



**HAL**  
open science

## **MSCopilot: New smartphone-based digital biomarkers correlate with Expanded Disability Status Scale scores in people with Multiple Sclerosis**

Ian-Christopher Tanoh, Elisabeth Maillart, Pierre Labauge, Mikael Cohen, Adil Maarouf, Sandra Vukusic, Cécile Donzé, Philippe Gallien, Jérôme de Sèze, Bertrand Bourre, et al.

### ► To cite this version:

Ian-Christopher Tanoh, Elisabeth Maillart, Pierre Labauge, Mikael Cohen, Adil Maarouf, et al.. MSCopilot: New smartphone-based digital biomarkers correlate with Expanded Disability Status Scale scores in people with Multiple Sclerosis. *Multiple Sclerosis and Related Disorders*, 2021, 55, pp.103164. 10.1016/j.msard.2021.103164 . hal-03648668

**HAL Id: hal-03648668**

**<https://hal.science/hal-03648668v1>**

Submitted on 22 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## Title page

**Title:** MSCopilot: new smartphone-based digital biomarkers correlate with Expanded Disability Status Scale scores in people with Multiple Sclerosis

### Full list of authors and affiliations:

Ian-Christopher Tanoh<sup>1\*</sup>, Elisabeth Maillart<sup>2\*</sup>, Pierre Labauge<sup>3</sup>, Mikael Cohen<sup>4</sup>, Adil Maarouf<sup>5,6</sup>, Sandra Vukusic<sup>7,8</sup>, Cécile Donzé<sup>9</sup>, Philippe Gallien<sup>10</sup>, Jérôme De Sèze<sup>11</sup>, Bertrand Bourre<sup>12</sup>, Thibault Moreau<sup>13</sup>, Céline Louapre<sup>2</sup>, Morgane Vallée<sup>14</sup>, Séverine Bieuvelet<sup>14</sup>, Lissandra Klaeyle<sup>14</sup>, Anne-Laure Argoud<sup>14</sup>, Saad Zinai<sup>14</sup>, Ayman Tourbah<sup>15</sup>

1 Ecole Polytechnique, Palaiseau, France.

2 Department of Neurology, Pitié-Salpêtrière Hospital Paris, France.

3 Department of Neurology, Montpellier University Hospital, Montpellier

4 Department of Neurology, Nice University Hospital, Nice, France.

5 CNRS, CRMBM, APHM, Aix-Marseille University, Marseille

6 Pôle de Neurosciences Cliniques, Marseille

7 Department of Neurology, Hospices Civils de Lyon, Bron

8 INSERM 1028 et CNRS UMR 5292, University Lyon 1, Lyon

9 Department of Physical and Rehabilitation Medicine, Groupe Hospitalier de l'Institut Catholique de Lille, Lille

10 Physical Rehabilitation Medicine, Pôle Saint Helier, Rennes

11 Department of Neurology, Hôpital Civil, Strasbourg University, Strasbourg

12 Department of Neurology, Rouen University Hospital, Rouen

13 Department of Neurology, Dijon University Hospital, Dijon

14 Ad Scientiam, Paris Brain Institute (ICM), Pitié-Salpêtrière Hospital, Paris, France

15 Service de Neurologie, Hôpital Raymond Poincaré, Garches, UFR Simone Veil, UVSQ, INSERM U 1195, APHP, Université Paris Saclay, France

\* Both authors equally contributed to this work

### Correspondence:

Ian-Christopher Tanoh, Ecole Polytechnique, Palaiseau, France (tel.: +336 48 06 04 76; e-mail: [ian-christopher.tanoh@polytechnique.org](mailto:ian-christopher.tanoh@polytechnique.org))

**Manuscript word count:** 2713

**Running title:** MSCopilot<sup>TM</sup> scores correlate with the EDSS scores

**Keywords:** Smartphone application, Multiple Sclerosis, Digital Health, Expanded Disability Status Scale (EDSS), Multiple Sclerosis Functional Composite (MSFC)

## **Abstract**

### **Background**

A previous clinical study showed the high specificity, sensitivity and reliability of MSCopilot, a software medical device designed by Ad Scientiam for the self-assessment of people with Multiple Sclerosis (PwMS), compared to the traditional Multiple Sclerosis Functional Composite (MSFC). We conducted further analyses to assess MSCopilot's performance with respect to the Expanded Disability Status Scale (EDSS).

### **Methods**

The data of 116 PwMS were analysed. We studied the correlations between MSCopilot scores and the EDSS, and their ability to distinguish PwMS with high and low EDSS through a study of the distribution of the digital test scores as well as logistic regression models. The same analyses were performed using the MSFC tests.

### **Results**

MSCopilot composite scores were as highly correlated to the EDSS ( $|r|=0.65$ ,  $p<0.01$ ) as their MSFC counterparts, confirming the known correlation of the MSFC with the EDSS. In a linear regression framework, the Walking digital tests have good explanatory power, especially for PwMS with  $EDSS>3.5$  ( $R^2_{adj}=0.47$ ). The mean values of each MSCopilot subscore were significantly different between patients with an  $EDSS > 3.5$  and others ( $p<0.05$ ), which could not be proved for the MSFC Cognition tests. MSCopilot4 was the best model to predict an EDSS score  $>3.5$  ( $AUC=0.92$ ).

### **Conclusion**

These analyses confirm the reliability of MSCopilot and show interesting correlations with the EDSS (similar results obtained with the MSFC). MSCopilot was able to highlight nuances in the different stages of MS the MSFC could not capture.

