation Screening Scale a valid and reliable instrument to detect delirium by nurses' observations during regular care?*

Yes, the Delirium Observation Screening Scale can be considered a valid and reliable instrument to rate nurses' observations during regular care. The DOS Scale in three samples of patients (geriatric patients, hip fracture patients and patients referred for psychiatric consultation) showed very good internal consistency (α : 0.92 - 0.97). Good criterion related validity with the DSM-IV diagnosis was seen in three samples by significant differences in mean ratings of delirious and non-delirious patients. Interrater reliability showed problems in a sample of nurses trained in care in geriatric medicine. In a sample of staff nurses rating patients referred for psychiatric consultation and a sample of staff nurses rating a videotaped delirious patients satisfactory results were accomplished. Content validity was established by a multidisciplinary group of experts. Criterion related validity was tested by correlation with the Mini Mental State Examination. Construct validity was tested in relation to the Informant Questionnaire of Cognitive Decline in the Elderly, a premorbid psychiatric diagnosis and the Barthel Index. The results, as shown in Table 1, confirm good criterion related validity and modest to good construct validity of the DOS Scale in these studies.

Table 1 Criterion- and construct validity of the DOS Scale

	Geriatric patients		Hip fracture patients	
	DOS Scale items	DOS Scale sum score	DOS Scale items	DOS Scale sum score
MMSE	- 0.52 - 0.74 ¹	- 0.66 ***	- 0.44 - 0.74 ¹	- 0.79 ***
IQCODE	0.10 - 0.41	0.33*	0.42 - - 0.72	0.74 ***
Psychiatric diagnosis	0.16 - 0.42	0.42 ***	0.25 - - 0.53	0.43 ***
Barthel Index	0.13 - 0.38	- 0.26 *	0.29 - - 0.57	- 0.55 ***
¹ DOS Scale items related to thinking, memory and orientation (4-11) $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ Calculated Spearman rank correlation tests				

The DOS Scale showed significant differentiation between symptoms of delirium and symptoms of other psychiatric problems.

- *Is identification of risk factors useful for early recognition of delirium?*
1. Yes, the study in elderly hip fracture patients confirmed that patients with premorbid ADL impairment ($p = 0.02$), psychiatric comorbidity ($p = 0.006$) or a high number of comorbid conditions ($p = 0.02$) are at risk. Delay of surgery of over 48 hours increased the risk for delirium, the difference was, however, not significant
 2. No, the study in geriatric patients showed a low percentage of delirium in high-risk patients. A preventive effect of structured care and the specialized multidisciplinary team in the department of geriatric medicine is a possible explanation.

Implications for clinical practice

Delirium is a serious, often unrecognized, condition in elderly hospitalized patients. As Inouye et al. (31, p 568) stated 'An effective case finding method would initiate an improvement in the quality of care for the elderly'. Findings of the studies described in this thesis show that by standard observation, staff nurses can observe delirium in an early stage. This standard observation can be done during regular care and rating takes a few minutes maximum. Due to the methodological issues discussed, a definitive answer to the usefulness of the DOS Scale algorithm in clinical practice can not be given. Although further testing of the scale and of the algorithm is recommended, use in clinical practice will, nevertheless, benefit the care for elderly patients. Nurses lack knowledge and experience in detecting delirium symptoms or as Rahm Hallberg described 'are not attuned to detecting delirium' (32, p 422). Use of the DOS Scale algorithm will influence and guide nurses' observations of elderly patients. Implementation of the scale in hospital care of elderly patients is a simple intervention that will lead to early recognition of delirium, resulting in

nurses' awareness of the importance of their observations for the prognosis of the elderly hospitalized patients. Based on this study, assessment shortly after admission of a patient's cognitive, somatic and functional status before admission is recommended. Elderly patients with ADL impairments, with psychiatric comorbidity and elderly patients with a high number of comorbid problems should be monitored closely for delirium when admitted to a hospital. Therefore having complete assessment of patients prehospital situation is necessary. Waiting to the moment of discharge, to assess a patient's situation before admission, is not acceptable.

Implications for further research

Further research should be done into recognition and prevention of delirium. Nursing research can add significantly in this area of research. As Gottlieb (33, p 158) stated 'the role of nurse-investigator-initiated research must be supported as a keystone, given that some of the most exciting interventions have been shaped by nursing observations and practice'.

Future study of the DOS Scale in elderly hospitalized patients should confirm if the 13 -item algorithm is a sensitive and specific method for early recognition of delirium. The DOS Scale is developed for observations during regular care and tested for use by nurses. Until now not much study is done on delirium in elderly patients at home. As a result of the operationalizations of the observations, the DOS Scale can supposedly be useful for observations by persons who take care for an elderly person at home. Further research in this area is recommended.

Last, but not least, the preventive effect of the care, as given in the structured approach of a specialized department of geriatric medicine, deserves further study. Awareness of frailty of elderly patients and the influence of patient centered pro-active care on the prognosis of the elderly patients, is a challenge for all working with elderly patients.

