International Psychogeriatrics (2011), 23:1, 102–106 © International Psychogeriatric Association 2010
doi:10.1017/S1041610210000323

Assessment of memory function: the relation between daily observation and neuropsychological test performance

A. Persoon,[1] R. P. C. Kessels,[2,3] L. Joosten-Weyn Banningh,[2] J. Verkoelen,[1] T. van Achterberg[4] and M. G. M. Olde Rikkert[1]

[1] *Department of Geriatrics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands*
[2] *Department of Medical Psychology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands*
[3] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands*
[4] *Scientific Institute for Quality of Healthcare, Nursing Science Section, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands*

ABSTRACT

Background: The aim of the study was to explore the value of a daily observation scale in the assessment of patients' memory function by nurses on a geriatric ward.

Methods: An observational study of 50 geriatric inpatients was carried out. The relationship between the memory items of the Nurses' Behavioral Rating Scale for Geriatric Inpatients (GIP) and four types of neuropsychological memory tests was examined: visual paired-associate learning (Visual Association Test, VAT), word-list learning (Eight Word Test, 8WT from the Amsterdam Dementia Screening, ADS), and the subtests Route Recall and Story Recall from the Rivermead Behavioural Memory Test (RBMT). Correlations with the overall measures assessing level of dementia such as the Mini-mental State Examination (MMSE), Clinical Dementia Rating scale (CDR) and the 15-item Geriatric Depression Scale (GDS-15) were examined as well.

Results: The Pearson's correlation coefficients between GIP and the four memory tests were between 0.45 and 0.71 ($p < 0.01$). The GIP correlations with the MMSE and CDR were 0.63 and 0.46, respectively ($p < 0.01$). No significant correlation was found with the GDS-15. Statistically significant differences in GIP memory scores between patients with dementia and non-demented patients were found ($p < 0.01$).

Conclusions: Results indicate that an observation scale of memory function may have value for providing information about the underlying memory impairment. The results of nurses' observations may be used in triage contributing to the diagnostic process by selecting patients requiring further neuropsychological assessment.

Key words: behavioral rating scale, memory, observation scale, validity, nursing

Introduction

Behavioral observation in an inpatient setting has been claimed to contribute to cognitive assessment, supplementing information obtained from medical history and cognitive testing (Spiegel *et al.*, 1991; Dellasega, 1998; Langley, 2000; Inouye *et al*, 2001; Foreman *et al*, 2003; Bouwens *et al.*, 2008). The level and nature of everyday cognitive functioning can provide valid information for medical diagnoses, nursing interventions, interpretation of neuropsychological tests, discharge arrangements and placement (Langley, 2000; Marson and Herbert, 2006). In nursing, daily practice involves observing admitted patients' behavior around the clock during patient-care activities, such as assisting with personal care, transfers or meal times. It has been claimed that reliable and valid assessment of cognitive functioning through daily observation is problematic, because overt behavioral deficits and disabilities may not be directly linked to a specific underlying neurocognitive impairment (Lezak *et al.*, 2004). For instance, observed "forgetting" may be the result of memory impairment, or it may also be due to cognitive deficits in other domains (e.g. impaired attention) or psychiatric symptoms (such as depression). In daily practice, however, monitoring of patients' behavior can provide a basis for initial screening for cognitive decline.

Correspondence should be addressed to: Anke Persoon, Department of Geriatrics – 925, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands. Phone: +31-24 361 6772; Fax: +31-24-361 7048. Email: A.Persoon@ger.umcn.nl. Received 23 Sep 2009; revision requested 4 Dec 2009; revised version received 1 Feb 2010; accepted 2 Feb 2010. First published online 3 June 2010.

https://doi.org/10.1017/S1041610210000323 Published online by Cambridge University Press

Table 1. NOSGER and GIP memory subscales

NOSGER memory scale (community and geriatric institution setting)

The patient:

- .. remembers a point in conversation after interruption.
- .. repeats the same point in conversation over and over.
- .. remembers names of close friends.
- .. remembers where clothes and other things are placed.
- .. confuses the identity of some people with others.

GIP memory subscale (geriatric institution setting)

The patient:

- .. knows other patients by name.
- .. knows close family members by name.
- .. remembers what has been said or asked.
- .. forgets that he/she is admitted/ institutionalized (and not here as a visitor).
- .. knows his/her own name.
- .. appears to forget his/her current activity.
- .. appears to recognize staff members.

