Introducing Multiple Screener: an unsupervised digital screening tool for cognitive deficits in MS

Dongen, L. van; Westerik, B.; Hiele, K. van der; Visser, L.H.; Schoonheim, M.M.; Douw, L.; ... ; Hulst, H.E.

Multiple Sclerosis and Related Disorders 38 (2020) 101479

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

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

Clinical trial

Introducing Multiple Screener: An unsupervised digital screening tool for cognitive deficits in MS

L. van Dongen[a], B. Westerik[a], K. van der Hiele[b], L.H. Visser[c,d], M.M. Schoonheim[a], L. Douw[a], J.W.R. Twisk[e], B.A. de Jong[f], J.J.G. Geurts[a], H.E. Hulst[a,*]

[a] Department of Anatomy & Neurosciences, Amsterdam Neuroscience, MS Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[b] Department of Psychology, Health, Medical and Neuropsychology Unit, Leiden University, Leiden, the Netherlands
[c] Department of Neurology, Elisabeth-Tweesteden Hospital, Tilburg, the Netherlands
[d] Department of Care Ethics, University of Humanistic Studies, Utrecht, the Netherlands
[e] Department of Epidemiology and Biostatistics, Amsterdam Public Health, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[f] Department of Neurology, Amsterdam Neuroscience, MS Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

ARTICLE INFO

Keywords:
Multiple Sclerosis
Cognitive impairment
Digital Screening tool
Norm scores

ABSTRACT

Background: Cognitive deficits affect up to 70% of all patients with Multiple Sclerosis (MS) and have a significant impact on quality of life. Cognitive assessments need to be performed by a neuropsychologist and are often time-consuming, hampering timely identification and adequate monitoring of cognitive decline in MS.
Objective: To develop a time-efficient, unsupervised, digital tool to screen for cognitive deficits in MS.
Methods: A digital (adjusted) version of the Brief International Cognitive Assessment for MS, including the Symbol Digit Modalities Test (SDMT, information processing speed), the California Verbal Learning Test (CVLT-II, verbal memory) and the Spatial Recall Test (SPART, visuospatial memory) was developed: Multiple Screener (intellectual property of Sanofi Genzyme).

Firstly, the clarity and feasibility of the tool was confirmed by 16 patients with MS (mean age 50.9 years (SD 9.4, range 37–68). Next, in 60 healthy controls (HCs, mean age 44.5 years (SD 14.0, range 18–67), intraclass correlation coefficients (ICC) were calculated to describe how strongly the digital version resembled the paper and pencil-based assessment. Finally, 236 HCs (mean age 42.8 years (SD 12.8, range 18–69) were included to obtain norm scores for each test.
Results: ICCs between digital and paper and pencil-based assessment were excellent to good (SDMT (ICC 0.79, confidence interval (CI) 0.67–0.87); CVLT-II (ICC 0.77, CI 0.64–0.85); SPART (ICC 0.61, CI 0.42–0.75)). For each test, a regression-based correction for the effect of age was applied on the raw scores before converting them to norm *Z*-scores. Additionally, the SDMT scores needed correction for education and the CVLT-II for education and sex (subgroups were created).
Conclusions: Performance on an adjusted, digital version of the BICAMS correlates highly with the standard paper-and-pencil based test scores in HCs. Multiple Screener is an unsupervised, digital tool, with available norm scores, ultimately allowing for easy monitoring of cognitive decline in patients with MS.

1. Introduction

Cognitive deficits affect up to 70% of all patients with MS and have a significant impact on work participation and quality of life (Rao et al., 1991; Chiaravalloti and DeLuca, 2008; Van der Hiele et al., 2014). Cognitive domains that are most frequently affected are information processing speed and memory (Rao et al., 1991; Chiaravalloti and DeLuca, 2008).