References

1. O'Keeffe S, Lavan L. The prognostic significance of delirium in older hospital patients. *Journal of the American Geriatric Society* 1997;45:174-178.
2. Edlund A, Lundström M, Lundström G, Hedqvist B, Gustafson Y. Clinical profile of delirium in patients treated for femoral neck fractures. *Dementia and Geriatric Cognitive Disorders* 1999;10:325-329.
3. Marcantonio ER, Flacker JM, Michaels M, Resnick NM. Delirium is independently associated with poor functional recovery after hip fracture. *JAGS* 2000;48:618-624.
4. Foreman MD. Confusion in hospitalized elderly. Incidence, onset and associated factors. *Research in Nursing and Health* 1989;12:21-29.
5. Pompei P, Foreman M, Rudberg MA, Inouye SK, Braund V, Cassel CK. Delirium in hospitalized older persons: outcomes and predictors. *Journal of the American Geriatric Society* 1994;42:809-815.
6. Inouye SK, Viscoli CM, Horwitz RI, Hurst LD, Tinetti ME. A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. *Annals of Internal Medicine* 1993;119(6):474-480.
7. O'Keeffe ST. Predicting delirium in elderly patients: development and validation of a risk-stratification model. *Age and Ageing* 1996;25:317-321.
8. Eden BM, Foreman MD, Sisk R. Delirium: comparison of four predictive models in hospitalized critically ill elderly patients. *Applied Nursing Research* 1998;11(1):27-35.
9. Inouye SK, Bogardus ST, Charpentier PA, Leo-Summers L, Acampora D, Holford TR, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. *The New England Journal of Medicine* 1999;340(9):669-676.
10. Berggren D, Gustafson Y, Eriksson B, Bucht G, Hansson L-I, Reiz S, et al. Postoperative confusion after anesthesia in elderly patients with femoral neck fractures. *Anesthesia and Analgesia* 1987;66:497-504.
11. Gustafson Y, Berggren D, Brännström B, Bucht G, Norberg A, Hansson LI, et al. Acute confusional states in elderly patients treated for femoral neck fracture. *JAGS* 1988;36:525-530.
12. Brännström B, Gustafson Y, Norberg A, Winblad B. ADL performance and dependency on nursing care in patients with hip fractures and acute confusion in a task allocation care system. *Scandinavian Journal of Caring Sciences* 1991;5(1):3-11.
13. Gustafson Y, Brännström B, Berggren D, Ragnarsson JI, Sigaard J, Bucht G, et al. A geriatric-anesthesiologic program to reduce acute confusion states in elderly patients treated for femoral neck fractures. *JAGS* 1991;39:655-662.
14. Bowman AM. Sleep satisfaction, perceived pain and acute confusion in elderly clients undergoing orthopaedic procedures. *Journal of Advanced Nursing* 1997;26:550-564.
15. Liptzin B. What criteria should be used for the diagnosis of delirium? *Dementia and Geriatric Cognitive Disorders* 1999;10(5):364-367.

16. Neelon VJ, Champagne MT, McConnel E, Carlson J, Funk SG. Use of the NEECHAM Confusion Scale to assess acute confusional states of hospitalized older patients. In: Key aspects of elder care; managing falls, incontinence and cognitive impairment; 1992. p. 278-289.
17. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington D.C.: American Psychiatric Association; 1994.
18. Tucker GJ. DSM-IV: proposals for revision of diagnostic criteria for delirium. *International Psychogeriatrics* 1991;3(2):197-208.
19. Tucker GJ. The diagnosis of delirium and DSM-IV. *Dementia and Geriatric Cognitive Disorders* 1999;10(5):359-363.
20. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington: American Psychiatric Association; 1980.
21. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington D.C.: American Psychiatric Association; 1987.
22. Hertogh CMPM, Eikelenboom P. Delirium bij ouderen en de DSM-IV: is er nu helderheid over verwardheid? *Tijdschrift voor Gerontologie en Geriatrie* 1995(26):51-56.
23. Inouye SK, Dyck CHv, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the Confusion Assessment Method. *Annals of Internal Medicine* 1990;113(12):941-948.
24. Milisen K, Abraham IL, Broos PLO. Postoperative variation in neurocognitive and functional status in elderly hip fracture patients. *Journal of Advanced Nursing* 1998;27:59-67.
25. Morency CR. Mental status change in the elderly: recognizing and treating delirium. *Journal of Professional Nursing* 1990;6(6):356-365.
26. Yeaw EMJ, Abbate JH. Identification of confusion among the elderly in an acute care setting. *Clinical Nurse Specialist* 1993;7(4):192-197.
27. Lipowski ZJ. *Delirium: acute confusional states*: Oxford University Press; 1990.
28. Polit DF, Hungler BP. *Nursing Research, principles and methods*. 6th ed. New York: J.B. Lippincott Company; 1999.
29. Zou Y, Cole MG, Primeau FJ, McCusker J, Bellavance F, Laplante J. Detection and diagnosis of delirium in the elderly: psychiatrist diagnosis, Confusion Assessment Method, or consensus diagnosis? *International Psychogeriatrics* 1998;10(3):303-308.
30. Whall AL. The measurement of need-driven dementia-compromised behavior. *Journal of Gerontological Nursing* 1999;25(9):33-37, 50-51.
31. Inouye SK, Schlesinger MJ, Lydon TJ. Delirium: a symptom of how hospital care is failing older persons and a window to improve quality of hospital care. *The American Journal of Medicine* 1999;106(May):565-571.
32. Rahm Hallberg I. Impact of delirium on professionals. *Dementia and Geriatric Cognitive Disorders* 1999;10(5):420-425.
33. Gottlieb GL. The future of delirium research. *Journal of Geriatric Psychiatry and Neurology* 1998;11(3):157-158.

