Journal of the Neurological Sciences 320 (2012) 91–96

Contents lists available at SciVerse ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

# Validation and interpretation of the Dutch version of the Multiple Sclerosis Neuropsychological Screening Questionnaire

Judith M. Sonder [a,b,*], Lidwine B. Mokkink [a], Femke A.H. van der Linden [a], Chris H. Polman [b], Bernard M.J. Uitdehaag [a,b]

[a] Department of Epidemiology and Biostatistics, VU University Medical Center Amsterdam, The Netherlands
[b] Department of Neurology, VU University Medical Center Amsterdam, The Netherlands

ARTICLE INFO

*Article history:*
Received 15 December 2011
Received in revised form 3 April 2012
Accepted 28 June 2012
Available online 16 July 2012

*Keywords:*
Multiple sclerosis
Outcome measurement
Cognitive impairment
Depression
Screening
Neuropsychology
Proxy measurement

ABSTRACT

*Background:* The Multiple Sclerosis Neuropsychological Screening Questionnaire (MSNQ) was developed for screening of MS patients at risk for cognitive impairment with a patient self-report (MSNQ-P) and an informant version (MSNQ-I). The objective of this study was to validate the Dutch versions and determine their interpretability.

*Methods:* The MSNQ was completed by 121 MS patients and their partners (informants). We investigated the factor structure, internal consistency and construct validity. Interrater reliability between MNSQ-P and MSNQ-I was investigated with the intraclass correlation coefficient (ICC) and Cohen's kappa. For interpretability of both MSNQ versions we calculated sensitivity, specificity and cut-off scores. Receiver operating characteristic (ROC) curves with related area under the curve (AUC) were used to evaluate the added value of combining both versions.

*Results:* We found a unidimensional factor structure. Cronbach's alphas were 0.93 and 0.94 for MSNQ-P and MSNQ-I, respectively. The ICC was 0.59 and kappas were ≤0.50. No cut-off score could be defined for the MSNQ-P because of low sensitivity. For the MSNQ-I, sensitivity was 0.75 and specificity 0.71 (AUC 0.80). The cut-off score was 21. ROC curve analyses showed no added value of the MSNQ-P when used in combination with the MSNQ-I.

*Conclusions:* The MSNQ-I is preferred over the MSNQ-P to screen MS patients for cognitive impairment.



## 1. Introduction

Cognitive impairment in multiple sclerosis (MS) is common, occurring in 40–70% of MS patients [1]. It can occur in any stage of the disease and it is associated with problems in activities of daily living, employment, social functioning and a reduced quality of life [2].

Decline in cognitive functioning can be tested using neuropsychological tests. However, there are several disadvantages of neuropsychological testing. First, it is time-consuming and may be burden for some patients. Second, it is expensive since there is a need for qualified staff. Therefore a short screening instrument is needed to identify patients who might be cognitively impaired and in whom additional formal neuropsychological testing is useful [3]. Such a short screening instrument may even be used to indicate possible cognitive decline in MS patients in (long term) follow-up studies.

The Multiple Sclerosis Neuropsychological Screening Questionnaire (MSNQ) [3] is a self-report questionnaire that was developed to screen MS patients at risk for cognitive impairment. The MSNQ consists of 15 items and reflects neuropsychological competence during activities in daily living. In addition to the MSNQ for patient self-report (MSNQ-P), an informant form was developed (MSNQ-I) to compare the opinion of the patient about cognitive problems in daily life with the opinion of an informant.

Studies in the United States [3–5] and Argentina [6] tested the psychometric properties of both versions of the MSNQ. In these studies moderate to high Cronbach's alphas were found [3–6]. Test–retest correlations were only calculated in one study and they were also found to be high [4]. Construct validity was investigated by calculating the relationship of the MSNQ with neuropsychological testing and depression. All previous studies found high correlations between the MSNQ-P score and depression and low correlations between MSNQ-P score and cognitive impairment. In contrast, the MSNQ-I was positively correlated with cognitive impairment and not with depression of the patient.

Comparison of MSNQ scores with an instrument that measures cognitive impairment provides information about the validity of the (change) scores of the MSNQ. Interpretability is the degree to which

* Corresponding author at: VU University Medical Center, Department of Epidemiology & Biostatistics, PO Box 7057, 1007 MB Amsterdam, The Netherlands. Tel.: +31 20 4442915; fax: +31 20 4444475.
*E-mail address:* judith.sonder@vumc.nl (J.M. Sonder).