Several cognitive test batteries have been validated to detect cognitive decline in patients with MS (Langdon et al., 2012; Benedict et al., 2012; Benedict et al., 2006), of which the BICAMS (Brief International Cognitive Assessment for Multiple Sclerosis) is the most well-known cognitive screener (Benedict et al., 2012). For some tests included in these batteries (i.e. symbol digit modalities test (SDMT) and the California verbal learning and memory test (CVLT-II)) it is known that they are able to detect clinical meaningful changes (Morrow et al., 2010;

* Corresponding author at: Human and Life Sciences Building (O|2), 13th floor, Dept. of Anatomy and Neurosciences, De Boelelaan 1108, 1081 HZ Amsterdam, Netherlands.
E-mail address: he1.hulst@amsterdamumc.nl (H.E. Hulst).

https://doi.org/10.1016/j.msard.2019.101479
Received 3 July 2019; Received in revised form 6 August 2019; Accepted 25 October 2019

Strober et al., 2018; Benedict et al., 2017). Still baseline and follow-up measurements of cognitive functioning are often lacking in clinical practice because objective cognitive assessment is time consuming and requires specialized personnel. This is definitely the case in the Netherlands, but also applies to other countries as was demonstrated in a recent investigation across the UK (Klein et al., 2018). This is in conflict with international recommendations for cognitive screening and cognitive management in MS that recommend baseline measurement and annual follow-up on cognitive functioning in stable patients (at least with SDMT and ideally using BICAMS) (Kalb et al., 2018). The lack of a baseline assessment (i.e. internal reference) hampers the detection of subtle deleterious changes in cognition, especially when patients perform above norm scores (Sumowski et al., 2018). Therefore, the aim of the current study was to develop a time-efficient, unsupervised, digital tool to measure cognitive deficits in MS. To enable implementation of the tool in a clinical setting, norm scores for each test were collected in a large sample of healthy controls to allow for a proper interpretation of the patients' cognitive test scores (e.g. indicate normal functioning or the presence of mild, moderate or severe impairment).

2. Methods

2.1. Digital cognitive assessment

The Multiple Screener tool (intellectual property of Sanofi Genzyme) was written in X-code, a general purpose high-level object-oriented programming language used in Apple's Mac OS X and iOS operating systems.

It is important to take psychological factors into account when screening for cognitive deficits. Therefore, the Multiple Screener tool consists of two segments: a) online questionnaires that measure well-being (Hospital Anxiety and Depression Scale for anxiety and depression; MS Neuropsychological Screening Questionnaire for subjective complaints; MS Impact Scale for fatigue) and b) three neuropsychological tests that measure cognition. As the online questionnaires are well-known, standardized, and validated, the focus of the current research is on the neuropsychological tests.

The digital tool included the Dutch-versions of three neuropsychological tests that measure the cognitive domains that are most commonly affected in MS (Chiaravalloti and DeLuca, 2008) (inspired by the BICAMS test battery): information processing speed (Symbol Digit Modalities Test (SDMT)) (Parmenter et al., 2007; Smith, 1991), verbal learning and memory (Dutch version of the California Verbal Learning Test – second edition (CVLT-II)) (Woods et al., 2006), and spatial memory (Spatial Recall Test (SPART)) (Rao, 1990). The SPART was included instead of the Brief Visuospatial Memory Test-Revised (BVMT-R) (Benedict, 1997), which is originally part of BICAMS, since a digital version of the BVMT-R was too difficult to score automatically.

The digital versions of the three tests were as similar as possible to the original, paper and pencil-based versions and permission from the manufacturers was obtained to digitalize the cognitive tests. Nevertheless, adjusting the paper and pencil-based versions of the three tests to digital alternatives resulted in some (minor) changes: writing became typing, and verbal explained test instructions were replaced by a textual (digital) explanation preceding each test. For a more detailed description of the digital tests and the visualization of the Multiple Screener tool, see supplementary material 1 and Fig. 1.

This study consisted of three parts:

- Part 1: *Clarity and Feasibility*. Ten patients with MS with varying disease duration, disease type, and level of cognitive functioning evaluated the Multiple Screener tool by filling out a questionnaire regarding the understandability, readability, speed, use of touchscreen, and audio parts of each test. Based on the feedback of these ten patients, adjustments were made to the Multiple Screener tool. The final version was re-evaluated (same questionnaire) by 6 tool-naïve patients with MS confirming the clarity and feasibility of the tool.
- Part 2: *Comparison between paper and digital assessment*. To ensure similarity between the original neuropsychological tests (paper and pencil-based) and the digital version, the tests included in the Multiple Screener tool (version A) were compared to the paper and pencil-based assessment (version B) in 60 healthy controls (HCs).
- Part 3: *Normative data*. For implementation in the clinical setting, norms were obtained for each test based on data from 236 HCs.