Summary

The studies described in this thesis were initiated to study early recognition of delirium. In elderly hospitalized patients delirium is a frequent form of psychopathology with serious consequences (Chapter 1). High mortality and morbidity, lengthened hospital stay and nursing home placement. Delirium is often overlooked or misdiagnosed for dementia or depression due to lack of knowledge and awareness of nurses and doctors. Early recognition of delirium enables diagnosis and treatment of the underlying etiological factor(s), and may prevent consequences of the disturbed behavior. The diagnostic criteria for delirium as described in the Diagnostic Statistical Manual (DSM) IV of the American Psychiatric Association are: a disturbance of consciousness with reduced ability to focus, sustain or shift attention, and changed cognition or the development of a perceptual disturbance. Delirium develops in a short period of time, symptoms fluctuate and first symptoms occur mostly in the evening or during the night. Therefore observations made by nurses may play a key role in the diagnosis of delirium. The central question of this thesis was:

Is it possible to come to the diagnosis of delirium with a limited number of standard observations by nurses?

First step was to review delirium instruments with regard to characteristics and psychometric qualities (process of testing and results) (Chapter 2). Twelve instruments were found, differences in these instruments were described in goal (diagnosis, screening symptoms severity), type of data on which the rating is based (observation, interview or test of patients), the rater qualities required, number of items and the rating time needed. Most instruments are based on the DSM criteria and measure signs and symptoms as described by these criteria. Much difference was found in the degree of which reliability and validity were tested and the quality of the testing procedures used.

Three instruments were observation based and developed for use by nurses; the Clinical Assessment of Confusion A, the Confusion Rating Scale and the NEECHAM Confusion Scale. None of them is based on the DSM criteria and due to methodological problems none of them was considered suitable to study the central question of this thesis. Therefore a new instrument was developed and tested for its psychometric qualities.

An additional research question was formulated:

Is the Delirium Observation Screening Scale a valid and reliable instrument to detect delirium by nurses' observation during regular care?

The Delirium Observation Screening (DOS) Scale is a 25 item scale based on the DSM-IV criteria (diagnostic and related). The DOS Scale was designed to rate behavior that can be observed by nurses while providing regular care, and involves behavior that is observable by

every nurse with basis knowledge of geriatrics. It can also be used with patients who are already cognitive disturbed. Completion of the instrument requires less than 5 minutes.

The DOS Scale was tested for its psychometric qualities in different studies:

- Content validity was established by a panel of seven experts using the two phase procedure described by Lynn (Chapter 3).
- Internal consistency, predictive validity, criterion related validity and construct validity were tested in two prospective studies of high risk patients; 82 geriatric patients (4 became delirious) and 92 hip fracture patients (18 became delirious) (Chapter 3). In these studies during the first week of admission, nurses rated the DOS Scale at the end of every shift (day, evening, night) based on their observation during this shift (8 hours). A diagnosis of delirium was made by a geriatrician based on the DSM-IV criteria. This diagnosis was used to test the predictive validity of the DOS Scale. The criterion related validity was tested in both studies with the Mini Mental State Examination (MMSE). In the study with the hip fracture patients, DOS Scale ratings were also compared with a research nurse' rated DOS Scale and Confusion Assessment Method (CAM). The research nurse rated the instruments after a 15 minutes visit to the patient. Construct validity was tested by correlation with variables related to the occurrence of delirium: cognitive functioning (the Informant Questionnaire of Cognitive Decline in the Elderly (IQCODE)), a pre-existing psychiatric diagnosis and activities of daily living (ADL, the Barthel Index).
- Internal consistency and predictive validity were also tested in a single rating of 57 patients (33 delirious) referred for psychiatric consultation (Chapter 5). Nurses rated the DOS Scale, without knowledge of the psychiatric diagnosis of the patients, at the end of a day shift. The diagnosis was made by a consultant psychiatrist.
- Interrater reliability was tested in three studies (Chapter 4). In two studies DOS Scale ratings were based on simultaneous observations by two nurses during regular care. Patients were referred for psychiatric consultation (n=6) and geriatric patients (n=36). In the third study the DOS Scale was rated based on observation of a videotape of a delirious patient by 44 nurses and 5 nurse specialists.

Since the DOS Scale was developed to study the possibility to come to the diagnosis of delirium with a limited number of standard observations by nurses, analysis were done on the level of items and on sum score.

The DOS Scale showed high internal consistency (Cronbach's $\alpha > 0.90$) in all studies done. The predictive validity, against a DSM-IV diagnosis of a geriatrician or psychiatrist, was good. Significant differences on sum score level and on the level of most items were found in the studies with the geriatric patients, hip fracture patients and patients referred for psychiatric consultation. Concurrent validity with the MMSE showed moderate correlation in the study

with the geriatric patients ($R_s = 0.66$, $p \leq 0.001$) and good correlation ($R_s = 0.79$, $p \leq 0.001$) in the study with the hip fracture patients. Correlation between the nurses' DOS Scale and the research nurse's DOS Scale was $R_s = 0.54$ ($p \leq 0.001$), correlation with the CAM was comparable.

Correlation with the IQCODE was $R_s = 0.33$ ($p \leq 0.05$) in the study with the geriatric patients and $R_s = 0.74$ ($p \leq 0.001$) in the study with the hip fracture patients. Correlation with a pre-existing psychiatric diagnosis was $R_s = 0.42$ and 0.43 ($p \leq 0.001$). Correlation with the Barthel Index was $R_s = 0.26$ ($p \leq 0.05$, geriatric patients) and $R_s = 0.55$ ($p \leq 0.001$, hip fracture patients). The outcomes on the level of items showed differences in degree of validity, in all studies. Differences in findings with regard to the construct validity might be explained by characteristics of the sample and the number of delirious patients in each sample.