0022-510X 
doi:10.1016/j.jns.2012.06.024

one can assign qualitative meaning to quantitative scores [7]. The Brief Repeatable Battery for Neuropsychological Tests (BRB-N) [8–10] is widely used to measure cognitive impairment. A recent study comparing neuropsychological test batteries concluded that the BRB-N is reliable and sensitive to use in MS patients [11]. The BRB-N measures different aspects of cognition in five subtests. In MSNQ validation studies [3,4,6] cut-off scores were defined, based on measures of cognitive impairment, for the interpretability of MSNQ scores. These cut-off scores were different in all studies. In most of the validation studies no discriminative value was found for the MSNQ-P because of low sensitivity [3,5,6].

The first objective of this study was to validate the Dutch version of the MSNQ-P and MSNQ-I in a large sample of Dutch MS patients and informants. The psychometric properties were investigated by determining the structure of the MSNQ, internal consistency and construct validity. In addition, interrater reliability and agreement between patient and informant scores were investigated.

The second objective was to determine the interpretability of the MSNQ. Using sensitivity and specificity we tried to calculate cut-off scores for both versions to identify patients at risk for cognitive impairment. For that purpose, the BRB-N subtest scores were used as an external criterion. Receiver operating characteristic (ROC) curves with related area under the curve (AUC) were used to evaluate the added value of combining both versions.

## 2. Methods

### 2.1. Study sample

The subjects in this study were 121 MS patients and 121 informants, partners of MS patients. Patients were participating in ongoing research projects at the MS Center of the VU University Medical Center Amsterdam and their visit was already scheduled for these projects. The medical ethical committee of the VU University Medical Center approved the study protocol. The characteristics of the study sample are shown in Table 1.

### 2.2. Translation of MSNQ

A forward and backward translation procedure of the original English version [3] was followed to develop the Dutch version of the MSNQ patient and informant form. Both versions were translated into Dutch by two independent researchers followed by an independent backward translation into English. The backward translation was compared with the original version. In only three items minor differences were observed which were dealt with in consensus by adapting the wording of the Dutch version.

**Table 1**
Characteristics of the study sample.

| | Patient | Proxy |
|---|---|---|
| Male[a] | 46 (38) | 73 (60) |
| Female[a] | 75 (62) | 48 (40) |
| Age in years[b] | 53 (45–63) | 55 (45–63) |
| Relapsing Remitting (RR)[a] | 43 (36) | na |
| Secondary Progressive (SP)[a] | 40 (33) | na |
| Primary Progressive (PP)[a] | 34 (28) | na |
| Clinically Isolated Syndrome (CIS)[a] | 4 (3) | na |
| Disease duration in years[b] | 15 (9.0–22.0) | na |
| EDSS[b] | 4.5 (3.0–6.5) | na |
| High education (college/university)[a] | 37 (30) | 38 (31) |
| Moderate education (secondary school)[a] | 40 (33) | 42 (35) |
| Low educated (primary school)[a] | 42 (35) | 40 (33) |
| Completion of MSNQ in hospital setting[a] | 121 (100) | 87 (72) |

EDSS, Expanded Disability Status Scale.
[a] N (%).
[b] Median (IQR).

### 2.3. Tests and procedures

Patients were asked to complete the MSNQ-P and other questionnaires (described below) and were invited to perform a number of neuropsychological tests. The informant (partner of patient) was asked to complete the MSNQ-I in the hospital waiting room or it was completed at home and returned using a prepaid envelope.

The MSNQ versions for patient self report and informant contain the same items but from another perspective. The 15 MSNQ items have 5 response options, 0 (does not occur) to 4 (very often, very disruptive). A total score is computed with a range from 0 to 60, a higher score indicates more cognitive problems.

The Hospital Anxiety and Depression Scale (HADS) [12] is a 14-item self-report screening scale to indicate the possible presence of anxiety (7 items) and depressive states (7 items). For both continuous scales the range of scores is 0 to 21. The HADS is validated in different groups of Dutch subjects (three groups of healthy controls > 18 years and general, somatic and psychiatric patient groups) [13].