2.2. Study population

This study was approved by the local medical ethical board, the Declaration of Helsinki was followed, and all participants gave written informed consent prior to participation. Educational level (Dutch equivalent) was divided into two categories: low-medium (vocational training or lower, first three years of higher general secondary education, and pre-university education) and high (bachelor's degree, master's degree, PhD, or advanced professional degree). All participants were native Dutch speakers.

For part 1, 10 patients with MS (mean age 51.7 years (SD 7.7, range 38–64); men: $N = 1$; educational level high: $N = 4$; Relapsing Remitting MS (RRMS): $N = 7$; Secondary Progressive MS (SPMS): $N = 2$; Unknown MS-type: $N = 1$) evaluated the clarity and feasibility of the first version of the tool. Six additional patients with RRMS (mean age 49.5 years (SD 12.4, range 37–68); men: $N = 1$; educational level high: $N = 5$) re-evaluated the final version of the tool. The patients were recruited through advertisements and previous participation in research. Inclusion criteria were 18 years and older, a confirmed MS diagnosis of at least three months (by a neurologist, McDonald criteria (Polman et al., 2011), and sufficient visual and motor capabilities to work on an iPad. Patients with neurological or psychiatric diseases or with a history or current substance abuse were excluded.

For part 2, 60 HCs (mean age 44.5 years (SD 14.0, range 18–67); men: $N = 20$; educational level high: $N = 38$) were assessed with the digital and the paper and pencil-based version of the tests. The order of the digital and paper and pencil-based versions was counterbalanced. Data of the participants that started with the digital version were included to obtain norm scores (part 3), while data from the 30 HCs that started with the paper and pencil-based assessment before performing the digital tests were not included in determining norms because of a possible practice effect.

For part 3, 236 HCs (mean age 42.8 years (SD 12.8, range 18–69); men: $N = 93$; educational level high: $N = 149$) were included to obtain norms. All HCs (part 2 and part 3) were recruited through advertising in the local newspapers (printed and online), online advertising on Facebook, and the researchers' network of family and friends. Inclusion criteria were an age of 18 years and older and sufficient visual and motor capabilities to work on an iPad. People with neurological or psychiatric diseases or with a history or current substance abuse were excluded.

2.3. Statistical analysis

SPSS version 22.0 (SPSS/IBM 2013) was used for statistical analyses. Data were checked for normality using the Shapiro–Wilk test and by visual inspection of histograms. If non-normally distributed, the data were transformed to obtain normality. A p-value of < 0.05 was considered statistically significant. Part 1 consisted of qualitative evaluation of the clarity and feasibility of the Multiple Screener tool including no statistics. For part 2, first, intraclass correlation coefficients (ICC, two-way mixed effects model, relative agreement, single rater) were calculated to describe how well the digital and paper and pencil-based forms of assessment resembled each other with regard to the construct tested. ICC scores less than 0.40 indicated poor reliability, scores between 0.40 and 0.59 indicated fair reliability, scores between 0.60 and 0.74 indicated good reliability, and scores between 0.75 and 1.00