Interrater reliability was analyzed using different measures. The three studies showed high percentages of ratings indicating 'normal behavior'. This might explain the low kappa's found on the level of items in the studies of simultaneous observations. Percentage agreement and Spearman correlation coefficients showed good results in the study with the patients referred for psychiatric consultation (mean 90%, $R_s = 0.89$, $p \leq 0.05$) and moderate in the study with the geriatric patients (mean 77%, $R_s = 0.42$, $p \leq 0.01$). In the third study with the videotaped patient no significant differences were found in the mean DOS Scale ratings of experts and nurses. On the level of items differences in agreement were found in all three studies. The findings can be explained by the psychometric qualities of the scale or by influences of test procedures. Rating behavior that is known to fluctuate and rating behavior within regular care can give several biases. For testing the DOS Scale, however, these possible biases were accepted since a more controlled test would not guarantee good reliability of the scale when used by nurses in regular care.

The question 'Is the Delirium Observation Screening Scale a valid and reliable instrument to detect delirium by nurses' observation during regular care?', was answered positively.

Therefore the results of the study of the hip fracture patients were further analyzed to come to a limited number of observations (Chapter 6). Ratings of DOS Scale items of patients before diagnosed delirious and of non-delirious patients were analyzed per shift (day, evening and night) on mean item ratings, on variances of item ratings and on proportion of ratings that indicate altered behavior. An algorithm was developed with thirteen items that showed the best discriminative value in these analysis. This algorithm showed a sensitivity of 100% and a specificity of 68% with nurses' ratings over two days. With this result a positive answer to the central question of this study can be given. Yes, to come to the diagnosis of delirium with a limited number of standard observations by nurses is possible.

Apart from systematic observation of symptoms, early recognition of delirium could possibly be enhanced by identification of patients at risk. Since two studies of this thesis were prospective in high risk groups (geriatric patients and hip fracture patients), data were collected also on predisposing and precipitating risk factors for delirium in these patients.

These data were used to answer the additional question:

Is identification of risk factors useful for early recognition of delirium?

Predisposing factors studied included age, gender, functional impairment, sensorial impairment, pre-existing cognitive impairment, comorbidity and use of medication. Precipitating factors studied included immobility, malnutrition, more than three medications newly prescribed, use of a bladder catheter and any iatrogenic event. In the study of the hip fracture patients factors related to the surgery were studied. These included waiting time till surgery, type and duration of surgery, type and duration of anesthesia, and complications.

In the group of geriatric patients only three patients developed delirium during the first week of admission and one patient on day eight (Chapter 8). This number was too low to meaningfully compare delirious and non-delirious patients. The group of non-delirious geriatric patients (n=78), however, rated high on predisposing and precipitating factors. Their mean age was 82.7 years, more than 50% needed much assistance or were fully dependent in their ADL's, 57% had sensory impairments and 69% had pre-existing cognitive impairment. Nearly all non-delirious geriatric patients had three or more diagnosis. One or more precipitating factors occurred during the first week of admission in 68% on the non-delirious patients. This study showed a low percentage of delirium in a high risk group. A preventive effect of the patient centered care in a department of geriatric medicine is a possible explanation for this finding.

In the study of the hip fracture patients significant differences in predisposing risk factors were found between the patients who developed delirium (n=18) and those who did not (n = 74) (Chapter 7). Patients with pre-existing ADL dependency ($p = 0.02$), psychiatric comorbidity (including dementia) ($p = 0.006$) and a high number of comorbid problems ($p = 0.02$) were more at risk for the development of delirium. Of the patients who had a prolonged waiting time before surgery (over 48 hours) 43% developed a delirium. This finding was, however, not significant.

The answer to the research question, 'is identification of risk factors useful for early recognition of delirium?' is yes in the study of the hip fracture patients and no in the study of the geriatric patients.

Based on the findings of these studies standard observation by nurses during regular care is recommended (Chapter 9). Use of the DOS Scale will influence and guide nurses'

observations of elderly patients, resulting in early recognition of delirium. Further study of the DOS Scale in different samples of elderly patients should confirm the sensitivity and specificity of the algorithm. Assessment shortly after admission of a patient's cognitive, somatic and functional status before admission is recommended. This information enables to focus at patients at risk and to provide patient centered care that might prevent patients from developing delirium. The possible preventive effect of care, as given in the structured approach of a department of geriatric medicine, deserves further study.

Samenvatting

Het onderzoek dat in dit proefschrift wordt beschreven is gericht op vroegtijdig herkenning van het delirium. Delirium is veel voorkomende vorm van psychopathologie bij oude patiënten na opname in het ziekenhuis (hoofdstuk 1). De gevolgen van een delirium kunnen aanzienlijk zijn: hoge mortaliteit en morbiditeit, langere verblijfsduur in het ziekenhuis en opname in een verpleeghuis. Als gevolg van een tekort aan kennis en bewustzijn bij verpleegkundigen en artsen wordt delirium vaak niet herkend of mis-gediagnosticeerd als dementie of depressie. Vroegtijdige herkenning van het delirium maakt diagnostiek en behandeling van de onderliggende etiologische factor(en) mogelijk en daarmee zouden de gevolgen van het gestoorde gedrag kunnen worden voorkomen.

In de Diagnostic Statistical Manual (DSM) IV van de American Psychiatric Association worden de volgende diagnostische criteria voor delirium omschreven: een stoornis in bewustzijn met verminderd vermogen om aandacht te concentreren, vast te houden of te verplaatsen en een verandering in cognitieve functie of de ontwikkeling van een waarnemingsstoornis. Delirium ontstaat in een korte tijd, de symptomen fluctueren en de eerste symptomen ontstaan meestal in de avond of gedurende de nacht. Daarom kunnen observaties van verpleegkundigen een bepalende rol spelen in de diagnosestelling ervan.

De centrale vraag van dit onderzoek was:

Is het mogelijk om tot de diagnose delirium te komen op basis van een beperkt aantal standaard observaties door verpleegkundigen?