The BRB-N was developed as a short observational instrument to identify disturbances of cognitive domains in MS patients. The BRB-N (version A) consists of five subtests. Immediate and delayed recall memory is assessed by the Selective Reminding Test (SRT) [14]; visuospatial immediate and delayed recall memory is assessed by the Spatial Recall Test (SPART) [15]; the Rao version [9] of the Symbol Digit Modalities Test (SDMT) [16] examines processing speed and concentration by primarily assessing complex and visual scanning and tracking; the Rao version [9] of the Paced Auditory Serial Addition Test (PASAT) [17] measures information processing speed and interference suppression in the 2 and 3 second tests; and the Word List Generation (WLG) test [8] assesses semantic verbal fluency. For each of the five subtests a score is obtained. An abnormal score for each subtest was defined with the most stringent criterion, i.e. two standard deviations below the mean reported for Dutch healthy subjects [18]. Patients were classified in three levels of cognitive functioning, no impaired subtest score (no cognitive impairment), 1 or 2 impaired subtest score(s) (group at risk) and 3 or more impaired subtest scores (cognitively impaired) [19].

The Expanded Disability Status Scale (EDSS) [20] was used as a standardized neurological examination. The EDSS assesses neurological impairment and disability and was performed by a trained doctor. EDSS score varies between 0 and 10, a high score indicates a more severe disability.

## 3. Data analysis

### 3.1. Validity

Factor analyses (principal component analyses) were done to investigate the factor structure of each MSNQ version separately. We hypothesized that the MSNQ items load on one general factor. In addition to the principal component analysis we used parallel analysis to compare the unrotated eigenvalues from a random sample, with the same number of cases and variables, with our data. If the eigenvalues from the principal component analysis were found to be higher than the eigenvalues from the parallel analysis, it was significant and we could retain the factor.

We calculated Cronbach's alpha for both scales to determine their internal consistency. Cronbach's alpha above 0.70 was considered to be appropriate [21].

Construct validity was tested by defining hypotheses. It examines whether the MSNQ score represents the theoretical concept of interest, i.e. neuropsychological competence [22]. We tested the following hypotheses based on results from earlier studies [3–6]. Since the data were not normally distributed, Spearman correlations were used.

1. There is a small positive correlation (i.e. correlation between 0 and 0.40) between MSNQ-P score and amount of impaired subtests on the BRB-N.
2. There is a moderate and positive correlation (i.e. correlation between 0.40 and 0.70) between MSNQ-I score and amount of impaired subtests on the patient BRB-N.
3. There is a moderate and positive correlation (i.e. correlation between 0.40 and 0.70) between MSNQ-P and patient HADS score.
4. There is a small positive correlation (i.e. correlation between 0 and 0.40) between MSNQ-I and patient HADS score.
5. The correlation between MSNQ-I and amount of impaired cognitive tests is higher than the correlation between MSNQ-P and impaired cognitive tests (difference ≥ 0.10).
6. The correlation between MSNQ-P and patient depression is higher than the correlation between MSNQ-I and patient depression (difference ≥ 0.10).

Interrater reliability between patient and informant on scale level was evaluated using the intraclass correlation coefficient two-way mixed model for absolute agreement ($ICC_{agreement}$) [23]. Cohen's weighted kappa ($\kappa_w$) [24] was calculated to evaluate interrater reliability between patient and informant on item level, using quadratic weights. The weighted kappas were calculated for all items. ICC and kappa above 0.70 were considered to be appropriate [21].

Agreement was measured with the Limits of Agreement. In a Bland and Altman [25] plot the differences between patient and proxy scores were plotted against the mean score of each pair on the MSNQ. The limits of agreement (LoA) were calculated using the mean difference per scale ($\bar{d}$) and the standard deviation of the mean difference ($SD_{difference}$):

$$LoA = \bar{d} \pm 1.96 SD_{difference}.$$

### 3.2. Interpretability

One of the objectives of this study was to calculate cut-off scores for the MSNQ that best classifies patients in the study sample as cognitively impaired. The scores on the BRB-N subtests were used as a criterion to classify patients in three levels of cognitive functioning. We calculated sensitivity and specificity for each MSNQ version separately and for that we used 3 or more impaired subtest scores as cutoff for cognitive impairment. In addition we repeated the analyses using 1 or more impaired subtest score(s) as cutoff for cognitive impairment. We considered the optimal cut-off score for both MSNQ versions to be the value for which [1-sensitivity] + [1-specificity] is the smallest. Because the goal of this instrument is to screen for cognitive impairment, we considered sensitivity to be more important than specificity.

Receiver operating characteristic (ROC) curves were constructed using logistic regression models in which the presence of cognitive impairment was the dependent variable and MSNQ-P and MSNQ-I were entered either alone or both together as independent variables. We compared the area under the curve (AUC) of MSNQ-P, MSNQ-I and the combined line.

