AM_471108/ AM_471109/ AM_471110 Submission form

Author information

Student name and number	Justus F. Huebotter
Title manuscript	Towards predicting cognitive profiles in MS: A translational approach on network dynamics in fMRI data
Course	AM_471108/ AM_471109/ AM_471110
Name VU supervisor	Menno Schoonheim
Name daily supervisor (if applicable)	Kim Meijer
Name on-site supervisor (if applicable)	

Journal information

Name of peer-reviewed journal	Neurology
Audience	Clinical focus
Aims & Scope	*"The journal's purpose is to advance the field by presenting new basic and clinical research with emphasis on knowledge that will influence the way neurology is practiced."*

Manuscript information

Type of manuscript	Research report (Internship)
Title - requirements	A maximum of 96 characters
Abstract - requirements	A maximum of 250 words
Introduction - requirements	A maximum of 250 words
Materials & Methods	
Results	
Discussion	
References	Reference style: APA / Neurology
Figures and tables	
Supplementary material (if applicable)*	
Additional requirements (if applicable)	

* Sometimes not all your results are suited for a manuscript. E.g. pilot experiments with small *n*, negative results, or data only slightly related to the research question described in the manuscript. Alternatively, you might have collected more data than can be included in the manuscript, thereby exceeding the maximum number of figures allowed. In such instances, you can decide to add these results as supplementary material. This way, you will meet the journal's requirements, yet be able to give a complete overview of all the work you have done.

NOTES

- To increase readability: *single*-spaced text (i.e. ignore journal requirements)
- Grade registration requires you to upload a < 4MB manuscript. This means you do *not* have to meet the journal requirements regarding image quality.

Towards predicting cognitive profiles in MS: A translational approach on network dynamics in fMRI data

OBJECTIVE: Here we expand the properties of resting state fMRI data beyond widely used network correlations to predict cognitive deficits. This was done to investigate if the experience gained in stock markets can be translated to obtain more insight into biological processes due to a high number of temporal similarities. This is towards gaining a better understanding for MS patients' individual cognitive development profile.

METHODS: Brain network timeseries were obtained by a data-driven approach using ICA-FIX, melodic and dual regression within the FSL 5 software. Properties of timeseries were obtained by calculating time series indicators derived from stock markets for a wide range of timescales. These properties are used as features in regression models, allowing comparison with a baseline model, based on the respective Pearson's correlation coefficient of predicted and true average cognition scores.

RESULTS: Robust predictors are found by cross-validated recursive feature elimination and ridge regression models. Results indicate that these predictors can outperform a traditional correlation-based model on the same data. Although group differences for all predictors are non-significant, some indicators show potential as biomarkers to identify cognitive deficits based on fMRI data.

CONCLUSIONS: Robust stock market indicator-based features are believed to be able to grant new insights in mechanisms in brain networks dynamics leading to cognitive deficits. The heterogeneity of cognitive ability and affected domains within the present dataset do not allow for predicting complete and individual cognitive profiles. Larger studies are necessary to further interpret and validate time-series indicators in fMRI.

Keywords: *Multiple sclerosis, functional magnet resonance imaging, dynamic functional connectivity, brain networks, cognition, time series analysis, machine learning, sliding window*

Introduction

Multiple sclerosis (MS) is heterogeneous in symptoms and progression, with 40 – 70% of affected patients suffering from some sort of cognitive deficits[1,2], potentially starting early in disease[3]. Given the current lack of effective treatments[4] and biomarkers,[5] it is important to work towards preventative measures rather than symptomatic treatments, to minimize the impact on the patients' daily life. Early and individual treatment, based on the respective cognitive profile of affected subjects, is desired to prevent progression of cognitive deficits as they get increasingly irreversible over disease progression.

Structural imaging techniques, such as magnet resonance imaging (MRI - traditionally used to detect lesion location and size[6]) or diffusion tensor imaging (to identify what white matter tracks are affected by found lesions) fail to explain severity or affected domain of cognitive deficits reliably.[7] While grey matter atrophy can be easily linked to compromised function of the corresponding cortex region, the often occurring white matter lesions are hard to link with an affected cognitive domain[8]. To create disease models explaining these, fMRI studies have investigated the correlation of brain network activity, especially with the default mode network[9,10] (DMN), or graph-theory based measures.[11] Understanding on what timescale relevant mechanisms occur is important when considering the dynamic properties of network activies.[12–15] Contradictory findings associate cognitive deficits with both increased and decreased (dynamic) functional connectivity (FC) of specific networks[9], highlighting

the need to expand the horizon of temporal properties investigated. However, statistical time series analysis is not limited to the biomedical sciences, with such data also used in financial markets.[16] Similarities are evident between temporal brain network activity and stock prices, such as wave-like oscillations on multiple overlaying timescales, reaction to external stimulation and non-independent course development of related areas.

Therefore, to predict cognitive disfunction, this work considers network activity correlations (functional connectivity measures) and additional dynamic properties of resting state (RS) fMRI data. Statistical values inspired by stock indices are translated for data-driven RS network timeseries and given the timescale for relevant properties is unknown[12–14], a range of temporal windows is covered. This work explores a translational approach to detecting and understanding underlying mechanisms occurring with cognitive deficits. This proof-of-concept is focused on cross-sectional prediction of average cognitive abilities, aiming to evaluate if new predictors can perform similarly to a traditional network-correlation based baseline regression model. Careful interpretation of the most important indicators could lead to insights on mechanisms that lead to specific cognitive deficits and enable the creation of individual cognitive profiles for MS patients.

Methods

Participants. All participants were part of the Amsterdam MS cohort, as previously described[17], consisting of 234 patients diagnosed with clinically definite MS according to the revised McDonald criteria[18] (MS; 32 % male, age 47.61 ± 11.02 years, symptom duration 14.6 ± 8.4) and 60 matched healthy controls (HC; 42 % male, age 46.45 ± 9.91 years). Of the patients with MS, 181 patients were diagnosed with relapsing remitting MS, 33 patients with secondary progressive MS, and 20 with primary progressive MS. All patients were relapse-free and without steroid treatment for at least two months prior to participation in the study

Standard protocol approvals, registrations, and patient consents. The study protocol was approved by the Vrije Universiteit Medical Center ethics review board, and all participants gave written informed consent before participation.

Neuropsychological testing. On the day of fMRI scanning, all participants underwent extensive neuropsychological testing with an extended version of the Brief Repeatable Battery of Neuropsychological tests (BRB-N)[19] as previously described.[10,17,20,21] The assessed cognitive domains include: Executive functioning (concept shifting test), verbal memory (selective reminding test), verbal fluency (word list generation), information processing speed (symbol digit modalities test), visuospatial memory (spatial recall test), attention (Stroop color-word test) and working memory (memory comparison test), as well as an average cognition composite score. The test scores were corrected for effects of age, sex and education, as well as standardized, based on the HC observations.[22] Only subjects without missing scores were included for prediction of the average cognition score.