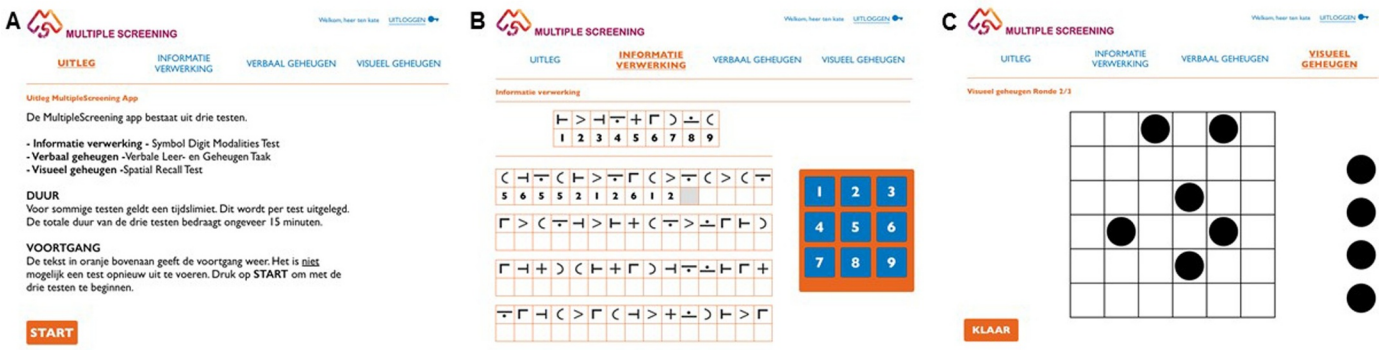

Fig. 1. Visualization of the Multiple Screening tool. The Dutch version of the California Verbal Learning Test–second edition (CVLT-II) is not depicted as this test has an auditory format. **A.** After login, the Multiple Screening tool starts with a general introduction (in Dutch) explaining that the app consists of three tests (information processing speed, verbal memory, and visual memory) of 15 min in total and that some tests have a time limit. The bar in the upper part of the displayed at the top of the screen. **B.** The Digit Modalities Test (SDMT). A reference key with combinations of geometric figures and digits (one to nine) is displayed at the top of the screen. The participant has 90 s to type in the correct digits underneath the geometric figures. The lies with symbols move up according to the speed of the participant. **C.** The Spatial Recall Test (SPART). The 36-square grid with 10 black checkers is displayed three times for ten seconds. After each time, an empty grid is displayed with ten black checkers next to it. Participants have to swipe the black checkers to the correct places in the empty grid to match what they observed.

Table 1
ICCs with 95% confidence intervals (CI).

Test	ICC	CI
SDMT[a]	0.79	0.67–0.87
CVLT-II	0.77	0.64–0.85
SPART	0.61	0.42–0.75

Two-way mixed effects model, relative agreement, single rater.
[a] Written version for paper and pencil-based version.

indicated excellent reliability (Cicchetti, 1994). Second, Bland-Altman plots were made to depict the agreement between digital and paper and pencil-based versions of the different tests. For part 3, a multiple linear regression was used to establish the effect of age, sex, and educational level on the test scores. A regression-based correction for the effect of age was applied on the raw scores before converting them to norm *Z*-scores (Oosterhuis et al., 2016; Parmenter et al., 2010). If sex or educational level significantly predicted test performance, participants were additionally grouped based on sex and/or educational level. For each test, group means and standard deviations were calculated, resulting in the norm scores. If subgroups were created, norms were calculated per subgroup (men/women; high/low-education).

3. Results

3.1. Part 1: Clarity and feasibility

Based on the evaluations of the first 10 patients with MS, minor technical errors were identified (i.e. too small font, minor spelling errors). Textual explanations of all three tests needed clarification for patients to perform the tests independently. For instance, based on the text, it was unclear to most patients where to start the SDMT after the example items had been shown. This was solved by adding a flashing square around the first item. Also, it was unclear that the patients needed to click 'next' after entering a word during the CVLT-II, and how to move the checkers during the SPART. These issues were solved by using a larger font, and including this information in the textual explanation of the tests. Following these adjustments, the clarity and feasibility (understandability, readability, speed, use of touchscreen, and audio parts) of the tests were re-evaluated by 6 tool-naïve patients with MS, who confirmed that the digital tool was self-explanatory. All 236 HCs were able to complete the digital version of the tests without the help of a test leader.

3.2. Part 2: Comparison between paper and digital assessment

Results of the ICCs comparing performance on the digital version and paper and pencil-based version of the three tests ($N = 60$ HCs) are presented in Table 1, and in the Bland Altman plots (Fig. 2). Good (SPART) to excellent (SDMT and CVLT-II) consistency was found between the paper and pencil-based and digital versions of the tests.

3.3. Part 3: Normative data

Norm scores are presented in Table 2.

3.3.1. SDMT

From the 236 HCs that were recruited, one person was excluded because of an unreliable test score (11 correct items versus 25 incorrect items). Data from one person was missing due to technical problems, resulting in a total of 234 HCs (mean age 42.7 years (SD 12.8, range 18–69)) that were included in the reference group of the SDMT. The results of the regression analysis indicated that age, sex, and education level explained 43.5% of the variance in SDMT score. Age significantly predicted the SDMT score (regression coefficient (β) = -0.40, $p < 0.001$), as did educational level ($\beta = 3.09$, $p = 0.002$), but not sex ($\beta = 0.70$, $p = 0.451$).

