Validation and attempts of revision of the MDS-recommended tests for the screening of Parkinson’s disease dementia

V. Isella a, *, C. Mapelli a, C. Siri b, D. De Gaspari b, G. Pezzoli b, A. Antonini b, c, M. Poletti d, U. Bonuccelli d, M. Vista e, I.M. Appollonio a

a Neurology Section, S. Gerardo Hospital, University of Milan Bicocca, Monza, Italy
b Parkinson Institute, Istituti Clinici di Perfezionamento, Milan, Italy
c Department of Neurology, IRCCS San Camillo, University of Padua, Venice, Italy
d Department of Neuroscience, University of Pisa, Italy
e Neurology Section, Campo di Marte Hospital, Lucca, Italy

A R T I C L E   I N F O

Article history:
Received 28 March 2013
Received in revised form 2 September 2013
Accepted 6 September 2013

Keywords:
MDS guidelines
Mini Mental Parkinson
Parkinson’s disease dementia
Mini Mental State Examination

A B S T R A C T

The Movement Disorders Society (MDS) formulated diagnostic criteria and assessment guidelines for the screening of dementia in Parkinson’s disease (PD). We carried out a validation of the cognitive measures suggested in the screening algorithm (i.e. the Mini Mental State Examination – MMSE – total score, serial 7s subtraction, 3-word recall, pentagons copy, and one minute letter fluency) in 86 patients with PD. Thirty-six percent of participants were diagnosed with dementia using the MDS algorithm, but with the Dementia Rating Scale instead of the MMSE. The original MDS procedure misclassified 11 patients (12.8%) as false negatives and 3 (3.5%) as false positives, leading to 65% sensitivity and 95% specificity. The main reason for misdiagnoses was insensitivity of the MMSE total score. Three attempts were made to reach a better screening performance, which warrants high sensitivity more than high specificity: 1. exclusion of the MMSE total score as a diagnostic requirement; 2. determination of a better cut off through Receiver Operating Characteristic curve analysis; 3. replacement of the MMSE with the equally undemanding, but more PD-specific, Mini Mental Parkinson. The first two strategies generally yielded high sensitivity, but poor specificity. The best outcome was achieved using a Mini Mental Parkinson total score < 27 as cognitive criterion: sensitivity was 87% and negative predictive value was 90%; however, specificity was only 67%. Our findings seem to suggest that MDS practical guidelines are specific, but might benefit from the use of more PD-oriented tools than the MMSE in terms of sensitivity.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The interest of movement disorders clinicians specialists in the cognitive aspects of Parkinson’s disease (PD) is very vivid, especially since treatment perspectives are becoming more and more concrete [1]. Much research effort is devoted to improving accuracy in the diagnosis of PD-associated cognitive disturbances. In 2007 the Movement Disorders Society (MDS) established criteria for the diagnosis of PD Dementia (PDD) [2], and also proposed practical suggestions for their verification [3]. More precisely, two procedures were outlined. The so called Level I algorithm was intended for a bedside screening of PDD, and included the Mini Mental State Examination (MMSE) and few other short tasks like letter fluency or the clock drawing. Level II testing was instead suggested mainly for research purposes, and relied on an extensive neuropsychological battery.

These assessment guidelines were based on literature evidence and ‘common experience’, but the need for a formal evaluation of their validity was highlighted by the Authors themselves [3]. Recently, Martinez-Martin et al. [4] demonstrated that the MDS procedure was more sensitive than DSM IV criteria in identifying PDD. On the other hand, Barton et al. [5] and Di Battista et al. [6] conducted a validation study where the diagnostic ability of the Level I checklist was confronted with a comprehensive neurological, neuropsychological and functional assessment [5] and with the Level II procedure [6]. In both cases the MDS algorithm showed a specificity > 95%, but other clinimetric properties of the checklist were problematic. In Barton et al.’s study [5], sensitivity was only 46%. Most cases misclassified as non-demented did not fulfill the criterion of an abnormal total score at the MMSE, confirming previous literature that suggested that the scale is quite insensitive to
the typical profile of PD-associated cognitive impairment [7]. In Di Battista et al.’s study [6] sensitivity was higher but still unsatisfying, being 78%. Besides, the positive predictive value was 70%, indicating that the checklist tended to overestimate PDD.