MRI scans. All subjects were scanned on a 3T General Electric system (Signa-HDxt, Milwaukee, WI, USA), utilizing an eight-channel phased-array head coil, as previously described.[10,17,20,21] 3D T1-weighted images were used for registration, using fast spoiled gradient echo sequence (repetition time 7.8ms, echo time 3ms, inversion time 450ms, flip angle 12°, 1.0mm sagittal slices, 0.9 × 0.9mm² in-plane resolution). Functional MRI data covering the whole brain was obtained by acquiring echo planar images (202 volumes, repetition time 2200ms, echo time 35ms, flip angle 80°, 3mm contiguous axial slices, 3.3 × 3.3mm² in-plane resolution) with eyes closed.

Image preprocessing. The preprocessing pipeline followed was previously described in detail[23] and included: fMRI image registration to the 3D T1-weighted images, brain extraction, motion correction (ICA-FIX)[24,25], spatial smoothening and removal of first two volumes, slow temporal drifts as well as other artifacts. This results in voxel activation series (200 timepoints each) within standard space for comparability between subjects with improved signal-to-noise ratio.

Extraction of resting-state networks. Independent component analysis was used to extract the spatial patterns from all subjects and resulted in 50 components, of which 14 were manually chosen for further analysis, which were identified and labeled as the following: Auditory, Cerebellum a, Cerebellum b, DMN I, DMN II, Executive control, Frontoparietal network (FPN) I, FPN II a, FPN II b, Sensorimotor, Visual a, Visual b, Visual c & Visual d, shown in Figure 1. Activation time series were obtained per subject using the previously described dual regression protocol within FSL 5.[24,25]

Obtaining statistical timeseries indicators. For the signal timeseries (S(t)) of each network of every subject several timescale dependent indicators were calculated using a specifically developed script in python 2.7 (https://www.python.org/). Time windows (Δt) for the rolling parameters varied between 5 and 100 timepoints (n) (11 – 220s). For every timepoint t_i, several statistical properties of S are obtained (for a detailed list of indicators and applied time windows see Table 1). Because indicators have variable initialization lengths, only their last 100 timepoints were used for further analysis. Single values for the predictive model were obtained for each subject, network and indicator by calculating the following measures: Mean, standard deviation, minimum, 25% quantile, median, 75% quantile, maximum, range, skewness & kurtosis. This resulted in a total of 47320 single value features as potential predictors for cognition per subject. This includes both sides (mirrored triangles) of the correlation matrix, but excludes correlations of networks with themselves (always 1.0). To counter the bias of regularization in regression models towards larger absolute values, all features were standardized by subtracting their mean value over all included subjects and then dividing them by the respective standard deviation. To get a first instance if indicators may be suitable predictors for a linear regression model, their correlation with subjects' average cognition scores were visualized. Additionally, to check if any indicators vary significantly in their group means (MS vs HC) to act as potential biomarkers, independent t-tests ($\alpha = 0.05$) were conducted.

Cognitive score prediction. To detect the most important predictors for average cognition, Ridge regression was used to create predictive models that minimize the sum of the squared prediction error as well as the sum of the squared coefficient weights by L2 regularization.[26] This technique leads to better generalizability of the model for unseen data and its dominance can be adjusted by the hyperparameter α. An increasing α shifts the penalty from bias to variance, which is utilized to minimize overfitting compared to ordinary least squares (OLS) regression.[27] Furthermore, this allows to handle data with higher multicollinearity, because two identical features are assigned similar weights in the linear model.[28] Performance of ridge regression in similar problems is comparable to other methods while having relatively low computational costs and being easy to tune.[29]

Individualized prediction framework. To ensure that every subjects' cognitive scores are predicted exactly once for every tested condition, included subjects are divided into five sets randomly. Model generation and subsequent score prediction is repeated five times. For each of the repetitions, four sets of subjects are combined to create and tune the model and the remaining set is used to validate it afterwards. Therefore, 5-fold cross validation (5F-CV) is applied to an outer and inner loop, with the outer 5F-CV loop estimating the generalizability of the model and the inner 5F-CV loop determining

the optimal parameter α between [2^{-2}, 2^{-1}, …, 2^{10}]. A schematic overview of the prediction framework is shown in Figure 2.

Model comparison. To check if the introduced indicators add value to a predictive model, performance is compared to a classic network correlation-based baseline model (CORR indicators only). To quantify the overall model performances, the Pearson correlation coefficient between predicted cognitive scores and the true counterparts is calculated, as done in similar studies.[29] This is done for each validation fold, as they are independent models with similar conditions and then averaged ($r_{mean} \pm r_{std}$) and for all predicted values (r_{total}). The analysis is done separately for three groups (all subjects, HC and MS).

Finding the most robust indicators / predictors. There are various methods for creating a set of robust predictors for any given cognitive score, each coming with different advantages and disadvantages. Usually, all left predictors are made available for the model to be based on, but with the number of features greatly exceeding the number of examples, overfitting is hard to avoid. Another option is to eliminate the features from a full set recursively (α between [10^2, 10^3, …, 10^7]), which comes at a great computational cost and does not automatically lead to a small or ideal set of robust predictors. This utilizes knowledge about the outcome variable of the prediction in the model building process and thereby makes results less comparable with similar experiments. However, here it is used as having a predictive model is secondary to investigating a subset of robust predictors in order to interpret its properties. Lastly, the top ten features based on their importance order from the recursive feature elimination are used to generate a model with reasonable complexity to predict cognitive scores. Comparing (pre-)selected predictors and associated weights within a model for any cognitive domain or group of subjects may grant insight in what quantified mechanism by respective predictors is specific for a clinical phenotype.

Created models. The previously described methods are applied to all subjects, patients only and healthy controls only. For each group, a baseline model (CORR indicators only), a full model (including all introduced indicators), a minimal model (using recursive feature elimination) and a model containing exactly 10 features (based on their feature importance) were created. When the full model performance exceeds the baseline model performance, this indicates added predictive value by the introduced stock market indicators. The minimal models (after feature elimination) were used to quantify the robustness of selected features. The top ten model are examples to give an indication of predictive performance a reasonable complexity, desired for real-world applications.

Results

Indicators correlate with cognition scores. Visual inspection of the data shows that indicator descriptives and subjects average cognition scores are partially correlated, where examples are shown in Figure 3. While some network correlation mean values show higher correlations with the average cognition scores than other mean indicators, their kurtosis correlates much weaker than those of e.g. BB, CCI or RSI-based indicators. Similar variation and comparably high correlations are found for other descriptives as well as for cognitive scores from specific domains (data not shown). The gradient of correlation intensity within a single indicator-network-combination displays the variation over different time windows from short (left) to long (right). In some indicator-network-combinations, a clear trend is visible, with either shorter (blue box) or longer windows (green box) steadily increasing the found correlation with the outcome variable. Interestingly, these local maxima of correlation do not all converge at a single temporal scale, understating the importance and controversy of this

parameter. These found linear correlations are a first instance, that some indicators may be suitable predictors for a linear regression model.