All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) version 14.0 (SPSS, Inc., Chicago, IL, USA) and Stata 11.1 (StataCorp, College Station, TX 2009).

## 4. Results

### 4.1. Missing items

Three patients and 4 informants had one missing MSNQ item. In these cases, the mean item score of the scale was imputed. In one case more than a single item was missing and this case was excluded from analyses.

**Table 2**
Factor loadings from principal component analysis.

| | | MSNQ-P | MSNQ-I |
|---|---|---|---|
| 1. | Distractibility | 0.71 | 0.69 |
| 2. | Problems with listening to others | 0.70 | 0.78 |
| 3. | Slowed processing | 0.74 | 0.77 |
| 4. | Forgetting appointments | 0.64 | 0.72 |
| 5. | Forgetting what is read | 0.78 | 0.83 |
| 6. | Forgetting shows or programs | 0.78 | 0.80 |
| 7. | Forgetting instructions | 0.79 | 0.86 |
| 8. | Needing frequent reminders | 0.76 | 0.86 |
| 9. | Failing to follow through on planned activities | 0.59 | 0.69 |
| 10. | Failing to answer questions coherently | 0.78 | 0.80 |
| 11. | Failing to track two things at once | 0.83 | 0.86 |
| 12. | Failing to follow conversations | 0.78 | 0.81 |
| 13. | Problems controlling impulses | 0.59 | 0.52 |
| 14. | Laughing or crying without cause | 0.57 | 0.55 |
| 15. | Excessive egocentric speech | 0.50 | 0.57 |

MSNQ-P, Multiple Sclerosis Neuropsychological Screening Questionnaire — patient version; MSNQ-I, Multiple Sclerosis Neuropsychological Screening Questionnaire — informant version.

### 4.2. Psychometric properties

The results of the factor analysis (principal component analysis) are displayed in Table 2. The items of both MSNQ versions loaded high on their first factor. The eigenvalue of this factor was 7.5 for MSNQ-P and explained 50% of the variance. The highest loading was 0.83 for item 11 (i.e. failing to track two things at once) and the lowest loading was 0.50 for item 15 (i.e. excessive egocentric speech). The first factor of the MSNQ-I explained 56% of the total variance and the eigenvalue was 8.4. The highest loading for the informant version was 0.86 for the items 7, 8 and 11 (i.e. forgetting instructions, needing frequent reminders and failing to track two things at once) and the lowest was 0.52 for item 13 (i.e. problems controlling impulses). In the parallel analyses was seen that the eigenvalues of this sample were much higher than the eigenvalue of a random sample. Some items also loaded on a second factor in the principal

**Table 3**
Spearman correlations.

| | Mean | SD | Correlations | |
|---|---|---|---|---|
| | | | MSNQ-P | MSNQ-I |
| MSNQ-P | 18.4 | 10.3 | na | 0.52 |
| MSNQ-I | 17.1 | 11.2 | 0.52 | na |
| Age | 52.8 | 11.7 | 0.20 | 0.19 |
| Gender patient | na | na | −0.03 | −0.24 |
| Education (level) | na | na | −0.13 | −0.13 |
| Disease duration (years) | 16.9 | 8.8 | 0.26 | 0.17 |
| EDSS[a] | 4.5 | 3.0–6.5 | 0.33 | 0.38 |
| HADS anxiety | 4.9 | 3.2 | 0.47 | 0.33 |
| HADS depression | 4.1 | 3.5 | 0.49 | 0.36 |
| BRB-N global (amount impaired out of 5 subtests) | 0.6 | 1.2 | 0.26 | 0.39 |
| SRT — immediate recall, long term storage | 46.0 | 14.8 | −0.28 | −0.29 |
| SRT — immediate recall, consistent long term storage | 35.6 | 15.9 | −0.22 | −0.34 |
| SRT — delayed recall | 46.1 | 16.0 | −0.17 | −0.36 |
| SPART — immediate recall (3) | 29.0 | 15.5 | −0.21 | −0.37 |
| SPART — delayed recall | 8.6 | 2.9 | −0.13 | −0.32 |
| SDMT | 19.8 | 5.1 | −0.06 | −0.02 |
| Pasat 3″ — correct | 6.8 | 2.4 | −0.12 | −0.14 |
| Pasat 2″ — correct | 47.1 | 15.1 | −0.27 | −0.34 |
| WLG | 24.4 | 7.4 | −0.17 | −0.30 |