Due to the low number of dementia cases in their samples, neither of the studies described above could validate alternative measures or procedures. In the present study we aimed to evaluate the validity of MDS recommended cognitive measures and, further, to test possible, more accurate variants of the original MDS algorithm. We adopted one of the validation strategies suggested by the MDS Task Force: “the criteria may be applied retrospectively to existing cohorts that have detailed investigations including neuropsychological assessments” [3]. We revised clinical, cognitive and functional data of a series of PD patients who had been administered the MMSE, the Dementia Rating Scale (DRS) [8] and the Mini Mental Parkinson (MMP) [9] as part of a validation study of the MMP. All the neurological and functional requirements of the MDS procedure were checklist, but using alternatively the MMSE, the DRS and the MMP as cognitive criterion. We were particularly interested in comparing the diagnostic performance of the MMSE with that of the MMP, which is equally brief and simple, but has a purportedly PD-oriented construct. Developed as a disease-specific screening tool, the MMP has been shown to be generally superior to the MMSE in PD patients in terms both of psychometric properties and validity [10,11].

2. Patients

MDS diagnostic criteria for PDD include a diagnosis of idiopathic PD according to the Queen Square Brain Bank criteria [12], and onset of the disease prior to the development of possible neuropsychological deficits [2]. These two elements were required for inclusion into the study, along with the availability of total and partial scores of Level I tests, DRS and MMP. Reasons for exclusion were delirium, brain injury, serious medical illness, and moderate to severe depression (score > 10 at the 15 item Geriatric Depression Scale — GDS).

The study was approved by local Ethics Committees. Patients were not paid for their participation and signed a written informed consent. Principles of the Declaration of Helsinki were followed.

3. Validation of the MDS Level I testing

3.1. Methods

Patients’ clinical records were revised retrospectively, taking into account neurological history, reports on impairment in non-motor activities of daily living, and total scores and subscores of DRS, MMSE, verbal fluency, and MMP. The clock drawing test and months backward are also part of Level I testing, as alternative measures of attention and executive functioning, but they were available only for a limited number of patients and were therefore disregarded. Functional impairment was investigated upon enquiry to patients and caregivers, who were asked about daily difficulties possibly related to cognitive dysfunction. Questions were focused on non-motor tasks such as managing finances, dealing with medications, remembering conversations and appointments. Informants were asked to provide concrete examples, in order to disentangle the contribution of motor disability.

The entire MDS checklist to our study population twice: using MMSE plus verbal fluency, and replacing them with the DRS. In the first case we used cut offs recommended by the Task Force: MMSE age- and education-adjusted score < 25, 7s subtraction < 3, 3-word recall < 2, pentagons copy = 0, one minute letter fluency < 9. As to the cut points for the DRS, guidelines recommend “reference to published norms” [3], which are not available for the Italian version of the scale. We therefore derived the cut offs in a pool of 76 healthy individuals, demographically comparable to our PD sample (men: 51%, mean age: 68.5 years ± 10.1, mean education: 8.9 years ± 3.4; mean MMSE score: 28.8 ± 1.1). Their average DRS score was 136.3 ± 5.5, which led to a cut point < 125. Cut offs for subs tests were as follows: Attention < 32, Initiative/Perseveration < 31, Construction < 4, Conceptualization < 26, Memory < 20.

Diagnosis of PDD according to published criteria [2] was made when patients obtained an abnormal score on the whole scale and on at least two of its subs tests, and when cognitive deficits were severe enough to impact daily activities.

3.2. Statistical analysis

Statistical analysis was carried out with PASW statistics, Release Version 18.0.0 (SPSS, Inc., 2009, Chicago, IL. www.spss.com). Chi-square analysis, student’s t test or univariate analysis of variance with age and education as covariates (ANCOVA) were used to compare means of discrete and continuous variables between PDD and non-demented patients, with a two tailed standard significance level set at p < 0.05.

Sensitivity, specificity, positive and negative predictive values (PPV, NPV) and total accuracy were calculated for the MDS recommended tests. A sensitivity value and an NPV > 0.80 were considered as an optimal screening performance, but specificity, PPV and accuracy were also taken into account to value the validity of the procedure.