Group difference between indicators. Although this study has a clear individual focus, looking at group differences can sometimes give valuable information on macroscopic mechanisms, especially when trying to find features which discriminate between affected and unaffected subjects (potential biomarkers). Feature-cognition correlation for HC and MS individually, shows very different patterns (data not shown). In general, the correlation between the indicators descriptives and average cognition is stronger within HC than in MS. This is quantified in Figure 4, showing the significance values of independent t-tests comparing group means for all indicator descriptives seperately. No significant differences between group means are found. Figure 5 shows a selection of indicator descriptives versus average cognition scores. Some features show similar correlations for both groups (e.g. A), while others show group specific trendlines (e.g. E). The highlighted indices of Figure 4, with the largest mean differences between HC (blue) and MS (red), are shown in Figure 5 F - I. Here it becomes visible that the distribution of cognition scores for MS patients are more widespread and generally lower compared to HC, which partially explains the finding that in the latter generally higher correlations occur. The highest correlations, positive and negative, between cognitive score and features for all subjects (r = 0.36 & -0.38), MS (r = 0.36 & -0.36) and HC (r = 0.53 & -0.61) are shown in A – E.

Robust features successfully predict cognitive scores. Figure 6 and Table 2 summarize the results of the predictive models. The baseline models with 18200 predictors shows lower performance than the respective models using all possible 47320 features for all subjects, MS and HC. No reliable baseline model for HC are generated under the tested conditions. The minimal models, established by recursive feature elimination, reduced the model complexity drastically while increasing predictive accuracy. Using the top ten predictors resulted in the best models compared in this study. Excluding the baseline, predicting cognitive scores of HC is more reliable than for MS or all subjects. Figure 7 visualizes two examples of the most important feature to predict all cognitive scores, the kurtosis of the RSI 90 based on the FPN II B activation signal, also found with the strongest correlation with the average cognition score (figure 5 A). The RSI 90 almost normalizes the signal based on the 90 most recent datapoints. The data for the HC subject (A - highest average cognition score) shows a more even feature value distribution than the MS subject (B - lowest average cognition score). However, due to two outliers in network activity of subject B the absolute covered value range is greater.

Discussion

When studying FC in the human brain, giving attention to temporal dynamics shows a promising approach to better understand cognition in HC and MS.[14,15] Common methodology is focused on network coactivation, eventually leading to FC states and their patterns of alternation. However, this does only indirectly explore activation patterns within single networks. It may be helpful to investigate such and potentially be able to link specific dynamic network activation pattern properties with cognitive disfunction, ideally with specific cognitive domains. Using the Amsterdam MS RS-fMRI dataset, the present study compared traditional network correlation-based features with new stock-index, single-network based features to predict average cognition scores. Results indicate that this translational approach can outperform baseline models on the same data. The found correlations of true and predicted cognitive scores are comparable to similar recent studies.[29]

Results for prediction are strongly dependent on the number of available datapoints. The large number of predictors increase the risk of overfitting to the respective train set. Both previous statements can

explain the poor prediction for only 54 HC subjects in the complete model. For minimal model complexity, the relatively higher scores for HC subjects can be explained by the more normally distributed cognitive abilities of this group. The patients with lower cognitive scores skew the cognition score distribution and evoke a bias in the machine learning algorithm towards the upper end of the spectrum, hence resulting in poorer predictive certainty on the less well represented affected subjects. One way to tackle this bias would be reducing the number of included subjects in such a way that present scores from training examples are always near equal distribution. However, this stands in direct conflict with the desire to increase the total number of included subjects to gain a more robust prediction and creates a tradeoff between these two biases.

The search for robust features via recursive feature elimination based on ridge regression with a fixed alpha value is suboptimal, because the necessity of regularization scales with the number of available predictors. For this task either a stepwise reduction of features with a predefined alpha value for each step, or a dynamic regularization parameter is desired, but both lie outside the scope of this project.

Interpretability of the predictors given most weight in the regression models is limited. This is due to the noisy data leading to large heterogeneity of affected cognitive domains throughout the subjects. The examples for the kurtosis of RSI 90 based on the FPN II b network in figure 7 is showing extreme values and is not representative for the entire cohort, as clearly visible in figure 5 A. Therefore, further analysis is desired to predict cognitive scores of all available domains on subsets of features grouped by the networks they are based on. This has the potential to tackle problems of suboptimal feature to sample ratio and establish a more direct link between specific network activity and domain of cognitive deficits.

Analyzing domain specific cognitive deficits and establishing robust features on the present dataset is difficult, because of the large heterogeneity of the clinical phenotypes present. Dividing the cohort into groups based on affected cognitive domain leaves not enough subjects for a robust analysis, and using the entire dataset induces a strong bias towards learning and predicting scores of the larger unaffected group for each domain respectively. A compromised approach could include subject classification instead of regression based on an arbitrary threshold in cognitive performance (typically at 2 z-scores below the overall mean). However, this drawback needs to be overcome to truly achieve robust prediction of complete individual cognitive profiles.

This study is primarily limited by the size of the available dataset and the high variance of cognitive scores across various cognitive domains. The fact that the dataset is based on RS-fMRI is unproblematic, as network activity found in the resting brain is also meaningful in tasks.[30] Optimal features for this cohort are to be validated on an independent dataset to estimate and minimize the effect of overfitting. Generally, the potential of this and similar projects is strongly limited by the size of available datasets. Ideally, large, open-source datasets are desirable to increase statistical power and comparability in benchmark tests. Recent advances in this direction, e.g. the OpenfMRI project (openfmri.org/dataset) or the UK Biobank (fmrib.ox.ac.uk/ukbiobank) will help to tackle these common problems of the field in the future.

Concluding result interpretation, a longitudinal study on larger, independent datasets, and utilizing features similar to the ones presented in this study are likely to enhance the understanding of the connection between temporal brain network activity patterns and clinical phenotypes in MS. It is not unlikely that a sophisticated methodology will help as a pre-diagnostic tool in clinical screenings to create cognitive profiles for individuals with risk for developing cognitive deficits, not limited to MS.

1. Chiaravalloti ND, DeLuca J. Cognitive impairment in multiple sclerosis. Elsevier; 2008;7:1139–1151.

2. Sumowski JF, Benedict R, Enzinger C, et al. Cognition in multiple sclerosis: State of the field and priorities for the future. Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2018;90:278–288.