In this study, we focused on the validity of daily observation of memory function by means of a standardized observation scale in relation to neuropsychological memory tests. Although many behavioral scales include items addressing memory function, observation scales including a specific memory subscale are limited in number. Examples of memory (sub)scales are the Cognitive Problem Scale (CPS), part of the Resident Assessment Instrument (RAI; Morris *et al.*, 1994) and the Disoriented Behaviour subscale of the Multidimensional Observation Scale for Elderly Subjects (MOSES; Helmes *et al.*, 1987). Furthermore, memory (sub)scales are typically validated against nonspecific cognitive measures, such as the Mini-mental State Examination (MMSE). Only the Nurses' Observation Scale for Geriatric Patients (NOSGER) memory subscale has been additionally validated against several specific neuropsychological memory tests (Spiegel *et al.*, 1991; Wahle *et al.*, 1996), demonstrating satisfying concurrent validity. As we are interested in the value of observation scales as a way to assess memory function, we also explored the value of other observation scales. In the present study, we examined the construct validity of the memory subscale of the Nurses' Behavioural Rating Scale for **G**eriatric **Inp**atients (GIP) (Verstraten, 1988). In Dutch nursing homes, the GIP is widely used in geriatric clinical practice and its psychometric qualities have been studied extensively (de Jonghe *et al.*, 1994; 1995; 1996). The GIP memory subscale focuses on geriatric inpatients as well as the NOSGER, but the subscale items differ (see Table 1).

Methods

Sample and setting

Consecutive patients admitted to the geriatric ward of the Radboud University Nijmegen Medical Centre from July to December 2007 were recruited if they fulfilled the inclusion criteria of understanding simple instructions and being able to complete the 45-minute neuropsychological test battery. Further, exclusion criteria were applied as follows: having delirium (Delirium Observation Scale/DOS < 3; Schuurmans *et al.*, 2003); being bedridden; having severe dementia (Clinical Dementia Rating Scale/CDR $= 3$; Hughes *et al.*, 1982); having severe hallucinations; and not being able to complete the 45-minute neuropsychological test battery. All nurses participating in the study were registered nurses ($n = 25$), and some were qualified clinical geriatric nurses with a two-year specialist training ($n = 15$).

GIP

For the memory observation scale, we used the six items from the Memory subscale from the GIP (range 0 to 3; lower scores indicating more memory problems). The GIP contains 82 items in total, subdivided into 14 subscales on a 4-point Likert scale, covering aspects of cognitive and non-cognitive functioning in geriatric patients. The content validity of the total GIP has been assessed; the construct validity showed satisfactory correlation with the Cognitive Screening Test (de Jonghe *et al.*, 1996). The internal consistency was good (Cronbach's $\alpha = 0.86$), while the inter-rater reliability of the memory subscale was fair to moderate ($\kappa = 0.47$). Furthermore, the memory subscale ratings were highly related to clinical diagnosis: mean scores in patient groups with dementia, psychosis, depression were 9.54, 3.61 and 1.72, respectively. In our study, the nurse observed the patient in routine clinical care during an eight-hour shift. The memory subscale sheet was filled in during day and evening shifts for three consecutive days, thereby yielding six assessments per patient. The GIP memory score was calculated by computing the mean of the six items per observation period, and then computing the mean of the six observation periods, resulting in a score between 0 and 3, with low scores indicating more memory problems.

Data gathering and analysis

In a 45-minute neuropsychological test battery, four different types of memory tasks were conducted: visual paired-associate learning, word-list learning, and story recall and spatial memory. The following

https://doi.org/10.1017/S1041610210000323 Published online by Cambridge University Press

tests, which are valid and reliable measures of different aspects of memory function and are widely used in Dutch clinical geriatric practice, were selected: Visual Association Test (VAT; Lindeboom *et al.*, 2002), the Eight-Word Verbal Learning Test (8WT, from the Amsterdam Dementia Screening; ADS; van der Pol-Vetten and Liem, 1992), and the subtests Route Recall and Story Recall from the Rivermead Behavioural Memory Test (RBMT; Wilson *et al.*, 1985), respectively. The VAT is a brief visual associative learning task based on imagery mnemonics (range 0–12). The 8WT measures the quality of learning a list of words in several rounds (range 12–55). The Route Recall (range 0–16) and Story Recall (range 0–42) subtests of the RBMT provide ecologically valid measures of spatial memory and verbal memory respectively (Lezak *et al.*, 2004). All tests were administered by a trained junior psychologist who was blind to the memory GIP observation ratings, on one of the three days of the GIP observation period.