3.3.2. CVLT-II

From the 236 HCs that were recruited, 225 HCs (mean age 42.6 years (SD 12.7, range 18–69)) were included to obtain norms for the Dutch equivalent of the CVLT-II. The score of one person was extremely low (31 out of 80) according to the educational level (master's degree) and was therefore considered a probable erroneous score and excluded from data-analysis. Four participants were excluded because of a highly unlikely drop in performance between two consecutive trials (one person dropped 5 points, one 6, one 11, and one 14). In $N = 6$, no results were recorded due to technical problems. Because CVLT-II scores were non-normally distributed (negatively skewed), each individual's score was subtracted from the highest score plus one (79), and a square root transformation was done, resulting in a normal distribution of scores. The multiple linear regression analysis indicated that age, sex, and education explained 26.7% of the variance in CVLT-II score. Age significantly predicted CVLT-II score ($\beta = 0.02$, $p < 0.001$), as did educational level and sex (both $\beta = -0.59$, $p < 0.001$).

3.3.3. SPART

For the SPART, 236 HCs (mean age 42.7 years (SD 12.8, range

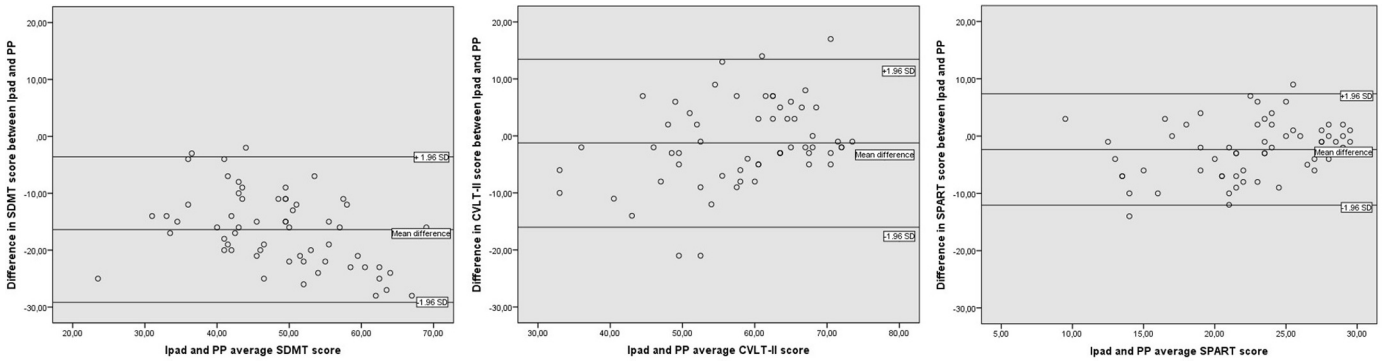

Fig. 2. Bland Altman plots.

18–69)) were included. Scores of one person were missing due to technical problems, resulting in a total of 235 HCs. The results of the regression analysis indicated that age, sex, and educational level explained 20.6% of the variance in visuospatial memory performance. Age was significantly related to SPART score ($\beta = -0.18, p < 0.001$), while educational level ($\beta = -0.19$, $p = 0.773$) and sex ($\beta = 0.54$, $p = 0.427$) were not.

3.4. From norm scores to interpretation

The norm scores displayed in Table 2 could be used to interpret the current cognitive status of a patient. Based on scatter plots (supplementary Fig. 1), we assumed that the relation between age and test score is linear in each test. Therefore, for each test, a regression-based correction for the effect of age was applied before converting the raw scores into norm *Z*-scores:

- SDMT raw score participant + (age participant * 0.40);
- CLVT square root (79 – raw score participant) + (age participant * −0.02);
- SPART raw score participant + (age participant * 0.18).

These formulas will be implemented in the backend of the tool, such that the neurologist can compare the age-corrected score of a patient directly to the age-corrected norm scores in Table 2. In order to interpret the performance on the SDMT and the CVLT-II, the neurologist should use the norm scores for the subgroups (i.e. educational level and sex) displayed in Table 2 (see supplementary material for an example on how to use the norm scores).

4. Discussion

The aim of the current study was three-fold: (1) to confirm the clarity and feasibility of the digital versions of three neuropsychological tests, resulting in the Multiple Screener tool, (2) to assess how well the digital versions of the neuropsychological tests resembled the original paper and pencil-based assessments, and (3) to obtain norms for each test enabling interpretation in a clinical setting.