3. Amato MP, Portaccio E, Goretti B, et al. Cognitive impairment in early stages of multiple sclerosis. Springer Milan; 2010;31:211–214.

4. Amato MP, Langdon D, Montalban X, et al. Treatment of cognitive impairment in multiple sclerosis: position paper. Springer-Verlag; 2013;260:1452–1468.

5. Rocca MA, Amato MP, De Stefano N, et al. Clinical and imaging assessment of cognitive dysfunction in multiple sclerosis. Elsevier; 2015;14:302–317.

6. Filippi M, Rocca MA, Barkhof F, et al. Association between pathological and MRI findings in multiple sclerosis. Elsevier; 2012;11:349–360.

7. Pinter D, Khalil M, Pichler A, et al. Predictive value of different conventional and non-conventional MRI-parameters for specific domains of cognitive function in multiple sclerosis. Elsevier; 2015;7:715–720.

8. Rossi F, Giorgio A, Battaglini M, et al. Relevance of Brain Lesion Location to Cognition in Relapsing Multiple Sclerosis. Oreja-Guevara C, editor. Public Library of Science; 2012;7:e44826.

9. Schoonheim MM, Meijer KA, Geurts JJG. Network Collapse and Cognitive Impairment in Multiple Sclerosis. 2015;6:1–5.

10. Eijlers AJC, Meijer KA, Wassenaar TM, et al. Increased default-mode network centrality in cognitively impaired multiple sclerosis patients. 2017;88:952–960.

11. Schoonheim MM, Geurts JJG, Landi D, et al. Functional connectivity changes in multiple sclerosis patients: A graph analytical study of MEG resting state data. Wiley-Blackwell; 2013;34:52–61.

12. Hindriks R, Adhikari MH, Murayama Y, et al. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? Academic Press; 2016;127:242–256.

13. Shakil S, Lee C-H, Keilholz SD. Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states. Academic Press; 2016;133:111–128.

14. Park HJ, Friston KJ, Pae C, Park B, Razi A. Dynamic effective connectivity in resting state fMRI. The Authors; Epub 2017.:1–15.

15. Hutchison RM, Womelsdorf T, Allen EA, et al. Dynamic functional connectivity: Promise, issues, and interpretations. Academic Press; 2013;80:360–378.

16. Tsay R. Analysis of financial time series [online]. Wiley; 2005.

17. Eijlers AJC, van Geest Q, Dekker I, et al. Predicting cognitive decline in multiple sclerosis: a 5-year follow-up study. Epub 2018 Jul 26.

18. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. 2011;69:292–302.

19. Rao SM. A manual for the Brief, Repeatable Battery of Neuropsychological Tests in Multiple Sclerosis National Multiple Sclerosis Society. 1991.

20. Schoonheim MM, Hulst HE, Brandt RB, et al. Thalamus structure and function determine severity of cognitive impairment in multiple sclerosis. Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2015;84:776–783.

21. Meijer KA, Eijlers AJC, Douw L, et al. Increased connectivity of hub networks and cognitive impairment in multiple sclerosis. 2017;88:2107–2114.

22. Amato MP, Portaccio E, Goretti B, et al. The Rao's Brief Repeatable Battery and Stroop test: normative values with

age, education and gender corrections in an Italian population. 2006;12:787–793.

23. Smith SM, Vidaurre D, Beckmann CF, et al. Functional connectomics from resting-state fMRI. Elsevier; 2013;17:666–682.

24. Griffanti L, Salimi-Khorshidi G, Beckmann CF, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. NIH Public Access; 2014;95:232–247.

25. Vidaurre D, Smith SM, Woolrich MW. Brain network dynamics are hierarchically organized in time. National Academy of Sciences; 2017;114:12827–12832.

26. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. 1970;12:55–67.

27. Zou H, Hastie T. Regularization and variable selection via the elastic net. Wiley/Blackwell (10.1111); 2005;67:301–320.

28. Vinod HD. A Survey of Ridge Regression and Related Techniques for Improvements over Ordinary Least Squares. The MIT Press; 1978;60:121.

29. Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. Academic Press; 2018;178:622–637.

30. Smith SM, Fox PT, Miller KL, et al. Correspondence of the brain's functional architecture during activation and rest. 2009;106:13040–13045.