3.3. Results

Ninety patients fulfilled criteria for inclusion, but four had to be excluded due to incomplete data concerning functional impairment. The final study sample was therefore composed by 86 patients. Thirty-one of them (36.0%) met MDS criteria for probable (n = 24) or possible (n = 7) PDD using the algorithm including the DRS. Their socio-demographic, neurological and neuropsychological features, compared with the non-demented group, are shown in Table 1. They were significantly older, less educated and more neurologically impaired than non-demented patients. After correcting for the effect of age and education, their MMSE and DRS scores were significantly worse than for the other group.

Using MMSE and letter fluency, 72/86 patients (83.7%) were correctly classified as PDD (n = 20) or non-demented (n = 52). Misclassifications were 11 false negative (12.8%) and three false positive (3.5%) cases. The latter patients all scored 25 at the MMSE, while their DRS scores were 126, 128 and 136. The false negative cases had an average MMSE score of 27.2 ± 0.7 (range: 26.0–28.3); for all of them was the total score above the recommended cut off. Eight obtained more than two abnormal subs scores included in the MDS checklist. Their average DRS total score was 113.7 ± 6.6 (range: 103–124). The scatterplot in Fig. 1 shows the wide overlap of their DRS scores with that of the 20 cases correctly classified as PDD.

Table 1

<table>
<thead>
<tr>
<th>Algorithm for the screening of Parkinson’s disease dementia: Movement Disorders Society original checklist and revised version.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Diagnosis of PD based on Queen’s Square Brain Bank criteria</td>
</tr>
<tr>
<td>2. Parkinson’s disease developed prior to the onset of dementia</td>
</tr>
<tr>
<td>3. Cognitive deficits decrease independence in (non-motor) activities</td>
</tr>
<tr>
<td>4. MDS version: MMSE total score &lt; 25 and at least two abnormal subscores:</td>
</tr>
<tr>
<td>- 7s subtraction (or months backward)</td>
</tr>
<tr>
<td>- Letter fluency (or clock drawing)</td>
</tr>
<tr>
<td>- 3-Word recall</td>
</tr>
<tr>
<td>- Pentagons copy</td>
</tr>
</tbody>
</table>

Revised version: MMP total score ≤ 27

PD — Parkinson’s disease, MDS — Movement Disorders Society, MMSE — Mini Mental State Examination, MMP — Mini Mental Parkinson.
Overall, the MDS recommended tests showed 65% sensitivity, 95% specificity, 87% PPV, 83% NPV and 84% accuracy.

4. Tentative revised version of the MDS Level I testing

4.1. Methods

The validation of cognitive measures included in the MDS screening procedure, using the DRS, yielded very high specificity, satisfying NPV, but poor sensitivity. The criterion of involvement of at least two cognitive domains was fulfilled in more than 70% of false negative cases, while in none of them the MMSE total score was below the recommended cut off.

Based on this finding, we explored three possible strategies aimed at achieving the optimal screening combination of sensitivity and NPV, without sacrificing specificity. First, we excluded an abnormal MMSE total score as requirement for the diagnosis of PDD, focusing on subtests only. Then we searched for a better cut point for MMSE total score through the construction of a Receiver Operating Characteristic (ROC) curve, which was used to classify patients in isolation, as well as in combination with subtests. Finally, we replaced MMSE and verbal fluency with the MMP. The optimal cut off for the MMP was determined through ROC curve analysis, and applied by itself as well as combined with scores from three subtests, chosen in analogy to those recommended by the MDS: serial sevens, delayed visual recall, and semantic-phonological fluency. Cut offs for each of the four subtests were again determined with ROC curves (data not shown): 7s subtraction $< 4$, visual recall $< 3$, fluency $< 2$.

4.2. Results

Validity values obtained by the original MDS procedure and by variants explored in our study are shown in Table 2.

Excluding the requirement of an abnormal total MMSE score for a diagnosis of PDD, increased sensitivity to 84%, did not significantly modify NPV, but greatly decreased specificity, PPV and total accuracy.