MSNQ-P, Multiple Sclerosis Neuropsychological Screening Questionnaire — patient version; MSNQ-I, Multiple Sclerosis Neuropsychological Screening Questionnaire — informant version; EDSS, Expanded Disability Status Scale; HADS, Hospital Anxiety and Depression Scale; BRB-N, Brief Repeatable Battery for Neuropsychological Tests; SRT, Selective Reminding Test; SPART, Spatial Recall Test; SDMT, Symbol Digit Modalities Test; PASAT, Paced Auditory Serial Addition Test; WLG, Word List Generation.

[a] Median, IQR.

**Table 4**
Weighted kappa: interrater reliability of MSNQ items.

| | $\kappa_w$ | 95% CI |
|---|---|---|
| MSNQ 1 | 0.32 | 0.11–0.49 |
| MSNQ 2 | 0.36 | 0.17–0.52 |
| MSNQ 3 | 0.25 | 0.07–0.44 |
| MSNQ 4 | 0.39 | 0.22–0.58 |
| MSNQ 5 | 0.50 | 0.33–0.62 |
| MSNQ 6 | 0.50 | 0.35–0.65 |
| MSNQ 7 | 0.43 | 0.27–0.58 |
| MSNQ 8 | 0.43 | 0.24–0.58 |
| MSNQ 9 | 0.28 | 0.11–0.44 |
| MSNQ 10 | 0.47 | 0.30–0.62 |
| MSNQ 11 | 0.49 | 0.32–0.64 |
| MSNQ 12 | 0.46 | 0.28–0.61 |
| MSNQ 13 | 0.42 | 0.26–0.57 |
| MSNQ 14 | 0.42 | 0.20–0.60 |
| MSNQ 15 | 0.35 | 0.17–0.55 |

MSNQ, Multiple Sclerosis Neuropsychological Screening Questionnaire; $\kappa_w$, weighted kappa; CI, confidence interval.

component analysis (eigenvalue MSNQ-P 1.35 and MSNQ-I 1.17), but when looking at the parallel analysis we could not retain this factor. Therefore, we can confirm the unidimensional structure of the MSNQ.

Cronbach's alpha analyses showed good internal consistency; 0.93 for MSNQ-P and 0.94 for MSNQ-I.

Spearman correlations were calculated between MSNQ-P and MSNQ-I, and all other variables. The results of these correlations are presented in Table 3 with variable means and standard deviations (SD).

All hypotheses were confirmed concerning the construct validity:

1. The correlation between MSNQ-P score and amount of impaired subtests on the BRB-N was 0.26, small positive.
2. The correlation between MSNQ-I score and amount of impaired subtests on the BRB-N was 0.39, almost moderate and positive, although just below the hypothesized value of 0.40.
3. Both patient depression (0.49) and anxiety (0.47) scales of the HADS showed a moderate and positive correlation with the MSNQ-P score.
4. The correlations between MSNQ-I and patient anxiety and depression score were 0.36 and 0.33 respectively, both small positive.
5. The correlation between MSNQ-P and impaired BRB-N subtests was 0.26, for MSNQ-I this correlation was 0.39. The difference in correlation between the MSNQ versions and amount of impaired BRB-N subtests was 0.13, more than 0.10.
6. The correlations between MSNQ-P and patient anxiety and depression were respectively 0.47 and 0.49, for the MSNQ-I these correlations were 0.33 and 0.36 respectively. The differences in correlations were both more than 0.10.

Interrater reliability between the patient and informant version on scale level was moderate. The ICC of MSNQ-P total score and MSNQ-I total score was 0.59 (95% CI: 0.46–0.69), which indicated that the MSNQ-P and MSNQ-I were not able to discriminate patients similarly. Interrater reliability on item level was also moderate. None of the kappa values reached the limit of 0.70. Weighted kappa values ($\kappa_w$) for item scores are shown in Table 4 with their confidence intervals (CI).

The Bland and Altman plot is displayed in Fig. 1. The mean difference was +1.3, meaning that the mean score of patients was 1.3 points higher than the mean score of informants (i.e. systematic error). The limits of agreement (i.e. random error) were between −17.7 and +20.4 (range −60 to +60).