## Introduction

Multiple Sclerosis (MS) is a chronic disease of the central nervous system, with a broad diversity of symptoms and different progression phenotypes. Over two million people worldwide suffer from MS [1]. The most accepted tool for quantifying neurologic disability is the Expanded Disability Status Scale (EDSS) [2]. It measures the magnitude of the MS symptoms on a scale from 0 to 10 (0: healthy, 10: deceased). This scale has been widely used since 1983 and has the advantage of considering the functional systems that may be affected by MS, and the convenience of reflecting the clinical status of each patient with a simple number [3]. However, the EDSS presents several shortcomings: for moderate scores, this scale resembles an ambulation index, which cannot be enough to monitor a disease as complex as MS. Moreover, a single score can provide several interpretations. For instance, an EDSS score of 4.5 indicates that a patient can walk 300m without aid, while it may also mean that a patient is fully ambulatory with four points from one functional system (others at zero or one point). Also, an increase of 1 point has different significations for each score. Indeed, small differences in a patient's state are reflected at lower EDSS, which is not the case for patients with higher scores. As a result, the EDSS is not sufficient to assess the state of patients through longitudinal monitoring.

To tackle these shortcomings, the Multiple Sclerosis Functional Composite (MSFC) was developed as a three-component, standardized, quantitative, performance scale evaluating the degree of impairment in people with MS (PwMS) [4]. It measures the walking speed, using a 25-foot walk ("Timed 25-Foot Walk" test, T25FW), arm and hand dexterity with a peg test ("9-Hole Peg Test", 9HPT), and cognitive functions ("Paced Auditory Serial Addition Test", PASAT). Each test provides a subscore, using z-scores, which are then averaged to obtain a MSFC composite score. In order to include an assessment of the vision ability, the Sloan Low Contrast Letter Acuity Test (SLCLAT [5]) can be added to the original tests, resulting in a four-component scale, referred to as the "MSFC4". To avoid confusion, the original three-component MSFC is referred to as "MSFC3" in the following. The Symbol Digit Modalities Test (SDMT) was proposed as an alternative to PASAT, notoriously disliked by patients [6], resulting in the "MSFC3-revised" and "MSFC4-revised" versions. However, a practice effect and the difficulty of interpreting a change in z-scores in clinical practice have limited the use of the MSFC in clinical trials [7].

The MSCopilot mobile application was developed by Ad Scientiam (Paris Brain Institute, Pitié-Salpêtrière Hospital, Paris, France) with the purpose of using multi-dimensional data of MS disabilities to facilitate the clinical assessments of patients and refine the quality and accuracy of both clinicians and patients' knowledge of the disease progression. MSCopilot is a four-module application inspired by the MSFC4-revised. It features a Mobile Walking Test (MWT), a Mobile Dexterity Test (MDT), a Mobile Cognition Test (MCT), and a Mobile Vision Test (MVT). In the following, MSCopilot3 refers to the MSCopilot without MVT, whereas MSCopilot4 refers to the four-component version. To assess the performances of our application, we conducted a clinical study with 116 PwMS and 69 Healthy Controls (HCs) which successfully validated our primary aim which was to prove that MSCopilot3 and MSFC3 composite scores allowed to distinguish equally PwMS from HCs [8]. As for the secondary endpoints, our study showed high correlations between the other versions of MSCopilot and their MSFC counterparts, the non-inferiority of the MSCopilot subscores compared to the MSFC's, and the good reproducibility of MSCopilot through a test-retest

study. Our study also demonstrated excellent acceptability from both patients and clinicians [9].

As a next step, we performed further analyses to assess the performance of MSCopilot with respect to EDSS. We present these results in this paper. Analyses were performed on the data collected on the 116 PwMS of the initial clinical study. We studied the correlations between the MSCopilot scores and EDSS, and their ability to distinguish PwMS with high and low EDSS. To this end, we studied the distribution of the digital test scores and built logistic regression models. For each of these analyses, we compared the performances of MSCopilot with those achieved by the various versions of the MSFC.

## **Materials and Methods**

### *Origin of the data*

The data used for the following analyses are the EDSS, MSCopilot and MSFC composite scores and subscores (scores for the included tests) of the 116 PwMS from the initial clinical study mentioned in the introduction.

This clinical study was national, multicentre, controlled, randomized, and low interventional. Between 02/10/2017 and 30/07/2018, 141 PwMS and 76 HCs were enrolled in 11 French MS centres. During a single visit, participants were screened for eligibility and at each centre, a Neurostatus level-C certified neurologist measured the EDSS score of the enrolled participants. Several neurologists from the same team were therefore involved to avoid bias. In case of re-tests during a second visit (n=46), the neurological exam was performed by the same neurologist. All the EDSS related tests were performed in a standardized manner (all study examiners were briefed on how to perform the test), and not merely extracted from patient files. On the same day, the participants were randomized (1:1) to perform first the MSFC tests and then the MSCopilot tests (group A) or the MSCopilot tests followed by the MSFC ones (group B). To avoid MSFC variants repetition, participants had to perform each individual test only once in the following order: 1) SDMT, 2) T25FW, 3) 9-HPT, 4) PASAT, 5) SLCLAT. The test order for MSCopilot was: 1) MWT, 2) MDT, 3) MCT, 4) MVT. 185 participants completed the full set of digital and traditional tests without major protocol deviation (116 PwMS; 69 HCs).

### *Statistical analyses*

For each analysis performed, only p-values of less than 0.05 were considered significant. When noted, the population studied in the following analyses was stratified by their EDSS scores (EDSS $\leq$ 3.5 vs EDSS $>$ 3.5).

Analyses of the correlations between the EDSS score and the MSFC or MSCopilot composite scores were carried out using the psych package from the R software. The significance of the correlations was computed with the correlation test, whereas the significance of the difference between two different correlations was calculated using a Williams's test. In addition to the entire PwMS population, this analysis was also performed separately on patients with EDSS $\leq$ 3.5 and patients with EDSS $>$ 3.5.

We then performed a multiple correlation analysis between the combination of the subscores and the EDSS scores. To do so, we implemented linear regression models, with the *lm* function from R. The covariates of these models are various subscores (either from MSFC or MSCopilot), while the predicted variable is the patient's EDSS. These linear models can be specified as follows:

$$EDSS_i = \beta_0 + \beta_1 S_i^{(1)} + \dots + \beta_p S_i^{(p)} + \epsilon_i,$$

where  $EDSS_i$  is the EDSS of the  $i$ -th patient,  $S_i^{(1)}, \dots, S_i^{(p)}$  are  $p$  subscores of the  $i$ -th patient used as covariates, and  $\epsilon_i$  is the error term which follows a normal distribution with mean zero and whose variance does not depend on the patient. Finally,  $\beta_0, \dots, \beta_p$  are the model's coefficients.