The ROC curves obtained for MMSE and MMP are shown in Fig. 2. Areas under the curves were overlapping: 0.80 (95% CI, 0.70–0.90) for the MMSE and 0.85 (95% CI, 0.78–0.94) for the MMP ($z = −0.73$, n.s.). The cut off with the best combination of high sensitivity and adequate specificity was $≤ 27$ for both tests. When the MDS original checklist was applied using this new MMSE cut point rather than the recommended one ($≤ 25$), sensitivity increased from 65% to 77%, PPV was still very high, while specificity, PPV and accuracy decreased but were still acceptable. Misclassifications were 15 (17.4%) false positive and 7 (8.1%) false negative cases. In the latter group, four patients had an MMSE total score above the cut off, and five were impaired in only one cognitive domain. Disregarding the requirement of two or more abnormal subscores further increased sensitivity and NPV (both were 87%), but greatly reduced specificity, PPV and accuracy (Table 2).

Replacing the MDS measures and cut offs with the MMP total score and subscores minimally improved sensitivity (71%); NPV was unchanged; specificity, PPV and accuracy decreased. Misclassifications were 12 (14.0%) false positive and 9 (10.5%) false negative cases. In the latter group, four patients had an MMP total score above the cut off, and eight were impaired in only one cognitive domain. Excluding impairment at subtests yielded 87% sensitivity and 90% NPV, and did not exceedingly reduce specificity, PPV and total accuracy (Table 3).

5. General discussion

In the present study we contrasted the screening ability for PDD of cognitive measures included in the MDS guidelines (MMSE and letter fluency) with the diagnostic classification obtained replacing them with the DRS. Our findings confirmed previous evidence [5,6] suggesting that MMSE plus verbal fluency have high specificity (95%), but insufficient sensitivity (65%). Barton et al. [5] found an even poorer sensitivity value (46%), probably due to differences in the study population and design. First of all, their sample had a much lower prevalence of demented patients (16.5% versus 36.0%). Secondly, they used a very extensive neuropsychological battery as

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Socio-demographic, neurological and neuropsychological features of patients with (PDD) and without (PD) dementia classified according to the procedure including the Dementia Rating Scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PDD n. 31</td>
</tr>
<tr>
<td>Sex (% of men)</td>
<td>58%</td>
</tr>
<tr>
<td>Age</td>
<td>73.7 ± 7.1*</td>
</tr>
<tr>
<td>Education</td>
<td>6.0 ± 3.1*</td>
</tr>
<tr>
<td>Disease duration (years)</td>
<td>7.7 ± 5.1</td>
</tr>
<tr>
<td>Unified Parkinson’s Disease Rating Scale motor score</td>
<td>29.2 ± 12.6*</td>
</tr>
<tr>
<td>Hoehn &amp; Yahr (% of patients $≥ 2.5$)</td>
<td>68%*</td>
</tr>
<tr>
<td>Mini Mental State Examination total score</td>
<td>22.7 ± 3.3*</td>
</tr>
<tr>
<td>Dementia Rating Scale total score</td>
<td>110.1 ± 9.8*</td>
</tr>
<tr>
<td>Geriatric Depression Scale − 15 item</td>
<td>3.6 ± 2.7</td>
</tr>
</tbody>
</table>

Mean ± standard deviation. *p < 0.05 versus PD.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Validity values obtained by the Movement Disorders Society procedure for the diagnosis of Parkinson’s Disease Dementia, and by variants tested in the study (gold standard: Dementia Rating Scale).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut off</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>MMSE</td>
<td>Total score + subscores*</td>
</tr>
<tr>
<td></td>
<td>Subscores only</td>
</tr>
<tr>
<td></td>
<td>Total score + subscores</td>
</tr>
<tr>
<td>MMP</td>
<td>Total score + subscores</td>
</tr>
<tr>
<td></td>
<td>Total score only</td>
</tr>
<tr>
<td></td>
<td>Total score only</td>
</tr>
</tbody>
</table>

PPV = positive predictive value, NPV = negative predictive value.

*a Movement Disorders Society procedure.
a gold standard, which was likely more accurate than the DRS. Finally, they validated the whole MDS checklist, including mood and functional criteria, while we focused on cognitive tools.

Besides being of limited diagnostic use to the clinician, a screening procedure with such a poor sensitivity is also problematic whenever subsequent Level II assessment [3] is planned. Confusing incongruencies in the outcome of the two diagnostic levels might emerge, as patients initially categorized as non-demented might actually obtain an abnormal performance at the DRS, which is part of Level II battery. Dubois et al. [3] did envisage the possibility of revision of Level I diagnostic conclusions, through Level II assessment, for cases who remained equivocal after the screening, but our misclassified patients could not be considered equivocal: their diagnostic categorization was not uncertain at the end of Level I testing.