This study demonstrated that patients with MS positively evaluated the clarity and feasibility of the digital, unsupervised, cognitive test battery including tests of verbal and visuospatial memory and information processing speed. In healthy controls, the three tests included in the Multiple Screener tool correlated highly with their corresponding paper and pencil-based versions, indicating that the tests measure the same cognitive abilities. The largest ICC was found for the SDMT (0.79), which is identical to the findings of a previous study that compared the paper and pencil-based SDMT to a digital SDMT assessment (Pearson's $r = 0.78$) (Rao et al., 2017). For the CVLT-II and the SPART, no studies on digital assessment are currently available, but the ICC showed excellent similarity with the paper and pencil-based versions of the CVLT-II especially taking into account that we used parallel versions. We do realize that for the SPART, the ICC was adequate. This is most likely a reflection of the task itself that can rely on a variety of strategies used by the participants resulting in heterogeneity between paper-and-pencil

Table 2
Norm scores per test.

Test		Subgroups		N	Mean	SD	'normal'	'possibly mildly impaired'	'probably impaired'
SDMT	Uncorrected	Total group[c]		234	40.14	8.94			
	Corrected[b]	Total group[c]		234	57.39	6.91			
		Education low-medium[d]		87	55.36	6.53	> 48.-83	42.30–48.83	< 42.30
		Education high[d]		147	58.59	6.84	> 51.-75	44.85–51.75	< 44.85
CVLT-II[a]	Uncorrected	Total group[c]		225	4.51	1.14			
	Corrected[b]	Total group[c]		225	3.62	1.07			
		Education low-medium	Men[d]	44	4.47	0.95	< 5.42	6.37–5.42	> 6.37
			Women[d]	39	3.62	0.93	< 4.55	5.48–4.55	> 5.48
		Education high	Men[d]	44	3.64	1.22	< 4.86	6.08–4.86	> 6.08
			Women[d]	98	3.22	0.87	< 4.09	4.96–4.09	> 4.96
SPART	Uncorrected	Total group[c]		235	20.86	5.25			
	Corrected[b]	Total group[c]		235	28.55	4.68	> 23.-87	19.19–23.87	< 19.19

[a] CVLT-II scores are square root transformed.
[b] Regression-based correction for the effect of age. SDMT: raw score participant + (age participant * 0.404); CLVT: square root (79 – raw score participant) + (age participant * −0.021); SPART: raw score participant + (age participant * 0.180).
[c] Displays mean and standard deviation from the whole group.
[d] Displays mean and standard deviation from all participants falling in this range.

and digital performance.

To ensure proper implementation of the Multiple Screener tool in a neurology department as screening tool, norms are essential to interpret the patients' test scores (e.g. indicate normal functioning or the presence of mild, moderate or severe impairment). Demographic variables such as age, sex, and education are known confounders of cognitive performance (van der Elst et al., 2006) and, as a result, scores need to be adjusted accordingly. Based on a representative sample of subjects in the Netherlands ($N = 236$), we were able to demonstrate that raw test scores needed correction for age in all three tests (SDMT, CVLT-II, SPART), as expected by the well-known and widely studied effect of healthy aging on cognition (Salthouse, 2010). In line with previous literature, performance with the SDMT was influenced additionally by the level of education, such that highly educated people processed information faster than less educated people (van der Elst et al., 2006). In line with previous literature, for the CVLT-II, next to age and educational level, sex was also a significant predictor of performance (Van der Elst et al., 2005). This is explained by an often shown female verbal advantage for learning and memory (Andreano and Cahill, 2009). No sex differences on SPART performance were observed, which is consistent with previous literature on visuospatial memory (Postma et al., 2004).

Although our sample size is sufficient to calculate norm scores (Boringa et al., 2001; Vanotti et al., 2016), ideally it would be best to have enough participants to create separate norm scores for subgroups, based on age, sex, and educational level. As our sample size was not large enough to create these subgroups, a combined approach was chosen in the current study: a regression-based correction for age was applied, and additionally, if sex and/or education significantly predicted test performance, subgroups based on sex and/or educational level were created. Regression based norms were shown to be valid for both the MACFIMS (Parmenter et al., 2010) and the SDMT (Burggraaff et al., 2017). Using regression-based correction for age allowed us to keep our sample size as large as possible, with $N = 39$ in the smallest subgroup (CVLT-II; women with a low educational level).