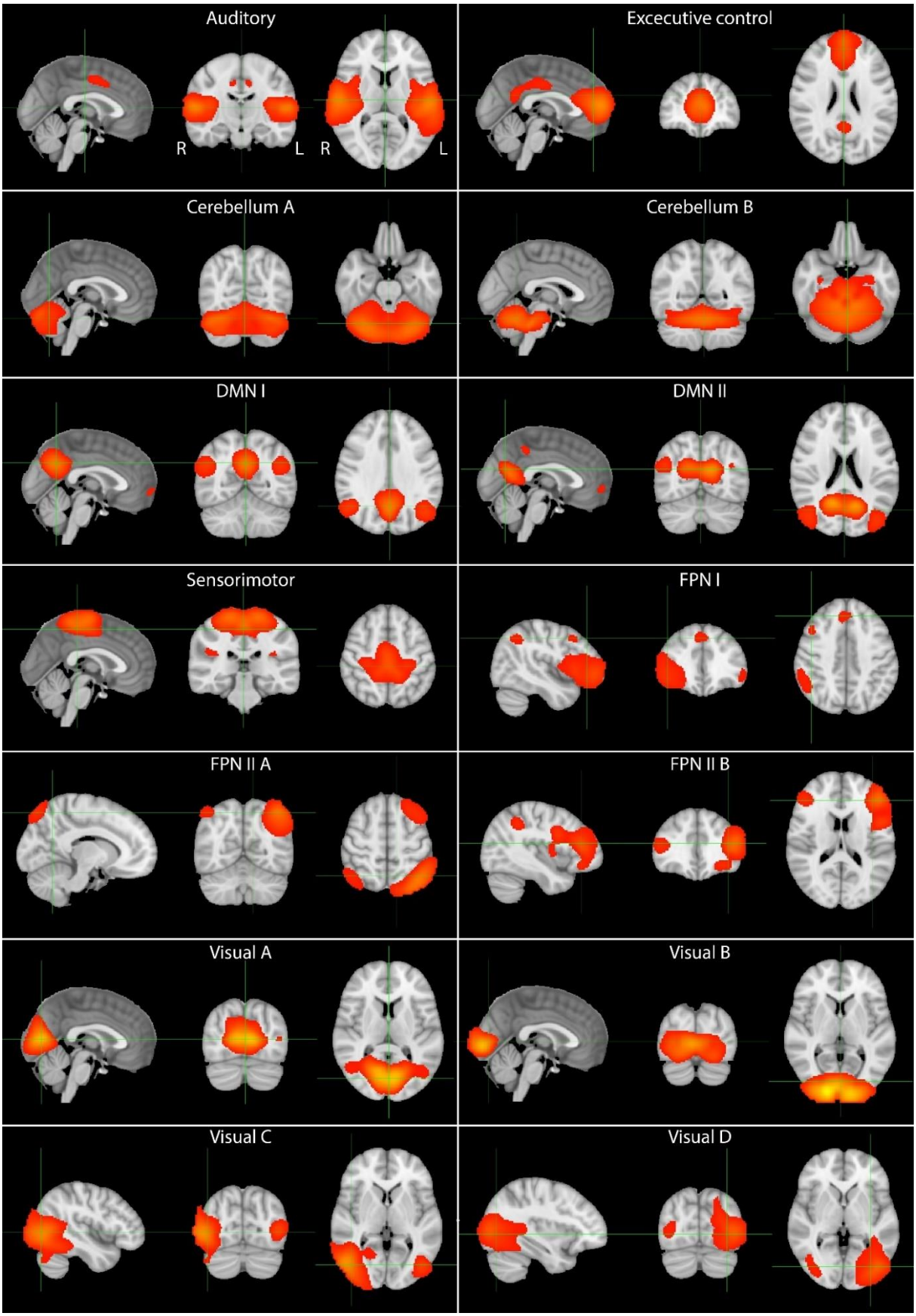

Figure 1: The 14 data-driven networks elicited by dual regression and melodic to provide their activation signals for this study.

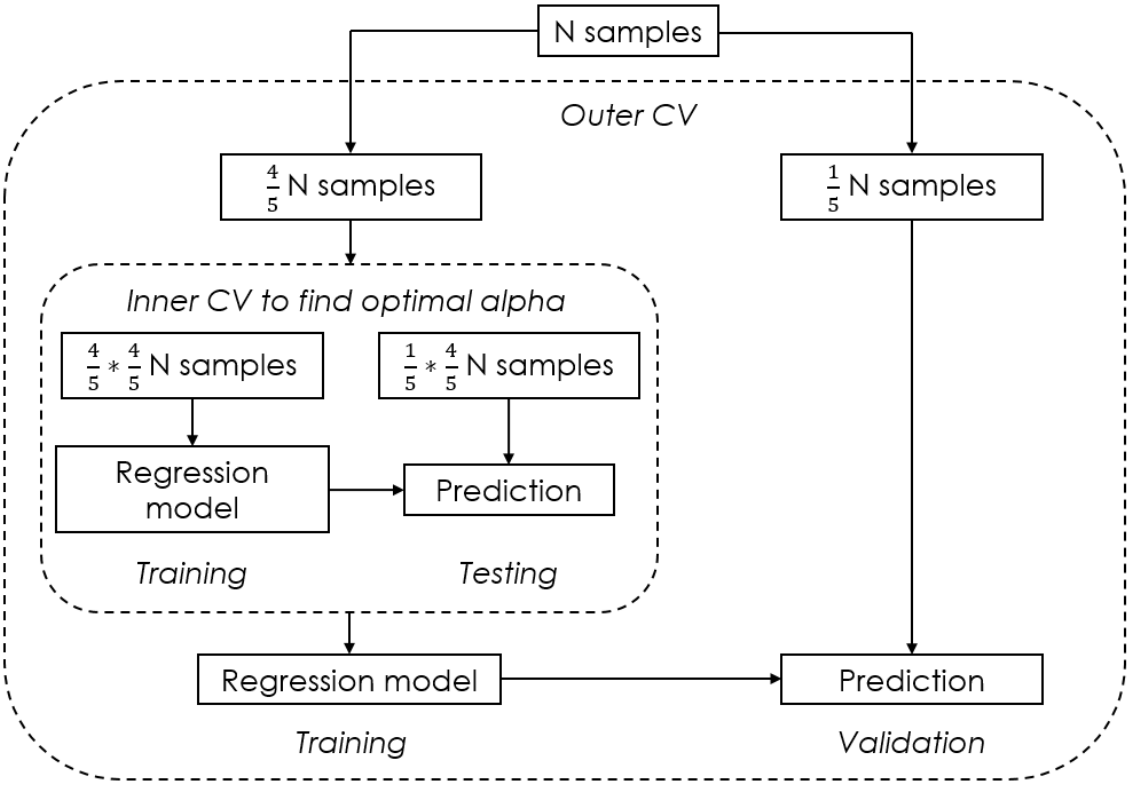

Figure 2: Schematic overview of the developed prediction framework for feature (sub-)set validation. The shown wrapper function consisting of an inner cross-validation, used to tune the hyperparameter alpha of the ridge regression algorithm, and an outer cross-validation to predict cognitive scores of the subjects and test the generalizability of the model. The process is repeated five times, so every subject has been part of the validation group exactly once.