Additionally, we determined the global cognitive status using the MMSE (Folstein *et al.*, 1975), and screened for depressive symptoms through the Geriatric Depression Scale (GDS-15; Yesavage *et al.*, 1982), both administered together with the neuropsychological tests. Severity of dementia was determined by means of the CDR score on the day of discharge. Comorbidity was measured through the Cumulative Illness Rating Scale-Geriatric (CIRS-G; range 0–52; higher scores indicate a higher level of comorbidity; Miller *et al.*, 1992).

Pearson's correlation coefficients were computed to determine the correlation between the scores on the memory subscale and the four neuropsychological test performances. Statistically significant coefficients higher than 0.4 were considered as meaningful correlations (Slick, 2006). We expected no significant correlation between the memory observation score and the depression scale. Differences in GIP memory scores in groups without dementia (CDR = 0) and with dementia (CDR = 1 or 2), in groups with different level of memory impairment (MMSE > 23, MMSE = 18–23; and MMSE<18), and in groups with and without depressive symptoms (GDS > 5, GDS ≤ 5) were analyzed by means of t-test and ANOVA test.

Sample

During the study period 129 patients were admitted to the geriatric ward, 63 of whom fulfilled the inclusion criteria. For 13 patients, it was not possible to collect data because of early discharge or logistic reasons, resulting in a total of 50 patients who completed the full test battery. Of these, 64% were female; the mean age was 83 years (SD = 6.7, range 72–97). The lowest level of education was primary school (30%), and the highest level was university education (10%). The mean number of different drugs per day was 4.6 (SD = 2.5, range 0–10).

Results

The mean GIP memory score was 2.3; the minimum score was 1.4 and the maximum 2.9,

Table 2. Correlations between daily observations, four neuropsychological memory tests, MMSE, CDR and GDS-15

				DAILY OBSERVATION	NEUROPSYCHOLOGICAL TESTS			
THEME	SCALE (RANGE)	n	X (SD)	GIP	VAT	8WT	RBMT (ROUTE)	RBMT (STORY)
Memory impairment	Visual learning/VAT (0–12)	49	6.6 (4.4)	0.76**	–	–	–	–
	Word learning/8WT (12–55)	47	33.3 (11.2)	0.60**	0.80**	–	–	–
	Route recall/RBMT (0–16)	48	6.8 (4.0)	0.52**	0.51**	0.54**	–	–
	Story recall/RBMT (0–42)	50	8.0 (8.1)	0.51**	0.59**	0.66**	0.65**	–
Cognitive function	MMSE (0–30)	48	21.3 (5.8)	0.64**	0.72**	0.67**	0.56**	0.56**
Severity of dementia	CDR (0–3)	40	1.0 (0.8)	−0.47**	0.57**	0.51**	−0.43*	−0.50**
Depression	GDS-15 (0–15)	48	4.2 (2.4)	−0.08	−0.15	0.39**	−0.29	−0.30*

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).
GIP = Nurses' Behavioral Rating Scale for Geriatric Inpatients; VAT = Visual Association Test; 8WT = Eight Word Test; RBMT = Rivermead Behavioural Memory Test; MMSE = Mini-mental State Examination; CDR = Clinical Dementia Rating scale; GDS = Geriatric Depression Scale.

https://doi.org/10.1017/S1041610210000323 Published online by Cambridge University Press

standard deviation 0.42 (range 0–3, lower scores indicating more memory problems) (n = 50). The intraclass correlation between the six measurement moments was 0.59. The mean severity score of comorbidity was 14 (SD = 5.4, range: 3–28) on the CIRS-G scale, reflecting the considerable disease burden. According to the CDR criteria, 36% of the respondents had no dementia (CDR = 0), 30% had questionable or mild dementia (CDR = 0.5 or 1), and 30% moderate dementia (CDR = 2). A quarter of the patients reported significant depressive symptoms on the GDS-15 (GDS > 5). Based on the MMSE, 38% of the respondents had no or mild indications for cognitive impairment (MMSE $\geq$ 24), 32% had moderate impairment (MMSE 18–23) and 30% had severe cognitive impairment (MMSE $\leq$ 17).