The significance and performance of the models were assessed by analysing the  $p$ -values of the Student's tests performed for the coefficients of the model, and the value of  $R^2_{adj}$ .  $R^2_{adj}$  was chosen over  $R^2$  since it adjusts for the number of explanatory terms in the model. We use  $R^2_{adj}$  to compare models with a different number of predictors. In addition to the entire PwMS population, this analysis was also performed separately on patients with  $EDSS \leq 3.5$  and patients with  $EDSS > 3.5$ .

For patients' discrimination, an analysis of the subscore was performed using the Wilcoxon's test (with the *wilcoxon.test* function from R), to assess the significance of the difference between the mean subscores for patients with  $EDSS > 3.5$  vs those with  $EDSS \leq 3.5$ .

Finally, for patients' classification, we implemented a logistic regression to predict if the EDSS of a PwMS is greater than 3.5 or not. Composite scores or subscores were used as covariates. For the test performance evaluation, PwMS were split into two datasets using a stratified sampling method (respectively by sex, age, education, Body Mass Index, EDSS score, Multiple Sclerosis type). One dataset was randomly chosen to train the logistic regression model (58 PwMS), while the second dataset was used to evaluate the performance of the model (58 PwMS). The performance was measured by the Area under the Receiver Operating Characteristic (ROC) curve (AUC). ROC curves were computed and drawn using the *pROC* package from R [10]. For two different models, the AUC equality was tested using a two-sided test provided by the *roc.test* function of the *pROC* package.

## Results

**Cohort:** 116 PwMS. 48% of the PwMS had an EDSS score greater than 3.5 and under 7, while the rest had a score between 0 and 3.5. The proportion of PwMS with Relapsing MS (RMS) was 74% and 26% with Progressive MS (PMS). Additional data regarding the characteristics of the global population (PwMS and HCs of the initial study) can be found in Table 1.

### *Correlation between Composite Scores and the EDSS*

Regardless of the population (all PwMS, PwMS with  $EDSS < 4$  or  $EDSS > 3.5$ ), the correlation between the MSCopilot3 or 4 composite scores and the EDSS was significantly

nonzero (correlation test p-value < 0.05). When considering all PwMS, the correlation between MSCopilot composite scores and the EDSS was good (Figure 1) with  $|r|=0.65$  for MSCopilot3 and  $|r|=0.62$  for MSCopilot4 ( $p<0.001$ ). We also focused on the lower range of the EDSS (<4). The correlations between MSCopilot composite scores and the EDSS remain significant (p-value=0.01), but the value of the Pearson correlations, though still moderate, are lower in absolute value (0.4 vs about 0.6 for all the PwMS).

The correlations of the MSFC composite scores and the EDSS were comparable to the performances of the MSCopilot composite scores and a significant difference between the correlations obtained for MSCopilot and the MSFC could not be proved (Williams's test p-value > 0.05). MSCopilot composite scores were as highly correlated to the EDSS as the MSFC ones.

### *Multiple Correlation between MSCopilot subscores and the EDSS*

Studying the correlations between subscores and the EDSS in PwMS emphasized the prominent role of the Walking test. We noticed that any combination of two tests including the Walking test (e.g., Walking + Dexterity) provided a significant model of the EDSS with a noticeable explanatory power (linear regression t-test p-values < 0.05 and  $R^2_{adj} \approx 0.45$ ). Replacing the Walking test with another test produced either insignificant models of the EDSS or very poorly correlated ones. When the focus was on PwMS with an EDSS > 3.5, among all possible test combinations, the MSCopilot Walking test was the only one that significantly modelled the EDSS, with a notable correlation (linear regression t-test p-value < 0.05 and  $R^2_{adj}=0.47$ ). Replacing the MSCopilot Walking test with the T25FW yielded similar results.

### *Patients' discrimination based on the distribution of the digital subscores*

We found that the means of each MSCopilot subscore were significantly different between patients with EDSS greater than 3.5 or not (Wilcoxon test p-value < 0.01; Figure 2 for MSCopilot Cognition test).

We compared these results with the performances achieved by the MSFC; the mean values were not significantly different between these two EDSS classes when we considered one of the two MSFC Cognition tests, with a Wilcoxon p-value for PASAT of 0.06 (Figure 3), and a Wilcoxon p-value for SDMT of 0.055 (Figure 4), whereas the Wilcoxon p-value for the MSCopilot Cognition test was 0.009.

### *Binary EDSS Classification based on digital scores*

We performed the classification of patients according to two EDSS classes (EDSS>3.5 or not) given the scores of the digital tests.

When considering MSCopilot subscores as covariates, the best performance was achieved by the Walking test (AUC=0.85), while the other tests provided lower AUC (0.77 for Dexterity, 0.72 for Cognition, 0.78 for Vision). As for MSFC subscores, the AUC obtained were lower

than for their MSCopilot counterparts (0.74 for T25FW and 9HPT, 0.57 for PASAT, 0.67 for SDMT, 0.78 for SLCLAT).

Better results were obtained when using composite scores as covariates. Indeed, the best model overall used MSCopilot4 as covariate and provided an AUC of 0.92, while the best MSFC-based model used MSFC4-revised as covariate with an AUC of 0.84 (Figure 5). However, we could not prove that the AUC of the MSCopilot4 model was significantly superior to the AUC of the MSFC4-revised model ( $p=0.29$ ). Using MSCopilot3 as a covariate also gave a very strong performance (AUC=0.90).

For each of these analyses, the AUC of the model using a MSCopilot test (whether composite score or subscore) as covariate was not significantly different from its MSFC counterpart ( $p>0.05$ ).