### 4.3. Interpretability

Patients were classified in three levels of cognitive functioning, no impaired subtest score (no cognitive impairment), 1 or 2 impaired subtest scores (group at risk) and 3 or more impaired subtest scores (cognitively impaired). Of all MS patients in this sample, 39 patients (32%) had one or more abnormal scored subtests and 12 patients (10%) were classified as cognitively impaired because of 3 or more abnormal subtests (Table 5).

We could not define a cut-off score for the MSNQ-P because sensitivity (0.42) was found to be too low and not discriminative for cognitive impairment (3 or more impaired subtest scores). The optimal cut-off score in our data for the MSNQ-I was 21, with a sensitivity of 0.75 and a specificity of 0.71. The ROC curves are shown in Fig. 2. The AUC of MSNQ-I was 0.80 (95% CI 0.65–0.95), equal to the combined line of MSNQ-P and MSNQ-I. Also when including the group at risk (1 or 2 abnormal subtest scores) in the cognitively impaired

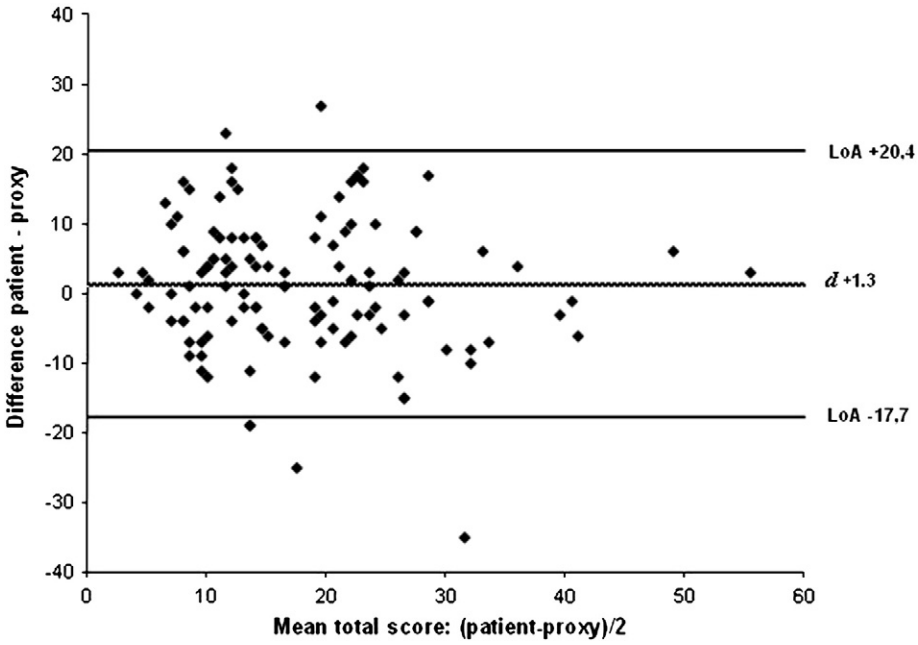


**Fig. 1.** Limits of Agreement MSNQ.

**Table 5**
Classification of BRB-N subtest results.

| BRB-N | N | |
|---|---|---|
| No impaired subtest score | 82 | 68% |
| 1 or 2 impaired subtest scores | 27 | 22% |
| 3 or more impaired subtest scores | 12 | 10% |

BRB-N, Brief Repeatable Battery for Neuropsychological Tests.

group, the MSNQ-I was more discriminative compared to the MSNQ-P and equal to the combination of MSNQ-I and MSNQ-P.

## 5. Discussion

In this study about the validation and interpretation of the Dutch version of the MSNQ we found a clear unidimensional factor structure for both versions. Internal consistency was good for both scales. Assessment of construct validity showed that all hypotheses based on previous studies [3–6] were confirmed. As expected, we found a higher correlation between the MSNQ-P and patient anxiety and depression (measured with the HADS) than between MSNQ-P and impaired cognitive subtests (measured with the BRB-N). In contrast, MSNQ-I correlates higher with the amount of impaired subtests compared to the correlation with anxiety or depression. Therefore, we recommend to use the MNSQ-I version.

The interrater reliability of the total score and the item scores between the patient and informant versions was moderate. The MSNQ-P is not able to identify cognitive disabled patients as good as the MSNQ-I does. Interrater agreement was poor. Although the systematic error between patient and informant version was low, the random error (i.e. Limits of Agreement) was large. Ninety-five percent of the differences between patient and informant were between −17.7 and +20.4. Since the range of the scale was 0 to 60, and therefore the difference between patient and informant could have a maximum variation between −60 and 60, the random error is huge. No significant differences were found between the group of informants participating in the hospital or at home (data not shown). However, it cannot be excluded that better results might have been obtained under more strictly controlled conditions.