A few improvements to the Multiple Screener tool are currently being implemented. Firstly, in this first version of tool, the explanation of how to perform the tests was provided in text. One could argue that an audio explanation would resemble the original neuropsychological tests (paper and pencil-based version) better. This adjustment should be implemented in the next version of the tool. Secondly, data from several HCs (SDMT: $N = 1$; CVLT-II: $N = 4$) were excluded because of highly unlikely scores. In the backend of the next version of the tool, it will be implemented that highly unlikely scores will give a NA score (missing value), to prevent false positives (classification of impaired instead of preserved cognition). Lastly, next to the demographical confounders of cognition, especially with regard to patients with MS, depressive mood and fatigue should be taken into consideration, as these are often present in patients with MS (Boeschoten et al., 2017). For that reason, the Multiple Screener tool will be combined with on-line questionnaires on anxiety, depression, and fatigue.

The ultimate goal of the tool is to be able to detect subtle cognitive deficits early in the disease and allow for the monitoring of (subtle) cognitive changes over time. We hope that the self-explanatory Multiple Screener tool will lower the threshold for the performance of cognitive assessment at baseline and for (annual) follow-up assessment, as was recently recommended by the international MS and cognition society (IMSCOGS) and the consortium of MS centers (CMSC) (Kalb et al., 2018). The unsupervised assessment of cognitive performance is time-efficient and comes with an advantage that scores will be automatically calculated and sent to the treating neurologist immediately. If in the future novel tests will be introduced that are more sensitive to MS-related cognitive deficits and/or decline, these will be ideally implemented in the tool. However, for now, the next step would be calculate the ICCs for the three tests in patients with MS and to validate the Multiple Screener tool towards the 'gold-standard' (MACFIMS). Sensitivity, specificity and the positive and negative predictive value need to be determined, as well as test-retest reliability in patients with MS with and without cognitive impairment, to finally obtain the optimal cut-off values for impairment.

5. Conclusions

In conclusion, digital cognitive testing in patients with MS is possible. The norm scores obtained for each test allow for further validation of the Multiple Screener tool in patients with MS. Ultimately, the use of this digital tool will provide clinicians with an indication of the cognitive performance of patients with MS, without the need of a test leader. Follow-up measurement will be easier to implement and could lead to timely identification of cognitive decline in patients with MS and subsequently allow for adequate counseling.

Funding

This research was supported by Sanofi.

Declaration of Competing Interest

L. van Dongen, B. Westerik, and J.W. Twisk have nothing to disclose with regard to this work. K. van der Hiele received honoraria for consultancies, presentations and advisory boards from Sanofi Genzyme and Merck Serono. L.H. Visser received honoraria for lectures, grants for research and honoraria for advisory boards from Sanofi Genzyme, Merck Serono, Novartis and Teva. M.M. Schoonheim serves on the editorial board of Frontiers of Neurology, receives research support from the Dutch MS Research Foundation, grant number 13-820, and has received compensation for consulting services or speaker honoraria from ExceMed, Genzyme and Biogen. L. Douw receives research support from Society in Science (Branco Weiss Fellowship). J.J.G. Geurts is an editor of MS journal and serves on the editorial boards of Neurology and Frontiers of Neurology and is president of the Netherlands organization for health research and innovation and has served as a consultant for Merck-Serono, Biogen, Novartis, Genzyme and Teva Pharmaceuticals. H.E. Hulst receives research support from the Dutch MS Research Foundation, grant number 12-548, and has received compensation for consulting services or speaker honoraria from Sanofi Genzyme, Merck Serono and Biogen Idec.

Acknowledgments

We would like to thank Julia Jelgerhuis for contributing to the data acquisition.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.msard.2019.101479.

References

Andreano, J.M., Cahill, L., 2009. Sex influences on the neurobiology of learning and memory. Learn. Mem. 16 (4), 248–266.
Benedict, R.H., et al., 2006. Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). J. Int. Neuropsychol. Soc. 12 (4), 549–558.
Benedict, R.H., et al., 2012. Brief International Cognitive Assessment for MS (BICAMS): international standards for validation. BMC Neurol. 12, 55.
Benedict, R.H., et al., 2017. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult. Scler. 23 (5), 721–733.
Benedict, R.H.B., 1997. Brief Visuospatial Memory Test-Revised: Professional Manual. Psychological Assessment Resources Inc., Odessa.
Boeschoten, R.E., et al., 2017. Prevalence of depression and anxiety in multiple sclerosis: a systematic review and meta-analysis. J. Neurol. Sci. 372, 331–341.
Boringa, J.B., et al., 2001. The brief repeatable battery of neuropsychological tests: normative values allow application in multiple sclerosis clinical practice. Mult. Scler. 7 (4), 263–267.