## Discussion

Our results show that the MSCopilot composite scores were as highly correlated to the EDSS as their MSFC counterparts. This encouraging outcome holds for both MSCopilot3 and 4 composite scores. The correlations between the EDSS and the MSFC, already demonstrated with the paper version [11], were later studied with an iPad version equipped with hardware enclosure (incorporating receptacles for metal pegs to mimic the 9HPT and a Bluetooth remote for the walking speed test) [12]. Our smartphone version might be more patient-friendly both in-clinic and in free-life.

The multiple correlation analysis emphasized the prominence of the MSCopilot Walking test when studying the correlation between subscores and the EDSS. It supports the claim that the EDSS mainly assesses the walking ability of patients, especially when the score is above 3.5 [2].

By significantly distinguishing patients with high or low EDSS, MSCopilot subscores validate, through their distributions, that not only the walking ability but also the state of vision, dexterity, and cognition are significantly different depending on the stage of the disease. MSCopilot seems to capture nuances that may be missed by the EDSS scaling. Moreover, unlike MSCopilot, we could not prove that any of the MSFC Cognition tests significantly distinguished patients with high or low EDSS. These findings must be qualified by the fact that 7% of the initially enrolled patients were excluded from the analyses (i.e., not included in the  $n=116$  PwMS) because they failed to complete the PASAT (too difficult or too tired). Consequently, the disappointing performance of the PASAT could reflect participants' fatigue and needs to be confirmed in other instances.

Finally, we implemented a logistic regression using subscores or composite scores as covariates to predict whether patients had an EDSS greater than 3.5 or not. Our results show that among all digital tests, the best performance was achieved when using the MSCopilot Walking test (AUC=0.85). This represents additional proof that walking faculties play a prominent role in the EDSS scaling. The other MSCopilot subscores also provided notable AUC (above 0.72), confirming their part in assessing the stage of the disease. The best overall classification was obtained by an MSCopilot model using the MSCopilot4 scores as our covariate (AUC=0.92). Though the AUC was 8% better than the one obtained with the best



MSFC-based model, we could not conclude that, in general, the MSCopilot-based model was significantly better ( $p=0.29$ ). The original clinical study was designed to discriminate PwMS from HCs and the participants were chosen accordingly [8]. Consequently, to train and test our logistic regression models, HCs were removed from the data, leaving only 58 PwMS to train, and 58 for testing. A wider study solely focused on PwMS could allow to conclude to statistical significance. One future goal would be to use our digital tests to have a more accurate modelling of lower EDSS and, ultimately, to predict the precise EDSS score of any PwMS. As we write, a new version (MSCopilot V2) is being developed. It would be interesting to compare it to the digital version of the Neurostatus and its improved consistency of standardized EDSS assessments [13].

As a CE-marked software medical device, MSCopilot facilitates the self-monitoring of patients with MS in free-life (at home). Currently, the application comes with pre-set testing frequencies: once a month for dexterity and walking, every 3 months for cognition and vision. These settings can be adapted for a specific usage (daily, weekly, etc.) depending on the research requirements. Using their own smartphone, the patients can perform their assessments alone, without the assistance of a nurse.

For the time being, MSCopilot is indicated as an add-on to the clinical examination. Several surveys are planned to evaluate the place of MSCopilot in the care pathway.

## **Conclusion**

Our analyses demonstrated that the MSCopilot3 or 4 composite scores were as highly correlated to the EDSS as their MSFC counterparts and confirmed the known correlations between the MSFC and the EDSS. Additionally, the MSCopilot Walking test provides a good understanding of the patients' EDSS. Our work also shows that MSCopilot subscores (Vision, Dexterity, and Cognition) will vary depending on the stage of the disease. To have a full grasp of this phenomenon, we built logistic regression models to predict whether a patient's EDSS score was high or low, based on its MSCopilot or MSFC test scores. The best model uses MSCopilot4 scores and can clearly distinguish patients with high or low EDSS. To capture subtle nuances and be able to forecast the precise EDSS score, an advanced version would be interesting to explore new and more accurate digital biomarkers for the Dexterity, Vision, and Cognition tests.

## **Acknowledgements**

The authors thank the patients and the volunteers, H el ene Brissart (Nancy, France), Claude Mekies (Toulouse, France), Catherine Vignal-Clermont (Paris, France) for their help in the digital tests' conception; Yann le Coz (Paris, France) for his help in digital tests elaboration; from the Paris Brain Institute, la Piti -Salp tri re Hospital, Paris, France, Jean-Christophe Corvol for the study organization; Vincent Perlberg and Fran ois-Xavier Lejeune for the statistical and analytical plans. This work was funded by Ad Scientiam (Paris, France) and has received an unrestricted educational grant from Roche SAS France.

Trial registration: NCT03148938.