In some earlier published studies about the MSNQ [3,5,6] researchers found low sensitivity for the MSNQ-P and no cut-off score could be defined. Also in this study we found a sensitivity that was too low to define a clear cut-off score for the patient version. The sensitivity and specificity of the informant version were better than the patient version. We could define a cut-off score of 21 for the Dutch version of the MSNQ-I, which was almost equal to results of earlier studies in North America [4]. Analyses showed that when combining MSNQ-I and MSNQ-P there was no added value of the patient version.

Some challenges were encountered when comparing our data with existing data because of the differences in the design and execution of previous studies. There were differences in measurement instruments and patient groups. Cognitive tests, used as the external standard, were not the same in all published studies about the sensitivity and determining cut-off scores for the MSNQ. Only the Argentinean study [6] used the BRB-N, but the classification of impaired patients was different compared to our study. In our study only a small number of patients were classified as cognitively impaired according to neuropsychological testing. Patients were participating in ongoing research projects and it cannot be excluded that this may have caused some selection bias towards less cognitively impaired patients. Furthermore it should be noted that the BRB-N as used in this study is missing a measure of higher executive functioning, which could have influenced the results. There were also differences in measuring depression in the published studies. Other researchers [3–6] used Beck's Depression Inventory — Fast screen (BDI-FS) [26], some in combination with the Center for Epidemiologic Studies Depression scale (CES-D) [3,4,27] while we used the HADS for measuring depression as well as anxiety. These issues may explain some of the observed differences in the results.

Future studies would benefit from selecting more MS patients with cognitive impairment, so that the relationship between cognitive testing and the MSNQ can be tested more specifically. In addition we would recommend to collect test–retest data. These studies should also address the origin of disagreement between patients and informants on the MSNQ. With the growing interest in self assessment scales understanding the background of disagreement is of increasing importance. It has been subject to research before and possible explaining factors are depression, anxiety and caregiver burden [28–30].

In summary, in this study we investigated the validity and interpretability of the MSNQ. The main outcome is that the MSNQ-I is more promising to screen for cognitive impairment in MS patients. The patient version has no added value, so when screening for cognitive impairment in MS the MSNQ informant version is preferred.

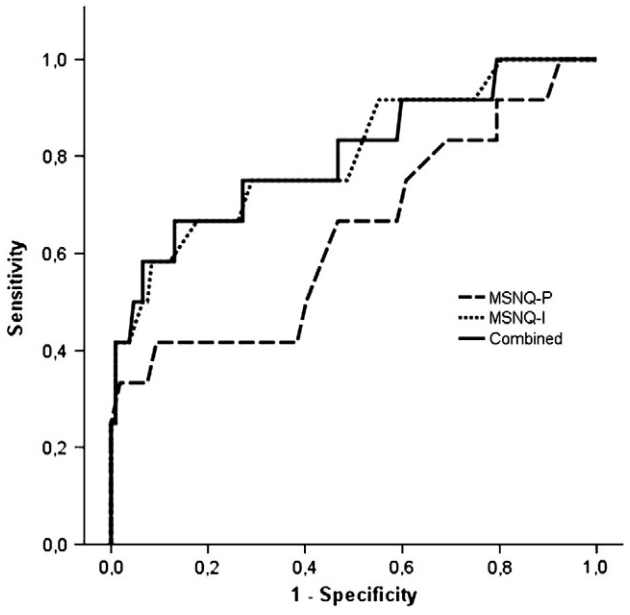


**Fig. 2.** ROC curves.

## Conflict of interest

I declare herewith that all authors report no financial or non-financial conflicts of interest.