Allereerst werden de eigenschappen en psychometrische kwaliteiten (test procedures en resultaten) van twaalf in de literatuur beschreven delirium instrumenten, bestudeerd (hoofdstuk 2). De instrumenten verschillen met betrekking tot het doel van de meting (diagnostiek, screening ernst van symptomen), het soort gegevens waarmee het instrument ingevuld wordt (observatie, interview of testen van patiënten), de benodigde deskundigheid om het instrument te kunnen gebruiken, het aantal items en de benodigde tijd om het instrument toe te passen. De meeste instrumenten zijn gebaseerd op DSM criteria en meten de symptomen zoals de DSM beschrijft. Veel verschil werd gevonden in de mate waarin betrouwbaarheid en validiteit was getest en de kwaliteit van de gevolgde procedures.

Drie instrumenten zijn gebaseerd op observatie en ontwikkeld voor gebruik door verpleegkundigen: de Clinical Assessment of Confusion A, de Confusion Rating Scale en de NEECHAM Confusion Scale. Maar geen van deze instrumenten is gebaseerd op de DSM criteria. Vanwege methodologische problemen werd geen van deze instrumenten geschikt geacht om de centrale vraag van dit onderzoek te beantwoorden. Daarom werd een nieuw instrument ontwikkeld en getest op psychometrische kwaliteiten.

Hiertoe werd een aanvullende onderzoeksvraag geformuleerd:

Is de Delirium Observatie Screening (DOS) Schaal een valide en betrouwbaar instrument om delirium te herkennen op basis van verpleegkundige observatie tijdens de reguliere zorgverlening?

De Delirium Observatie Screening (DOS) Schaal is een 25 item schaal gebaseerd op de DSM-IV criteria (diagnostische en gerelateerde). De DOS Schaal werd ontwikkeld om gedrag vast te leggen dat door verpleegkundigen geobserveerd kan worden tijdens reguliere zorgverlening. Het betreft gedrag dat door iedere verpleegkundige met basis kennis van de geriatrie geobserveerd kan worden. Het instrument is ook geschikt voor toepassing bij patiënten met reeds bestaande cognitieve stoornissen. Het invullen van het instrument kost minder dan 5 minuten.

De psychometrische kwaliteiten van de DOS Schaal werden in verschillende studies onderzocht:

- Inhoudsvaliditeit werd vastgesteld door een panel van zeven experts met behulp van de twee fasen procedure zoals beschreven door Lynn (hoofdstuk 3).
- Interne consistentie, predictieve validiteit, criterium gerelateerde validiteit en constructvaliditeit werden getest in twee prospectieve studies bij patiënten met een hoog risico op delirium: 82 geriatrische patiënten (4 werden delirant) en 92 heupfractuur patiënten (18 werden delirant) (hoofdstuk 3). In deze studies vulden de verpleegkundigen gedurende de eerste week van opname de DOS Schaal aan het einde van iedere dienst (dag, avond, nacht) in op basis van hun observaties gedurende die dienst (8 uur). Een geriater stelde op basis van de DSM-IV criteria de diagnose delirium. Deze diagnose werd gebruikt ter bepaling van de predictieve validiteit. De criterium gerelateerde validiteit werd in beide studies getest met behulp van de Mini Mental State Examination (MMSE). In de studie met de heupfractuur patiënten werd daarnaast de vergelijking gemaakt met de DOS Schaal en de Confusion Assessment Method (CAM) die beiden waren ingevuld door een onderzoeksverpleegkundige. De onderzoeksverpleegkundige vulde de schalen in op basis van een ongeveer 15 minuten durend bezoek aan de patiënt. Construct validiteit werd getest door correlatie met variabelen die samenhangen met het voorkomen van delirium: cognitieve functie (de Informant Questionnaire of Cognitive Decline in the Elderly (IQCODE)), een preëxistente psychiatrische diagnose en activiteiten van dagelijks leven (de Barthel Index).
- Interne consistentie en predictieve validiteit werd ook getest door een eenmalige vaststelling van de DOS Schaal bij 57 patiënten (33 delirant) die waren doorverwezen voor psychiatrische consultatie in een ziekenhuis (hoofdstuk 5). Verpleegkundigen vulden

de DOS Schaal aan het einde van een dagdienst in, zonder de door de consulent psychiatrie gestelde diagnose te weten.

- Interbeoordelaarsbetrouwbaarheid werd getoetst in drie studies (hoofdstuk 4). In twee studies werd de DOS Schaal onafhankelijk ingevuld op basis van simultane observatie door twee verpleegkundigen tijdens de reguliere zorgverlening. Zij observeerden patiënten die waren verwezen voor psychiatrische consultatie (n=6) en geriatrische patiënten (n=36). In de derde studie werd gebruik gemaakt van een video-opname van een delirante patiënt. Vierenveertig verpleegkundigen en 5 verpleegkundig specialisten bekeken individueel de video en vulden vervolgens de DOS Schaal in.

Alle analyses werden gedaan op het niveau van items en somscore. Dit om uiteindelijk met behulp van de DOS Schaal items de mogelijkheid te onderzoeken om tot de diagnose delirium te komen met een beperkt aantal standaard observaties door verpleegkundigen.