## References

- [1] C. Cotsapas, M. Mitrovic, and D. Hafler, 'Multiple sclerosis', in *Handbook of Clinical Neurology*, vol. 148, Elsevier, 2018, pp. 723–730.
- [2] J. F. Kurtzke, 'Rating neurologic impairment in multiple sclerosis An expanded disability status scale (EDSS)', *Neurology*, vol. 33, no. 11, pp. 1444–1444, Jan. 1983, doi: 10.1212/WNL.33.11.1444.
- [3] B. P. Çınar and Y. G. Yorgun, 'What We Learned from The History of Multiple Sclerosis Measurement: Expanded Disability Status Scale', *Noro Psikiyatri Arsivi*, vol. 55, no. Suppl 1, pp. S69–S75, 2018, doi: 10.29399/npa.23343.
- [4] G. R. Cutter *et al.*, 'Development of a multiple sclerosis functional composite as a clinical trial outcome measure', *Brain J. Neurol.*, vol. 122 (Pt 5), pp. 871–882, May 1999.
- [5] L. J. Balcer and E. M. Frohman, 'Evaluating loss of visual function in multiple sclerosis as measured by low-contrast letter acuity', *Neurology*, vol. 74 Suppl 3, pp. S16-23, Apr. 2010, doi: 10.1212/WNL.0b013e3181dbb664.
- [6] L. A. S. Walker, A. Cheng, J. Berard, L. I. Berrigan, L. M. Rees, and M. S. Freedman, 'Tests of Information Processing Speed', *Int. J. MS Care*, vol. 14, no. 2, pp. 92–99, 2012, doi: 10.7224/1537-2073-14.2.92.
- [7] Meyer-Moock, S., Feng, Y.-S., Maeurer, M., Dippel, F.-W. & Kohlmann, T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurol.* 14, 58 (2014).
- [8] E. Maillart *et al.*, 'MSCopilot, a new Multiple Sclerosis self-assessment digital solution: results of a comparative study versus standard tests', *Eur. J. Neurol.*, Sep. 2019, doi: 10.1111/ene.14091.
- [9] E. Maillart and al., 'Acceptability in clinical practice of MSCopilot, a smartphone application for the digital self-assessment of patients living with MS', *Mult. Scler. J.*, vol. 24, no. 2\_suppl, pp. 328–529, Oct. 2018, doi: 10.1177/1352458518798590.
- [10] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011 12: 77.
- [11] R. A. Rudick *et al.*, 'Assessing disability progression with the Multiple Sclerosis Functional Composite', *Mult. Scler. Houndmills Basingstoke Engl.*, vol. 15, no. 8, pp. 984–997, Aug. 2009, doi: 10.1177/1352458509106212.
- [12] Rudick RA, Miller D, Bethoux F, Rao SM, Lee JC, Stough D, Reece C, Schindler D, Mamone B, Alberts J. The Multiple Sclerosis Performance Test (MSPT): an iPad-based disability assessment tool. *J Vis Exp.* 2014 Jun 30;(88):e51318. doi: 10.3791/51318. PMID: 25046650; PMCID: PMC4209820.
- [13] D'Souza M, Yaldizli Ö, John R, Vogt DR, Papadopoulou A, Lucassen E, Menegola M, Andelova M, Dahlke F, Schnyder F, Kappos L. Neurostatus e-Scoring improves consistency of Expanded Disability Status Scale assessments: A proof of concept study. *Mult Scler.* 2017 Apr;23(4):597-603. doi: 10.1177/1352458516657439.

## **Table caption**

**Table 1.** Baseline demographic characteristics of the per protocol population

## **Figure caption**

**Fig 1.** Correlation MSCopilot3 vs EDSS (Patients, N=116)

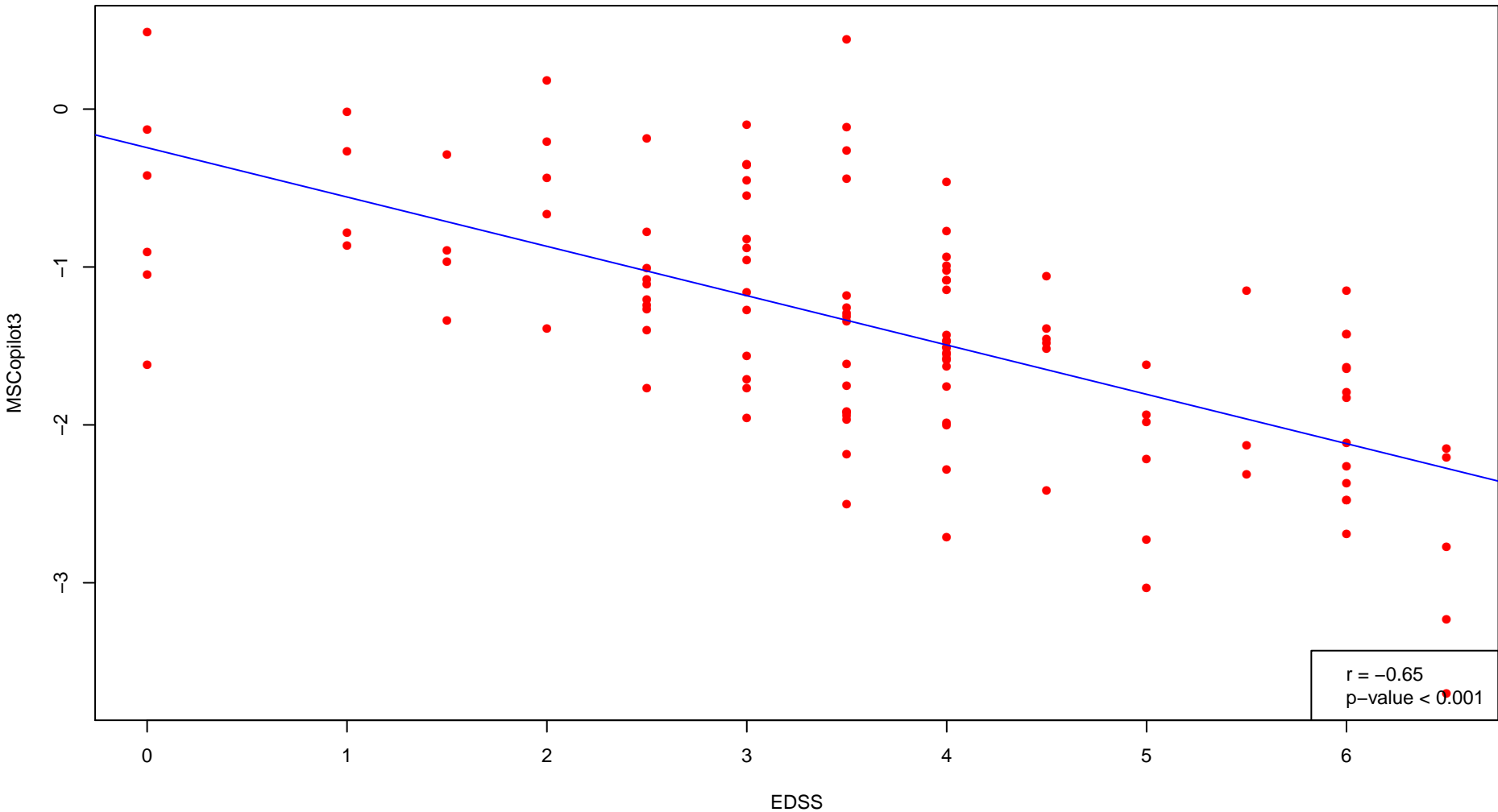
**Fig 2.** Distribution of the MSCopilot cognition test (N=185)