Analyses revealed that the correlation between the results on the memory observation scale and the performance on four neuropsychological memory tests were statistically significant and higher than 0.4 (between 0.45 and 0.71), as shown in Table 2. The correlations among the four memory tests themselves varied between 0.51 and 0.80 ($p < 0.05$). Furthermore, the correlation with the MMSE was 0.63, and with the dementia severity scale (CDR) it was −0.46, both statistically significant ($p < 0.01$). No significant correlation was found between the memory observation scale and depression score GDS-15 ($p > 0.05$).

The mean GIP memory scores differed statistically significantly ($p < 0.01$) between patients with dementia and those without dementia: 2.0 and 2.4, respectively. The mean GIP memory scores also differed statistically significantly ($p < 0.5$) between different groups of memory impairment: 2.6, 2.4 and 1.9 respectively. The mean GIP memory scores in the group with or without depressive symptoms did not differ, being 2.3 and 2.4, respectively.

Discussion

This study is one of the first studies to explore the value of an observation scale for daily memory function. Correlations between the daily observation scale scores and neuropsychological tests performances were all in the fair to moderate range. The correlations between GIP and the neuropsychological tests were similar in strength to the correlations among the four memory tests, suggesting that the daily observation may be a valid way to assess memory function. This is in line with the NOSGER study in which results of observation items were correlated with results of neuropsycological tests (Spiegel *et al.*, 1991; Wahle *et al.*, 1996). However, the correlations between observation scores and measures for global cognitive decline (MMSE and CDR) were moderate as well. This may suggest that the observation scale may be tapping global cognitive status rather than specific memory impairment. To examine this in greater depth, we recommend comparing the results of the observation scale with neuropsychological tests examining other cognitive domains (such as attention or executive function).

A strength of the observation scale is that no correlation was found with the depression scores. As differentiation between memory impairment and depression is rather complicated in clinical practice, this finding is an important added value.

Overall, the results of our study are in line with the results of the NOSGER validation study which found that an observation scale for memory function may have value for providing information on underlying memory impairment. The bedside nurses needed no extra education to fill out the GIP form and they reached fair to moderate levels of correlation with neuropsychological tests. The results of our study can be generalized to geriatric inpatients with considerable variety in the severity of cognitive decline and overall disease burden. We recommend further research for the validity of observation scales in other settings of elderly care (psychiatric hospital, home care, nursing homes, homes for the elderly, and outpatient clinics).

A valid and reliable assessment of cognitive function by nurses is important because nurses tailor their interventions to the patient's cognitive abilities, and are often asked to contribute to the diagnostic process as well. The observational results may be used in triage, thereby contributing to diagnosis by pointing to the need for neuropsychological assessment. Daily observation is an easy, inexpensive and highly feasible assessment procedure for nurses and can be conducted continuously during daily care. It is not burdensome to the patient and can even be performed with very ill patients. Therefore, memory observation scales have significant potential to contribute to the quality of care for the increasing populations of older patients.

Conflict of interest

None.

Description of author's roles

Anke Persoon, Roy Kessels and Liesbeth Joosten-Weyn Banningh designed the study. Jill Verkoelen collected the data and analyzed the data together

https://doi.org/10.1017/S1041610210000323 Published online by Cambridge University Press

with Anke Persoon. Anke Persoon and Roy Kessels wrote the paper. Marcel Olde Rikkert and Theo van Achterberg critically reviewed the study design and the paper.

References

Bouwens, S. F. *et al.* (2008). Relationship between measures of dementia severity and observation of daily life functioning as measured with the Assessment of Motor and Process Skills (AMPS). *Dementia and Geriatric Cognitive Disorders*, 25, 81–87.

de Jonghe, J. F., Kat, M. G. and de Reus, R. (1994). Validity of the Behavior Observation Scale for Intramural Psychogeriatrics: a comparison with the BOP (Evaluation Scale Elderly Patients) and NOSIE-30 in a psychogeriatric assessment clinic for the elderly. *Tijdschrift voor Gerontologie en Geriatrie*, 25, 110–116.

de Jonghe, J. F., Kat, M. G., Rottier, W. P. and de Reus, R. (1995). The Behavior Observation Scale for Intramural Psychogeriatrics and clinical diagnosis; a comparison with the BOP (Assessment Elderly Patients) and NOSIE-30. *Tijdschrift voor Gerontologie en Geriatrie*, 26, 24–29.

de Jonghe, J. F., Calis, P. J. and Boom-Poels, P. G. (1996). Behavioral dimensions in elderly patients: factor structure of the Behavioral Observation Scale for Intramural Psychogeriatrics. *Tijdschrift voor Gerontologie en Geriatrie*, 27, 159–164.