In alle studies had de DOS Schaal een hoge interne consistentie (Cronbach's $\alpha > 0.90$). De predictieve validiteit, in vergelijking met de DSM-IV diagnose van een geriater of psychiater, was goed. Significante verschillen werden gevonden in som scores en in de meeste items in de studies met geriatrische patiënten, heupfractuur patiënten en patiënten verwezen voor psychiatrische consultatie. Concurrente validiteit met de MMSE liet een redelijke correlatie zien in de studie met de geriatrische patiënten ($R_s = 0.66$, $p \leq 0.001$) en een goede correlatie in de studie met de heupfractuur patiënten ($R_s = 0.79$, $p \leq 0.001$). Correlatie tussen de door de verpleegkundigen gescoorde DOS Schaal en de door de onderzoeksverpleegkundige gescoorde DOS Schaal was $R_s 0.54$ ($p \leq 0.001$), correlatie met de CAM was vergelijkbaar.

Correlatie met de IQCODE was $R_s 0.33$ ($p \leq 0.05$) in de studie met de geriatrische patiënten en $R_s 0.74$ ($p \leq 0.001$) in de studie met de heupfractuur patiënten. Correlatie met een preëxistente psychiatrische diagnose was respectievelijk $R_s 0.42$ en 0.43 ($p \leq 0.001$). Correlatie met de Barthel Index was $R_s - 0.26$ ($p \leq 0.05$, geriatrische patiënten) en $R_s - 0.55$ ($p \leq 0.001$, heupfractuur patiënten). De resultaten op item niveau lieten in alle studies verschillen zien in de mate van validiteit. Verschillen in de resultaten met betrekking tot de construct validiteit kunnen verklaard worden vanuit kenmerken van de bestudeerde patiënten en het aantal delirante patiënten in iedere studie.

Interbeoordelaars betrouwbaarheid werd met verschillend maten bepaald. De drie studies lieten hoge percentages overeenkomstige scores zien die 'normaal gedrag' indiceerden. Dit kan verklaren waarom er lage kappa waarden werden gevonden op item niveau in de studies met simultane observatie. Het percentage overeenstemming en de Spearman correlatie coëfficiënten lieten goede resultaten zien in de studie met de patiënten verwezen voor psychiatrische consultatie (gemiddeld 90%, $R_s 0.89$, $p \leq 0.05$) en redelijke resultaten in de studie met de geriatrische patiënten (gemiddeld 77%, $R_s 0.42$, $p \leq 0.01$). In de derde studie

met de video-opname lieten de gemiddelde scores van de verpleegkundigen en de experts geen significante verschillen zien. In de drie studies werden op item niveau verschillen gevonden in de mate van overeenstemming. Dit kan veroorzaakt zijn door de psychometrische kwaliteit van de schaal of door invloeden van de test procedures. Het beoordelen van gedrag dat fluctueert en het beoordelen van gedrag gedurende de reguliere zorgverlening kan de oorzaak zijn van verschillende biases. Om de DOS Schaal te testen werden deze mogelijk biases geaccepteerd, een meer gecontroleerde meting biedt geen garantie voor betrouwbaarheid van de schaal bij gebruik door verpleegkundigen gedurende reguliere zorgverlening.

De vraag of de Delirium Observatie Screening Schaal een valide en betrouwbaar instrument is om delirium te herkennen op basis van verpleegkundige observatie tijdens de reguliere zorgverlening werd positief beantwoord.

Op basis van deze bevindingen werden de resultaten van de studie met de heupfractuur patiënten verder geanalyseerd, om het aantal observaties te beperken (hoofdstuk 6). Gemiddelde item scores, variantie in item scores en percentages van item scores die wijzen op afwijkend gedrag van patiënten voor de diagnose delirium was gesteld, werden vergeleken met scores van niet-delirante patiënten per dienst (dag, avond, nacht). Een algoritme werd ontwikkeld op basis van dertien items die de beste discriminerende waarde lieten zien. Op basis van verpleegkundige observatie gedurende twee dagen, gaf dit algoritme een sensitiviteit van 100% en een specificiteit van 68%. Met dit resultaat werd een positief antwoord gegeven op de centrale vraag van dit onderzoek. Ja, het is mogelijk om tot de diagnose delirium te komen op basis van een gelimiteerd aantal standaard observaties door verpleegkundigen.

Los van systematische observatie van symptomen, zou vroegtijdige herkenning van het delirium mogelijk verbeterd kunnen worden door identificatie van risicopatiënten. Omdat twee prospectieve studies plaatsvonden in groepen van patiënten met een hoog risico (geriatrische patiënten en heupfractuur patiënten) werden bij hen aanvullende gegevens verzameld die betrekking hadden op predisponerende en precipiterende risicofactoren voor delirium. Deze gegevens werden gebruikt om een aanvullende onderzoeksvraag te beantwoorden:

Is identificatie van risicofactoren zinvol voor vroegtijdige herkenning van het delirium?

Daarbij werden de volgende predisponerende factoren onderzocht: leeftijd, geslacht, functionele beperkingen, sensorische beperkingen, preëxistente cognitieve beperkingen, co-morbiditeit en medicatie gebruik. De precipiterende factoren die werden onderzocht waren:

immobiliteit, ondervoeding, toevoeging van drie of meer nieuwe medicijnen, een blaaskatheter en een iatrogene gebeurtenis. In de studie met de heupfractuur patiënten werden factoren onderzocht die in verband stonden met de operatie. Deze factoren betroffen de wachttijd voor de operatie, type en duur van de operatie, type en duur van de anesthesie en eventuele complicaties. In de groep van geriatrische patiënten ontwikkelden slechts drie patiënten een delirium gedurende de eerste week van opname en een patiënt op de achtste dag (hoofdstuk 8). Dit aantal was te laag voor een zinvolle vergelijking tussen delirante en niet-delirante patiënten. De groep van niet delirante geriatrische patiënten (n=78), scoorde evenwel hoog op predisponerende en precipiterende risico factoren. De gemiddelde leeftijd was 82.7 jaar, meer dan 50% had veel hulp nodig bij de ADL of was volledig ADL afhankelijk, 57% had sensorische beperkingen en 69% had preëxistente cognitieve beperkingen. Vrijwel alle niet-delirante geriatrische patiënten hadden drie of meer diagnoses. In de eerste week van opname trad bij 68% van de niet-delirante patiënten een of meer precipiterende factoren op. In dit onderzoek werd een laag percentage delirium gevonden in een populatie met een hoog risico. Een mogelijke verklaring voor deze bevinding kan liggen in het preventieve effect van de patiëntgerichte zorg van de afdeling geriatrie.