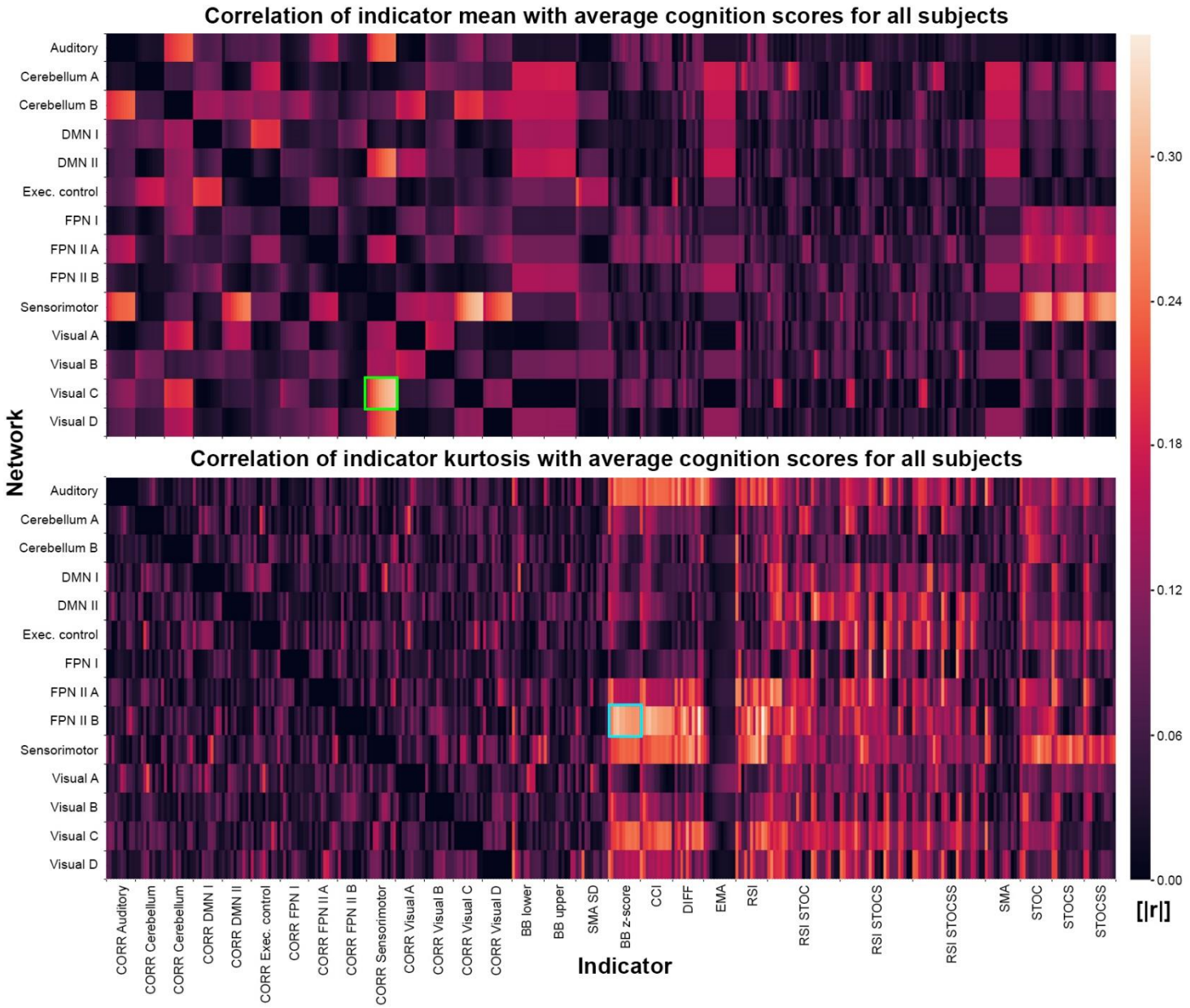

Figure 3: Absolute correlation of selected features of all subjects with their average cognition score. While some mean network correlations show higher correlations (Pearson's correlation coefficient) with the cognitive scores than other mean indicators, their kurtosis correlates much weaker than those of e.g. BB, CCI or RSI-based indicators. The gradient of correlation intensity within a single indicator-network-combination displays the variation over different time windows from short (left) to long (right). In some cases, a clear trend is visible, with either shorter (blue) or longer (green) windows steadily increasing the found correlation with the outcome variable.

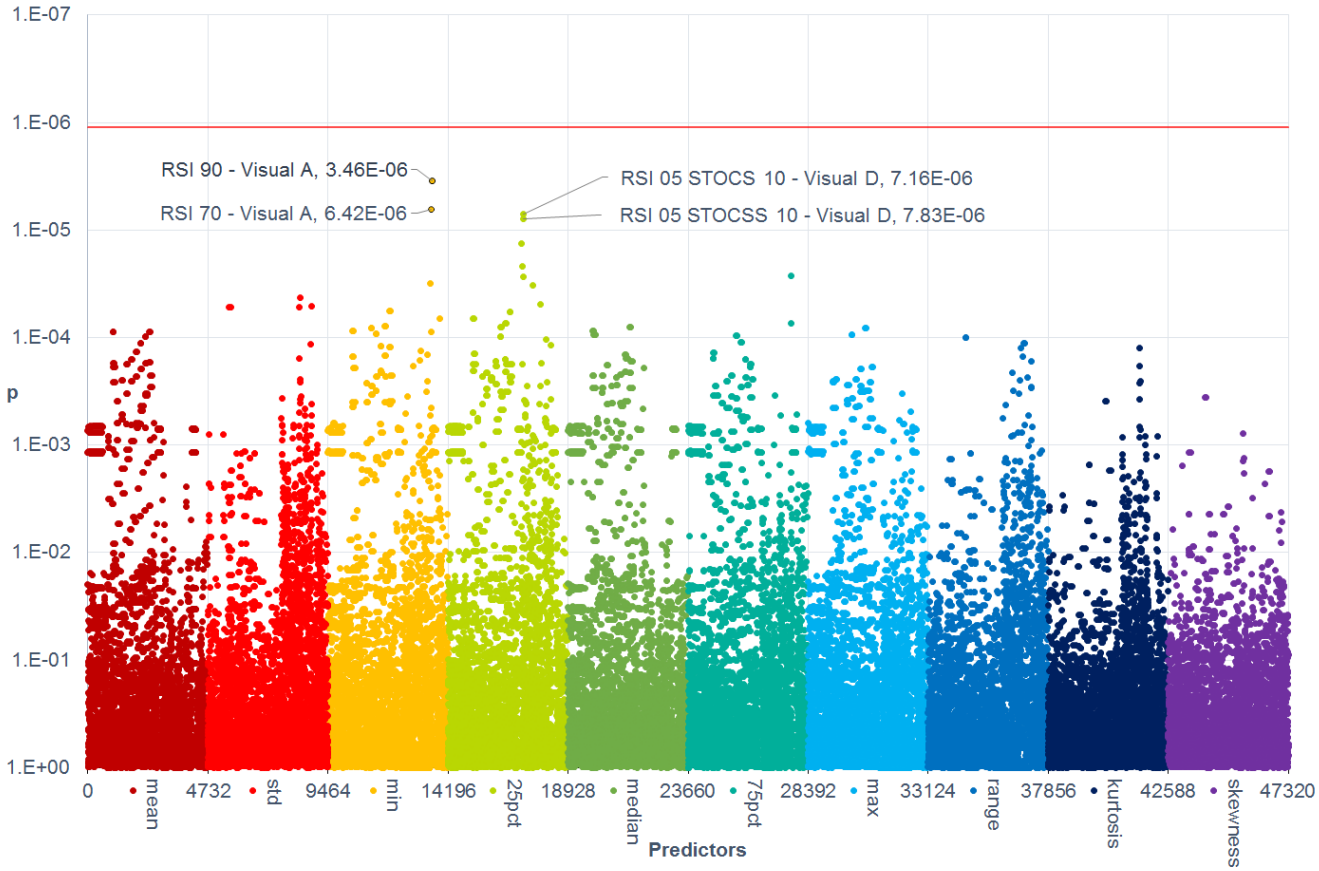

Figure 4: Manhattan plot for group differences of all 47320 predictors. Due to multiple comparisons, the significant p-value is set to 1.1E-06 by Bonferroni correction. No predictors show significantly different mean values between patients and healthy controls (independent t-test, alpha = 0.05), but four RSI-based candidates are highlighted. Such group differences are potential non-invasive biomarkers for MS and therefore ought to be investigated further.