Burggraaff, J., Knol, D.L., Uitdehaag, B.M.J., 2017. Regression-based norms for the symbol digit modalities test in the Dutch population: improving detection of cognitive impairment in multiple sclerosis? Eur. Neurol. 77 (5–6), 246–252.
Chiaravalloti, N.D., DeLuca, J., 2008. Cognitive impairment in multiple sclerosis. Lancet Neurol. 7 (12), 1139–1151.
Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. Psychol. Assess. 6 (4), 284–290.
Kalb, R., et al., 2018. Recommendations for cognitive screening and management in multiple sclerosis care. Mult. Scler. 24 (13), 1665–1680.
Klein, O.A., et al., 2018. Assessment and management of cognitive problems in people with multiple sclerosis: a National Survey of Clinical Practice. Int. J. Clin. Pract., e13300.
Langdon, D.W., et al., 2012. Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). Mult. Scler. 18 (6), 891–898.
Morrow, S.A., et al., 2010. Predicting loss of employment over three years in multiple sclerosis: clinically meaningful cognitive decline. Clin. Neuropsychol. 24 (7), 1131–1145.
Oosterhuis, H.E., van der Ark, L.A., Sijtsma, K., 2016. Sample size requirements for traditional and regression-based norms. Assessment 23 (2), 191–202.
Parmenter, B.A., et al., 2007. Screening for cognitive impairment in multiple sclerosis using the Symbol Digit Modalities Test. Mult. Scler. 13 (1), 52–57.
Parmenter, B.A., et al., 2010. The utility of regression-based norms in interpreting the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS). J. Int. Neuropsychol. Soc. 16 (1), 6–16.
Polman, C.H., et al., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol. 69 (2), 292–302.
Postma, A., et al., 2004. Sex differences for selective forms of spatial memory. Brain Cogn. 54 (1), 24–34.
Rao, S.M., 1990. Cognitive Function Study Group, N. A Manual for the Brief Repeatable Battery of Neuropsychological Tests in Multiple Sclerosis. National Multiple Sclerosis Society, New York.
Rao, S.M., et al., 1991. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. Neurology 41 (5), 685–691.
Rao, S.M., et al., 2017. Processing speed test: validation of a self-administered, iPad(R)-based tool for screening cognitive dysfunction in a clinic setting. Mult. Scler. 23 (14), 1929–1937.
Salthouse, T.A., 2010. Selective review of cognitive aging. J. Int. Neuropsychol. Soc. 16 (5), 754–760.
Smith, A., 1991. Symbol Digit Modalities Test. Western Psychological Services, Los Angeles.
Strober, L., et al., 2018. Symbol Digit Modalities Test: a valid clinical trial endpoint for measuring cognition in multiple sclerosis. Mult. Scler 1352458518808204.
Sumowski, J.F., et al., 2018. Cognition in multiple sclerosis: state of the field and priorities for the future. Neurology 90 (6), 278–288.
Van der Elst, W., et al., 2005. Rey's verbal learning test: normative data for 1855 healthy participants aged 24-81 years and the influence of age, sex, education, and mode of presentation. J. Int. Neuropsychol. Soc. 11 (3), 290–302.
van der Elst, W., et al., 2006. The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. J. Clin. Exp. Neuropsychol. 28 (6), 998–1009.
Van der Hiele, K., et al., 2014. A pilot study on factors involved with work participation in the early stages of multiple sclerosis. PLoS One 9 (8), e105673.
Vanotti, S., et al., 2016. A new assessment tool for patients with multiple sclerosis from Spanish-speaking countries: validation of the Brief International Cognitive Assessment for MS (BICAMS) in Argentina. Clin. Neuropsychol. 30 (7), 1023–1031.
Woods, S.P., et al., 2006. The California Verbal Learning Test–Second edition: test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. Arch. Clin. Neuropsychol. 21 (5), 413–420.