In het onderzoek van de heupfractuur patiënten werden significante verschillen gevonden in de predisponerende risicofactoren bij patiënten die een delirium ontwikkelden (n=18) en bij diegenen waarbij dit niet het geval was (n=74) (hoofdstuk 7). Patiënten met reeds bestaande functionele beperkingen ($p = 0.02$), psychiatrische co-morbiditeit (inclusief dementie) ($p = 0.006$) en met een groot aantal comorbide problemen ($p = 0.02$) hadden meer kans op het ontwikkelen van een delirium. Van de patiënten die een langere wachttijd voor de operatie hadden (meer dan 48 uur) ontwikkelden 43% een delirium. Deze laatste bevinding was niet significant.

Het antwoord op de vraag 'Is identificatie van risicofactoren zinvol voor vroegtijdige herkenning van het delirium?' werd positief beantwoord in de studie met de heupfractuur patiënten en negatief in de studie met de geriatrische patiënten.

Op basis van de resultaten van genoemde studies wordt gestandaardiseerde observatie door verpleegkundigen gedurende reguliere zorgverlening aanbevolen (hoofdstuk 9). Het gebruik van de DOS Schaal beïnvloedt en stuurt de verpleegkundige observatie van oude patiënten waardoor delirium eerder herkend wordt. Verder onderzoek van de DOS Schaal in verschillende populaties van oude patiënten is nodig om de sensitiviteit en specificiteit van het algoritme te bevestigen. Vaststelling van de cognitieve, somatische en functionele status van een oude patiënt kort na opname wordt tevens aanbevolen. Deze informatie maakt het mogelijk om de aandacht te richten op risicopatiënten en maakt het tevens mogelijk de zorg

aan te passen aan de behoeften van de individuele patiënt. Dit kan mogelijk voorkomen dat patiënten een delirium ontwikkelen. Ook het veronderstelde preventieve effect van de zorg zoals gegeven op een afdeling geriatrie verdient nader onderzoek.

Dankwoord/ Acknowledgement

In 1994 begon ik te denken over dit onderzoek. De vele delirante oude mensen die ik tijdens de consulten in mijn functie als verpleegkundig specialist geriatrie in het ziekenhuis tegenkwam, vormden de aanleiding. Oude mensen, die naar het ziekenhuis kwamen voor diagnostiek en behandeling en die ergens in de tijd, als gevolg van een lichamelijk probleem, psychisch flink ontregeld raakten.

Tussen het denken over het onderzoek en dit proefschrift liggen vele jaren en de bijdragen van vele mensen die ik daar graag voor wil bedanken.

Als eerste wil ik mijn beiden promotoren noemen, prof. dr. Lillie Shortridge-Baggett en prof.dr. Sijmen Duursma. Lillie, I admire you're optimism and enthousiasm. You never doubted my ability to bring these studies to a good end. Also, you gave me the opportunity to work with your students and to meet leading colleagues in the United States. I want to thank you for the years of inspiration and trust. Sijmen, je bent voor mij een echte leermeester geweest. Je kritische houding, je precisie, de gesprekken over meer dan alleen het onderzoek en de kansen die je mij gegeven hebt, zijn voor mij van zeer grote waarde. Ik wil je bedanken voor alles wat ik van je heb mogen leren.

Het onderzoek startte op de afdeling Geriatrie van het UMC Utrecht. Ik wil alle collega's, in het bijzonder de verpleegkundigen en artsen van de afdeling, hartelijk danken voor hun enthousiaste meedenken, meewerken en vertrouwen in de goede afloop. De combinatie verpleegkundig specialist-onderzoeker is niet de meest gemakkelijke maar wat mij betreft wel de mooiste. Ik ben altijd blij geweest te kunnen werken in de klinische praktijk met patiënten en collega's die ervoor zorgen dat je als onderzoeker niet te ver afraakt van de essentie van de zaak: de zorg voor oude mensen. Marianne Lensink, toenmalig hoofdverpleegkundige, was als leidinggevende steeds een bron van motivatie en ik ben blij dat onze samenwerking is voortgezet in een goede vriendschap.

In de eerste jaren van het onderzoek heb ik veel steun gehad aan collega onderzoekers Bert Lendemeijer en Berno van Meijel. Als pioniers vanuit het AZU probeerden we onderzoek in de praktijk te houden, en tegelijkertijd de praktijk te interesseren voor onderzoek. Ik hoop dat onze gesprekken over de leer en het leven nog veel vervolg zullen krijgen.

Een aanzienlijk deel van het onderzoek vond plaats op de afdeling Chirurgie van het Diaconessenhuis Utrecht. Ik wil Marie-Louise Boogers en Geert Jan Clevers van harte

danken voor hun bereidheid met mij de uitdaging van dit onderzoek aan te gaan. De verpleegkundigen en het secretariaat van de afdeling wil ik in het bijzonder danken voor hun inzet om de gegevens te verzamelen. Ruth Pel, die in deze fase aan het onderzoek verbonden was als 'mijn' research nurse, wil ik heel hartelijk danken voor haar betrokkenheid, nauwkeurigheid en gezelligheid. Ik hoop dat we nog lang in hetzelfde vakgebied kunnen samenwerken.