It might be argued that in our study the MDS procedure might have been penalized because we did not consider clock drawing and months backward tasks (which were recommended as alternative to letter fluency and 7s subtraction, respectively), but we doubt this might be the case. Verbal fluency was failed by most patients correctly classified as PDD, and false negative cases who scored normally at serial subtraction would have probably performed well at months repetition as well. Hence the inclusion of these two tests would have hardly improved the guidelines’ sensitivity. In line with previous evidence arguing against the use of MMSE for the cognitive assessment of PD patients [5,6,13–15], we identified the insensitivity of the MMSE as the main reason for missed PDD cases. Fig. 1 clearly shows that false negative and true positive patients were equal in dementia severity as measured by the DRS, yet fell on either side of MMSE recommended cut point. As tentative solutions to this problem, we excluded an abnormal MMSE total score as a criterion for diagnosis, and searched for a better MMSE cut off through ROC curve analysis. The latter attempt yielded the best sensitivity–specificity trade off, but sensitivity was still unsatisfying. The best screening performance was achieved when MDS recommended tests were replaced by the MMP, whose total score yielded 87% sensitivity and 90% NPV. However, specificity was only 67%.

These values were attained when the MMP total score was taken into account on its own, disregarding the individual performance at serial subtractions, visuo-spatial recall and letter-category fluency. This suggests that other MMP subtests might be more sensitive to cognitive impairment associated with PD. Further analyses might identify more valid combinations of items. Still in terms of future directions, upcoming investigations should also confront the MMSE with other PD-oriented screening tools (e.g. the MOCA [16–18] or the PDD-Short Screen [19]). The choice of the MMP was driven by the similar feasibility and briefness of this scale with respect to the MMSE, and by previous studies supporting its validity in PD [10,11]. The MMP has in fact demonstrated better sensitivity, but tends to produce an excessive number of false positive cases. Finally, it would be of great interest if future investigations dealt with PD patients at an earlier disease stage (mean illness duration was approximately seven years in our sample), and followed them longitudinally to determine the value of screening checklists in early PDD diagnosis.

One potential caveat of the present, retrospective, study is that we cannot be completely confident that collection of functional information by clinicians was systematic, nor homogeneous, in terms of number and type of daily activities investigated. As a consequence, some case of PDD might have gone undetected. It has to be said, though, that the sensitivity of a test is expected to increase with higher prevalence rates of the disease. This means that an underestimation of dementia prevalence (due to inaccurate delineation of functional impairment relating to cognition) would only reinforce our negative conclusions about the MMSE. This comment also applies to possible limitations deriving from the use of the DRS as a kind of gold standard for the diagnosis of PDD. The DRS can be seen as an extended screening measure, likely less sensitive than a neuropsychological battery as comprehensive as the one used in previous studies [5,6]. The actual effect of this lack of diagnostic power, though, would have been again an underestimation of PDD cases, resulting in a possible strengthening of the final outcome of our study. A more definite limitation of the present results concerns their generalizability. Our population showed a high prevalence of demented patients, which might be unusual in movement disorders clinics, so a different outcome might be expected in such a setting. Early onset PD was underrepresented, and participants had been preselected not to have significant depression, which is extremely frequent in PD patients with cognitive impairment; our findings might not hold true for those categories of PD patients.

Since the publication of MDS practical guidelines for the screening of PDD [3], few research groups formally assessed their validity [4–6]. Thanks to the present study, movement disorders and dementia specialists may have a more precise idea of the level of screening accuracy they can achieve when applying the MDS algorithm to their suspect PDD cases. We also empirically explored three possible alternative procedures, obtaining a relatively more satisfying procedure by replacing the MMSE with the MMP, but this finding needs to be confirmed in future, prospective, investigations.

Full financial disclosures of all authors for the past year

Prof. U. Bonuccelli:
Advisory Boards: GlaxoSmithKline, Lundbeck, Novartis, UCB.
Grants: Regione Toscana Health Authority.

Acknowledgments

None.
References