**Fig 3.** Distribution of the MSFC PASAT test (N=185)

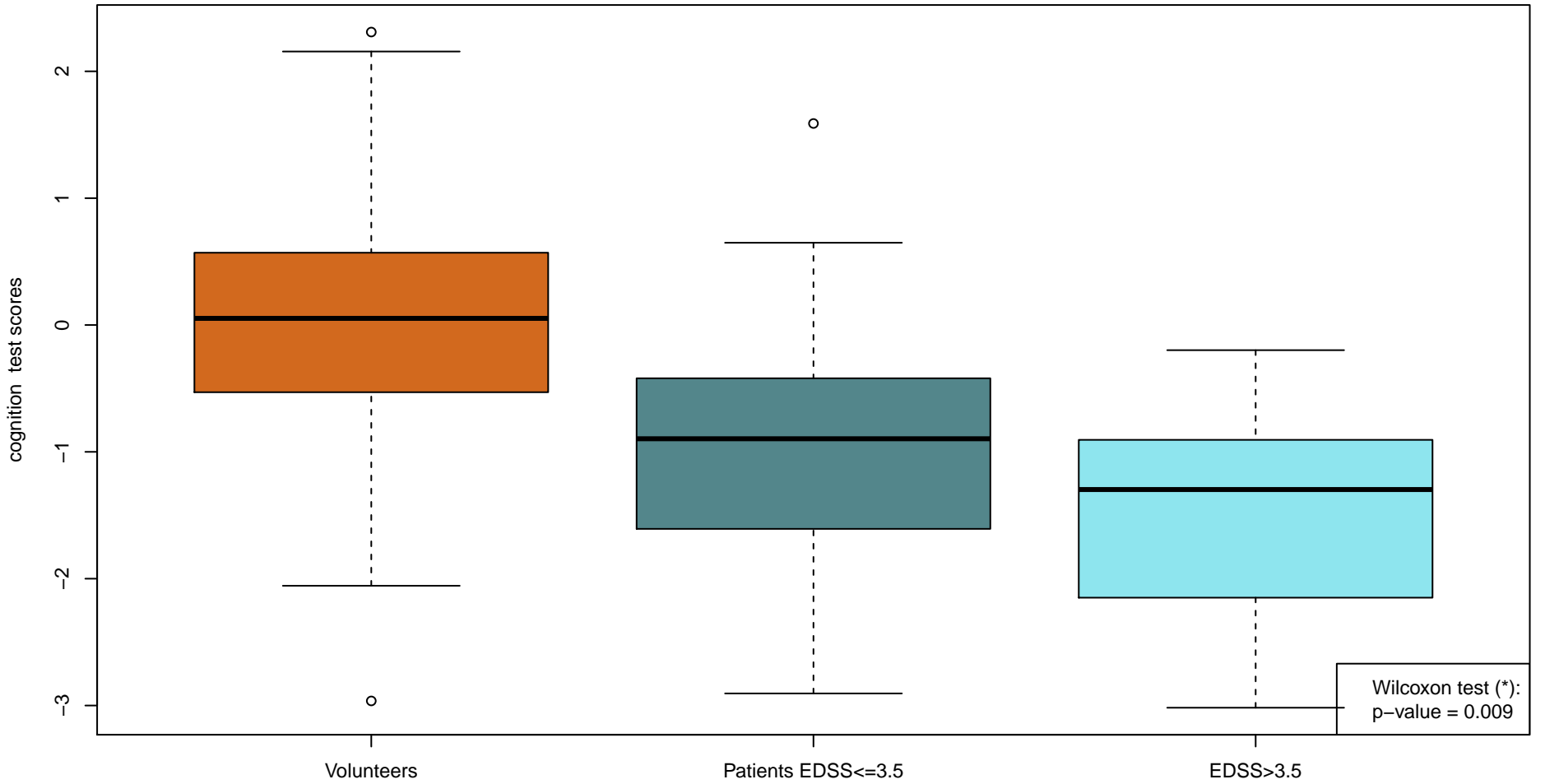
**Fig 4.** Distribution of the MSFC SDMT test (N=185)

**Fig 5.** ROC curve, MSCopilot4 vs. MSFC4-revised (Patients, N=58)

Correlation MSCopilot3 vs. EDSS (Patients, N=116)