Vanaf de start op de afdeling Geriatrie en gedurende de jaren van onderzoek was Jan van der Woude altijd een goede partner. Zijn jarenlange kennis en ervaring in de zorg voor delirante mensen maakten hem de persoon om ideeën bij te toetsen en mee uit te werken. Het avontuur met de video's van delirante patiënten had zonder hem nooit kunnen plaatsvinden. Mijn dank voor deze mooie samenwerking.

In de laatste fase van het onderzoek, het analyseren en schrijven, ben ik 'verhuisd' naar de divisie Verplegingswetenschap. Graag wil ik al mijn collega's van de divisie danken voor hun luisterend oor, advies en morele steun. In het bijzonder Hannie Ketelaar en Mieke Grypdonck, die zich bijzonder inspanden om mij binnen het UMC Utrecht een plaats te geven om mijn onderzoek af te kunnen maken. Marjolein van Vliet, mijn maatje in de laatste fase van het proces dat promoveren heet, wil ik danken voor het delen van de smart die daardoor half werd.

Naast genoemde personen zijn er nog vele andere professionals die direct of indirect een bijdrage hebben geleverd aan de tot stand koming van dit proefschrift. Mijn dank gaat uit naar ieder van hen. Een aantal mensen wil ik graag bij naam noemen en persoonlijk danken, te weten: de leden van de delierwerkgroep van de Leo Cahn Stichting, in het bijzonder Roos van der Mast, Piet Eikelenboom en Joris Slaets, voor het delen van hun kennis en hun enthousiasme voor het delirium onderzoek; de studenten Verplegingswetenschap Nijnke de Goeijen en Yke Kooistra die een deel van het onderzoek mee vormgaven, Gabriël Roodbol die als student van de Masters of Science in Nursing eveneens een bijdrage leverde aan het onderzoek, Patricia Deschamps and Susan Markham who collaborated in the literature review as part of their Masters preparation, Ferdy Balk die vanuit de Psychiatrische Consultatieve Dienst meewerkte aan het onderzoek, Thomas Borggreve die heel knap een delirante patiënt wist te simuleren ten behoeve van de video, alle verpleegkundigen en experts die deze video beoordeelden, Leo Nijo die heel precies alle gegevens invoerde, Rogier Donders die als statisticus een belangrijke rol speelde bij de uiteindelijke bewerking van de gegevens en tenslotte de achtereenvolgende managers van de Divisie Interne

Geneeskunde, Ed Baten en Janneke van Vliet, voor hun visie op en bijdrage aan het tot stand komen van verplegingswetenschappelijk onderzoek in de praktijk.

Naast alle genoemde personen had dit onderzoek nooit tot stand kunnen komen zonder de bijdrage van enerzijds de patiënten en hun familie en anderzijds de subsidiegever. De patiënten en hun familie wil ik danken voor bereidheid om mee te werken aan een onderzoek over een beladen onderwerp, in een voor hen moeilijke periode in hun leven. Het Landelijk Centrum Verpleging en Verzorging wil ik niet alleen danken voor de toegekende subsidie, maar ook voor de persoonlijke stimulans in de vorm van de Leergang voor Leiderschap 'Orion' waaraan ik ten tijde van dit onderzoek kon deelnemen.

Jaren heb ik geleefd met dit onderzoek. Een verhouding die gekenmerkt werd door enthousiasme, twijfels, vallen en weer opstaan. Een grote drijfveer bleef steeds de liefde voor mijn vak: het verplegen van oude mensen. In het bijzonder verpleegkundige zorg op basis van weten in plaats van op basis van gewoonte. De drive om te laten zien dat de bijdrage van verpleegkundigen in de door medische technologie en ambtelijke bureaucratie gedirigeerde gezondheidszorg er een is die verder reikt dan coördinatie en menselijke warmte. Dat verplegen een vak is met inhoud, een vak om trots op te zijn. Ik prijs me gelukkig dat de mensen in mijn directe omgeving mij ruimte bieden voor deze passie.

Graag wil ik mijn familie en vrienden bedanken voor hun stimulans en steun. In het bijzonder mijn ouders, lieve Cor en Ank, jullie hebben me laten zien dat je het leven zelf leeft, dat je je keuzes zelf maakt en dat geloof in eigen kunnen het begin van alles is. En dat dat goed is. Van mijn vriendinnen wil ik Desiree en Rianne bij naam noemen. Twee verschillende vriendschappen die voor mij heel belangrijk zijn. Lieve Dees, hoewel onze levens heel verschillend zijn vinden we bij elkaar vrijheid, vertrouwen, steun in moeilijke momenten en veel lol. Je bent er altijd voor me, dat is een heerlijk gevoel. Lieve Rianne, mijn vriendin in de wereld van onderzoek, proefschriften, werk en alles wat het leven van de hoog opgeleide ambitieuze vrouw moeilijk en mooi maakt. Je hulp bij het schrijven van dit proefschrift was van grote waarde en ik ben blij dat jij mij, samen met Rolf, terzijde zal staan tijdens de verdediging.

Mijn grootste dank gaat uit naar Rolf, mijn lief. Toen we elkaar leerden kennen had ik net de eerste ideeën over dit onderzoek. Dat dit een promotietraject zou worden, was duidelijk. Je was enthousiast en bent dat gebleven. Als ik twijfelde, jij niet. Je relativerende kijk, humor maar vooral je eindeloze liefde en vertrouwen maken me sterk.