## References

[1] Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. Neurology May 1991;41(5):685–91.
[2] Phillips LH, Saldias A, McCarrey A, Henry JD, Scott C, Summers F, et al. Attentional lapses, emotional regulation and quality of life in multiple sclerosis. Br J Clin Psychol Mar 2009;48(Pt 1):101–6.
[3] Benedict RH, Munschauer F, Linn R, Miller C, Murphy E, Foley F, et al. Screening for multiple sclerosis cognitive impairment using a self-administered 15-item questionnaire. Mult Scler Feb 2003;9(1):95–101.
[4] Benedict RH, Cox D, Thompson LL, Foley F, Weinstock-Guttman B, Munschauer F. Reliable screening for neuropsychological impairment in multiple sclerosis. Mult Scler Dec 2004;10(6):675–8.
[5] O'Brien A, Gaudino-Goering E, Shawaryn M, Komaroff E, Moore NB, DeLuca J. Relationship of the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) to functional, emotional, and neuropsychological outcomes. Arch Clin Neuropsychol Nov 2007;22(8):933–48.
[6] Vanotti S, Benedict RH, Acion L, Caceres F. Validation of the Multiple Sclerosis Neuropsychological Screening Questionnaire in Argentina. Mult Scler Feb 2009;15(2):244–50.
[7] Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clin Ther Sep 1996;18(5):979–92.

[8] Rao SM, collaboration with the Cognitive Function Study Group of the Nations Multiple Sclerosis Society. A manual for the Brief Repeatable Battery of Neuropsychological Tests in multiple sclerosis. Section of Neuropsychology, Medical College of Wisconsin, 1000 N. 92 street, Milwaukee, WI 53226; 1990.
[9] Rao SM. A manual for the brief, repeatable battery of neuropsychological tests in multiple sclerosis. Unpublished manuscript, Section of Neuropsychology, Department of Neurology, Medical College of Wisconsin, 1991.
[10] Bever Jr CT, Grattan L, Panitch HS, Johnson KP. The Brief Repeatable Battery of Neuropsychological Tests for Multiple Sclerosis: a preliminary serial study. Mult Scler Nov 1995;1(3):165–9.
[11] Strober L, Englert J, Munschauer F, Weinstock-Guttman B, Rao S, Benedict RH. Sensitivity of conventional memory tests in multiple sclerosis: comparing the Rao Brief Repeatable Neuropsychological Battery and the Minimal Assessment of Cognitive Function in MS. Mult Scler Sep 2009;15(9):1077–84.
[12] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand Jun 1983;67(6):361–70.
[13] Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol Med Mar 1997;27(2):363–70.
[14] Buschke H, Fuld PA. Evaluating storage, retention, and retrieval in disordered memory and learning. Neurology Nov 1974;24(11):1019–25.
[15] Barbizet J, Cany E. Clinical and psychometrical study of a patient with memory disturbances. Int J Neurol 1968;7(1):44–54.
[16] Smith A. Symbol digit modalities test: manual. Los Angeles: Western Psychological Services; 1982.
[17] Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. Percept Mot Skills Apr 1977;44(2):367–73.
[18] Boringa JB, Lazeron RH, Reuling IE, Ader HJ, Pfennings L, Lindeboom J, et al. The brief repeatable battery of neuropsychological tests: normative values allow application in multiple sclerosis clinical practice. Mult Scler Aug 2001;7(4):263–7.
[19] Sepulcre J, Vanotti S, Hernandez R, Sandoval G, Caceres F, Garcea O, et al. Cognitive impairment in patients with multiple sclerosis using the Brief Repeatable Battery-Neuropsychology test. Mult Scler Apr 2006;12(2):187–95.
[20] Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology Nov 1983;33(11):1444–52.
[21] Nunnally JC, Bernstein IH. Psychometric theory3rd ed. New York: McGraw-Hill; 1994.
[22] Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med Apr 15 1993;118(8):622–9.
[23] McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30–46.
[24] Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70:213–20.
[25] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet Feb 8 1986;1(8476):307–10.
[26] Beck AT, Steer RA, Brown GK. BDI-Fast Screen for medical patients: manual. San Antonio, TX: Psycholocial corporation; 2000.
[27] Radloff LS. The CES-D scale: a self report depression scale for research in the general population. Appl Psychol Meas 1977;1:385–401.
[28] Carone DA, Benedict RH, Munschauer III FE, Fishman I, Weinstock-Guttman B. Interpreting patient/informant discrepancies of reported cognitive symptoms in MS. J Int Neuropsychol Soc Sep 2005;11(5):574–83.
[29] Sonder J, Bosma L, van der Linden F, Knol D, Polman C, Uitdehaag B. Proxy measurements in multiple sclerosis: agreement on different patient-reported outcome scales. Mult Scler Feb 2012;18(2):196–201.
[30] van der Hiele K, Spliethoff-Kamminga NG, Ruimschotel RP, Middelkoop HA, Visser LH. The relationship between self-reported executive performance and psychological characteristics in multiple sclerosis. Eur J Neurol Apr 2012;19(4):562–9.