Dellasega, C. (1998). Assessment of cognition in the elderly: pieces of a complex puzzle. *Nursing Clinics of North America*, 33, 395–405.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

Foreman, M. D., Fletcher, K., Mion, L. C. and Trygstad, L. (2003). Assessing cognitive function. In M. Mezey, T. Fulmer, I. Abraham and D. A. Zwicker (eds.), *Geriatric Protocol for Best Practice* (pp. 99–115). New York: Springer Publishing Company.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A. and Martin, R. L. (1982). A new clinical scale for the staging of dementia. *Britisch Journal of Psychiatry*, 140, 566–572.

Helmes, E., Csapo, K. G. and Short, J. A. (1987). Standardization and validation of the Multidimensional Observation Scale for Elderly Subjects (MOSES). *Journal of Gerontology*, 42, 395–405.

Inouye, S. K., Foreman, M. D., Mion, L. C., Katz, K. H. and Cooney, L. M., Jr. (2001). Nurses' recognition of delirium and its symptoms: comparison of nurse and researcher ratings. *Archives of Internal Medicine*, 161, 2467–2473.

Langley, L. K. (2000). Cognitive assessment of older adults. In R. L. Kane and R. Kane (eds.), *Assessing Older Persons: Measures, Meaning, and Practical Applications* (pp. 65–128). New York: Oxford University Press.

Lezak, M. D., Howiesen, D. B., Loring, D. W., Hannah, H. J. and Fischer, J. S. (eds.) (2004). *Neuropsychological Assessment*. Oxford. Oxford University Press.

Lindeboom, J., Schmand, B., Tulner, L., Walstra, G. and Jonker, C. (2002). Visual association test to detect early dementia of the Alzheimer type. *Journal of Neurology, Neurosurgery and Psychiatry*, 73, 126–133.

Marson, D. and Hebert, K. R. (2006). Functional assessment. In D. K. Attix and K. A. Welsh Bohmer (eds.), *Geriatric Neuropsychology: Assessment and Interventions* (pp. 158–197). New York: The Guilford Press.

Miller, M. D. *et al.* (1992). Rating chronic medical illness burden in geropsychiatric practice and research: application of the Cumulative Illness Rating Scale. *Psychiatry Research*, 41, 237–248.

Morris, J. N. *et al.* (1994). MDS Cognitive Performance Scale. *Journals of Gerontology*, 49, M174–M182.

Persoon, A., Joosten, L., van de Vrie, W., Olde Rikkert, M. G. M. and van Achterberg, T. (2007). Agreement between geriatric nurses in the assessment of cognitive funtioning of older hospitalized patients. *Journal of the American Geriatrics Society*, 55, 1306–1307.

Schuurmans, M., Shortridge-Baggett, L. M. and Duursma, S. A. (2003). The delirium Observation Screening Scale: a screening instrument for delirium. *Research and Theory for Nursing Practice*, 1, 31–50.

Slick, D. J. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. S. Sherman and O. Spreen (eds.), *A Compendium of Neuropsychological Tests: Administration, Norms and Commentary* (pp. 3–26) Oxford: Oxford University Press.

Spiegel, R. *et al.* (1991). New behavioral assessment scale for geriatric out- and in-patients: the NOSGER (Nurses' Observation Scale for Geriatric Patients). *Journal of the American Geriatrics Society*, 39, 339–347.

van der Pol-Vetten and Liem, S. D. (1992). Usefulness and validity of the abbreviated Amsterdam Dementia Screening Test (ADS3) in nursing homes. *Tijdschrift voor Gerontologie en Geriatrie*, 23, 27–30.

Verstraten, P. F. (1988). 14 Observation scales for psychogeriatric behavior problems. *Tijdschrift voor Gerontologie en Geriatrie*, 19, 147–151.

Wahle, M., Häller, S. and Spiegel, R. (1996). Validation of the NOSGER (Nurses' Observation Scale for Geriatric Patients): reliability and validity of a caregiver rating instrument. *International Psychogeriatrics*, 8, 525–547.

Wilson, B. A., Cockburn, J. and Baddeley, A. (1985). *The Rivermead Behavioural Memory Test*. Reading, U.K.: Thames Valley Test Company.

Yesavage, J. A. *et al.* (1982). Development and validation of a Geriatric Depression Screening scale: a preliminary report. *Journal of Psychiatric Research*, 17, 37–49.

https://doi.org/10.1017/S1041610210000323 Published online by Cambridge University Press