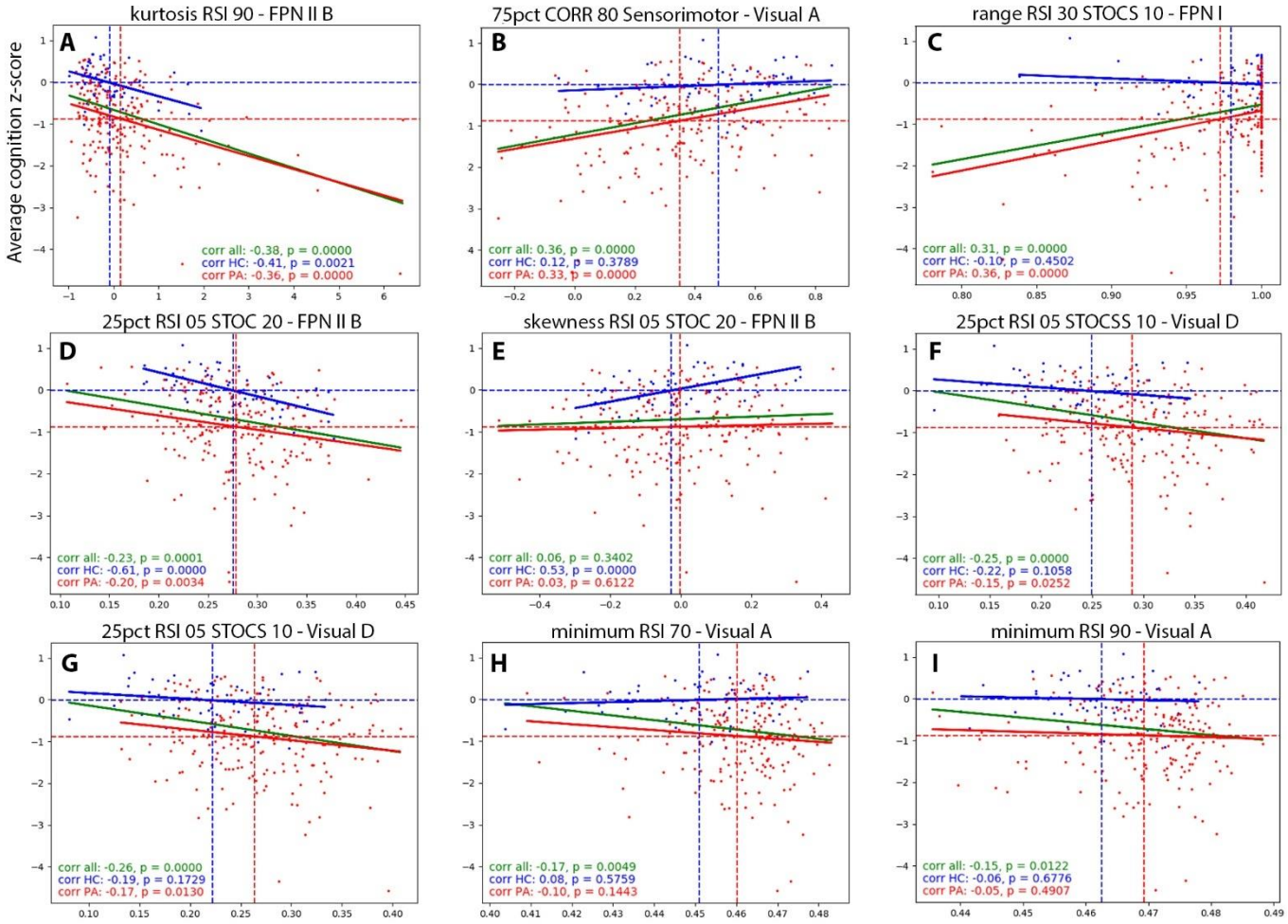

Figure 5: Scatterplots of average cognition over selected features. Only subjects with complete test score profiles are included (N = 267; 54 HC). HC: blue, MS: red, all: green, group means: dotted lines, correlations: solid lines. The mean cognitive score of the patients is approximately one standard deviation below the mean of HC. The highest correlations, positive and negative, between cognitive score and index descriptive (features) for all subjects (A & B), MS (A & C) and HC (D & E). Features based on bound indices such as the RSI show an accumulation of examples at extreme values, as shown in C. Some features show similar correlations for both groups (e.g. A), while other show group specific trendlines (e.g. E). The highlighted indices of figure 4 with the largest mean differences between HC (blue) and MS (red) are shown in F, G, H & I.