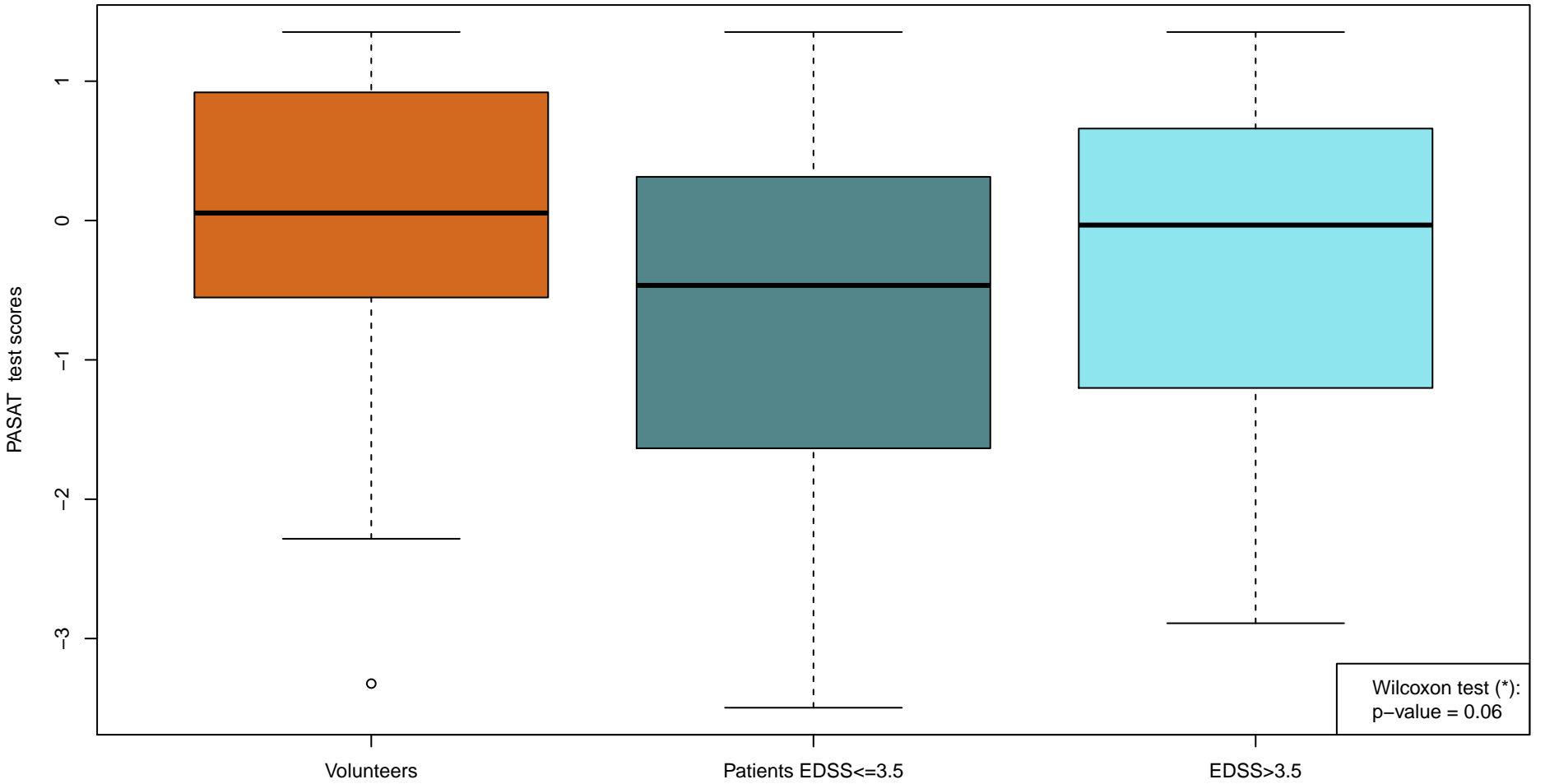


# Distribution of the MSCopilot cognition test (N=185)



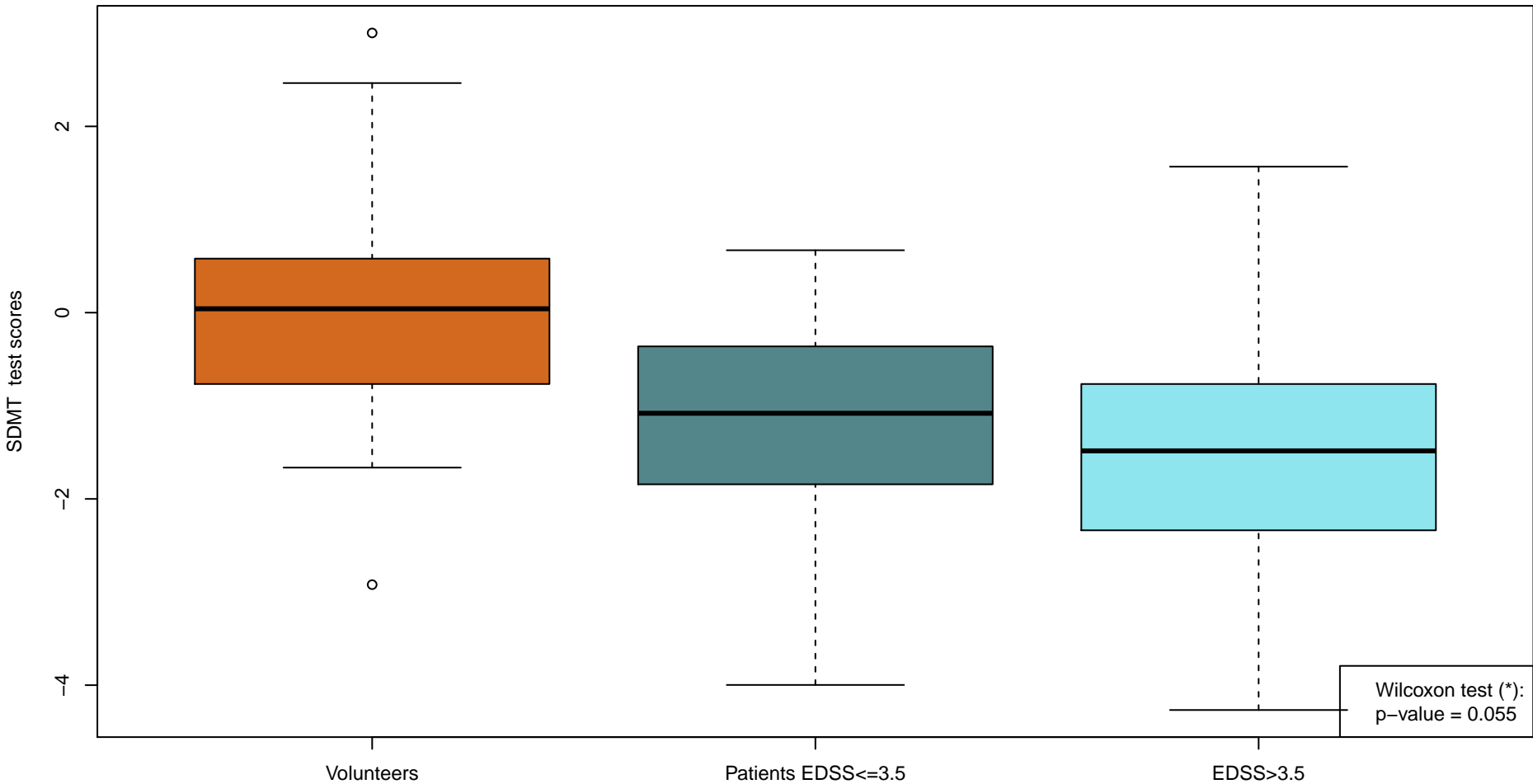
(\*) Wilcoxon test was carried out without the volunteer category

# Distribution of the MSFC PASAT test (N=185)



(\*) Wilcoxon test was carried out without the volunteer category

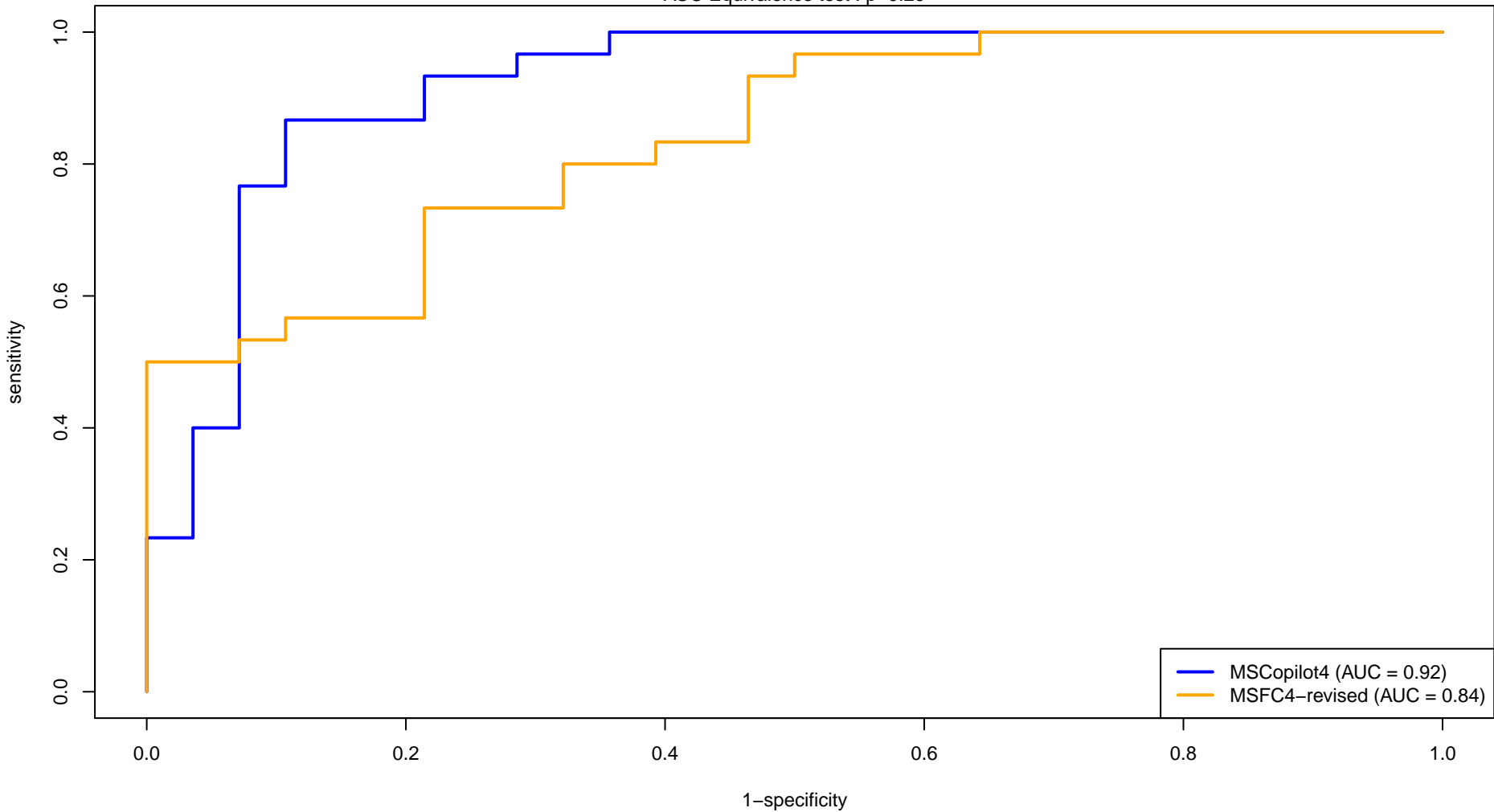
# Distribution of the MSFC SDMT test (N=185)



(\*) Wilcoxon test was carried out without the volunteer category

# ROC curve, MSCopilot4 vs. MSFC4-revised (Patients, N=58)

AUC Equivalence test :  $p=0.29$





**Table 1.** Baseline demographic characteristics of the per protocol population.

	MS patients (N=116)	Healthy controls (N=69)
Gender		
Male, n (%)	45 (38.8)	24 (34.8)
Female, n (%)	71 (61.2)	45 (65.2)
Age, years		
Mean (SD)	46 ( $\pm$ 10)*	39 ( $\pm$ 11)
Range	43 (65-22)	40 (62-22)
BMI, mean (SD)	24.1 ( $\pm$ 5)	24.4 ( $\pm$ 4.5)
Education level		
Primary	6 (5.2)	0 (0)
Secondary	38 (32.8)	17 (24.6)
2nd year university level	44 (37.9)	27 (39.1)
5th year university level	26 (22.4)	22 (31.9)
$\geq$ 8th year university level	2 (1.7)	3 (4.4)
Duration of MS, years		
Mean (SD)	12 ( $\pm$ 7)	
Median (range)	11 (30)	
Disease phenotype, n (%)		
RRMS	86 (74)	
PMS	30 (26)	
Last measured EDSS score		
Mean (SD)	3.6 ( $\pm$ 1.6)	
Median (range)	3.5 (6.5)	
EDSS [0-2], n (%)	19 (16)	
EDSS [2.5-3.5], n (%)	41 (35)	
EDSS [4-7], n (%)	56 (49)	

\*  $p=6.645e-06$  vs healthy controls

Abbreviations: RRMS: Relapsing-remitting multiple sclerosis; PMS: Progressive multiple sclerosis; BMI: Body mass index; SD: Standard deviation; EDSS: Expanded disability status scale.