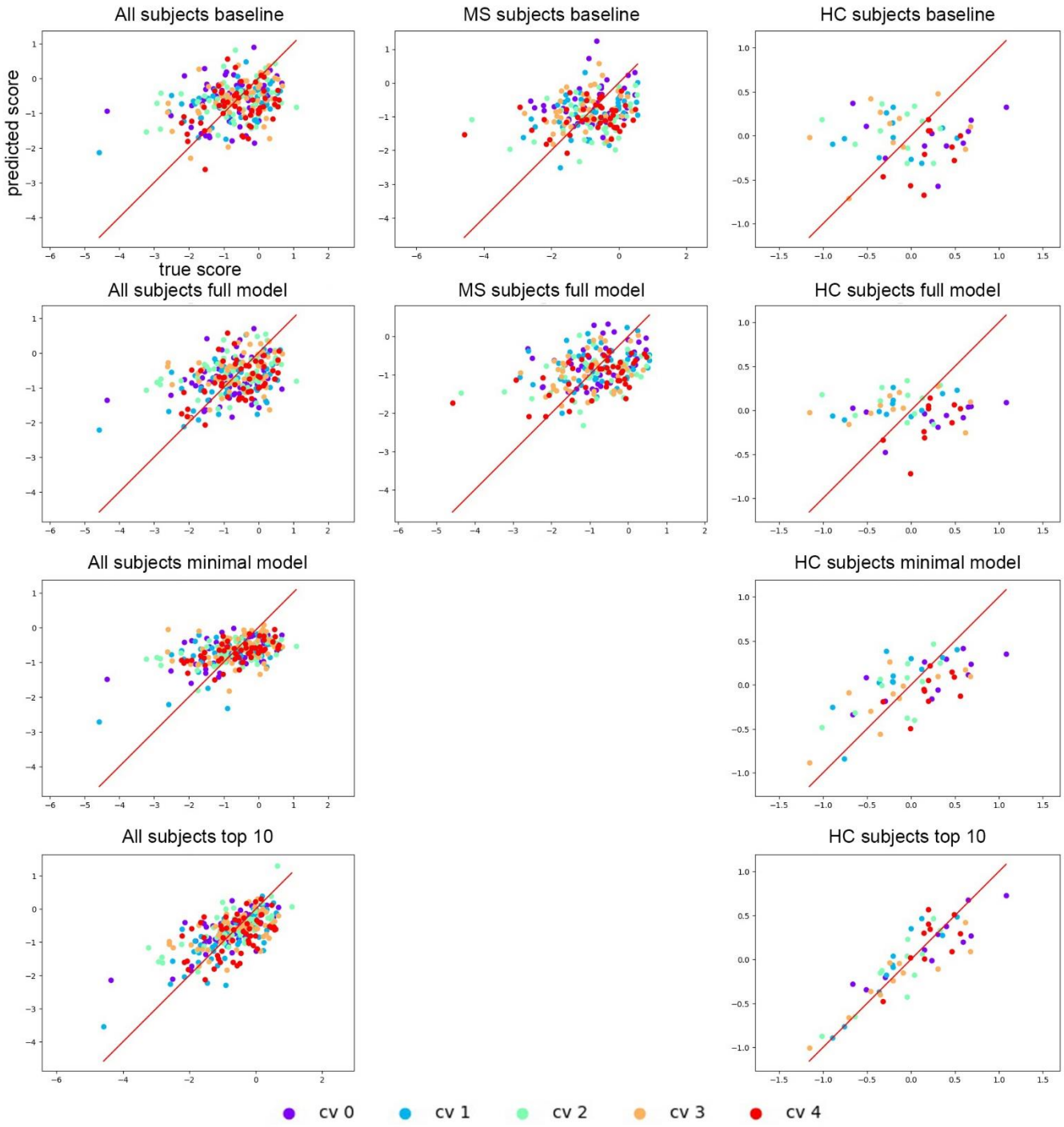

Figure 6: Predicted over true average cognition scores for all subjects, MS and HC. Predictive performance increases from top (baseline) to bottom (top 10 predictors from recursive feature elimination). Identity line in red.

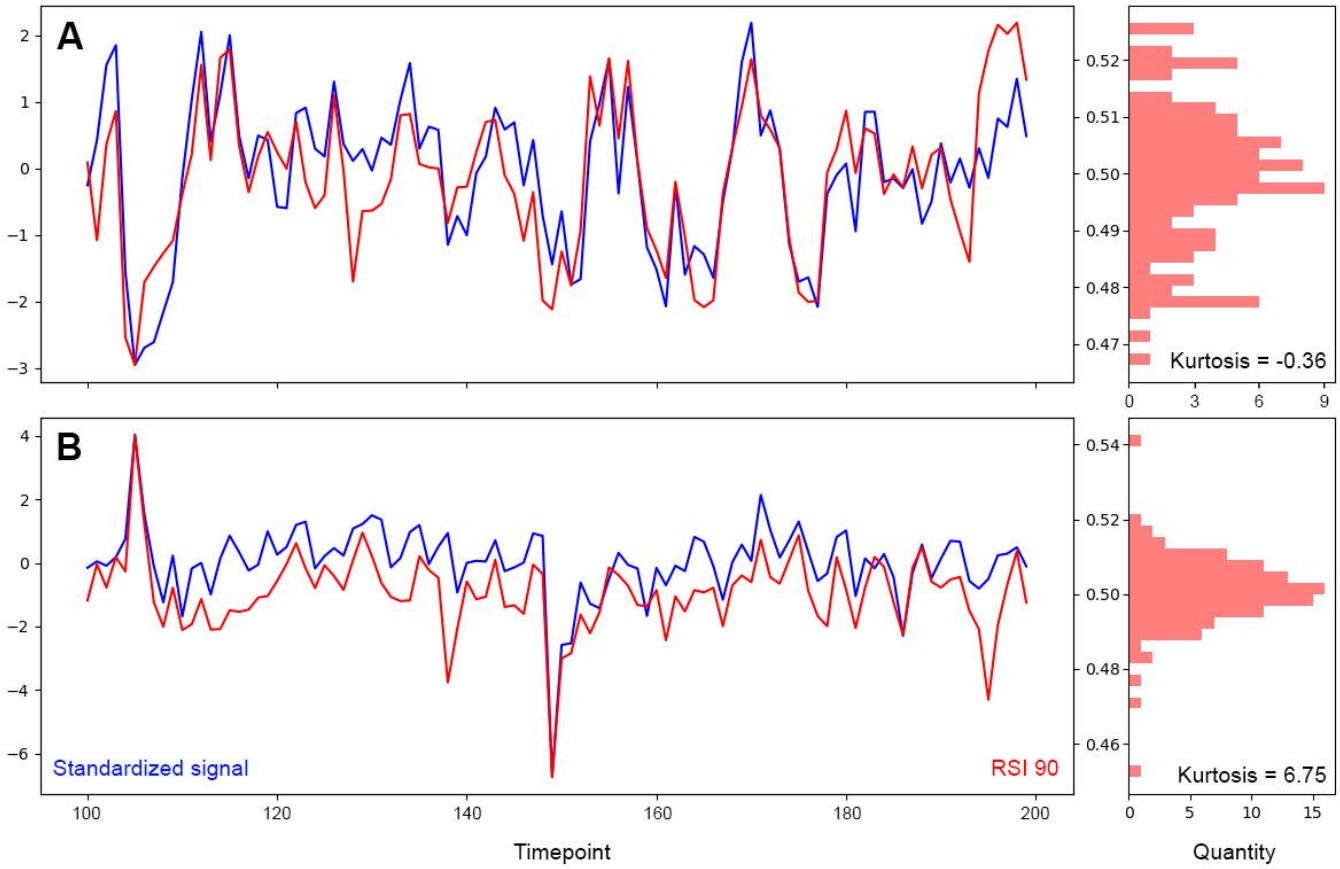

Figure 7: Examples of standardized signal and RSI 90 of FPN II b. A) HC subject with the highest average cognition score; B) MS subject with the lowest average cognition score. The RSI 90 follows the signal curve closely. Subject B shows two extreme signal values diverging strongly from the network activation mean. Histograms of the RSI (right) show a platykurtic value distribution for subject A and a leptokurtic distribution for subject B. Although the absolute range for RSI 90 is greater for subject B than for subject A, most values are concentrated around the mean value.

Indicator [abbreviation]	**Equation**	**Time windows** Δt (2.2 s each)
Signal $[S]$ - unbound	$S(t_i)$	
Pearson's correlation coefficient $[CORR]$ - bound	$CORR(t_i, \Delta t) = \frac{\Delta t \sum_{t_i-\Delta t}^{t_i} S(t_i)\, S_n(t_i) - [\sum_{t_i-\Delta t}^{t_i} S(t_i)][\sum_{t_i-\Delta t}^{t_i} S_n(t_i)]}{\sqrt{[\Delta t \sum_{t_i-\Delta t}^{t_i} S(t_i)^2 - (\sum_{t_i-\Delta t}^{t_i} S(t_i))^2][\Delta t \sum_{t_i-\Delta t}^{t_i} S_n(t_i)^2 - (\sum_{t_i-\Delta t}^{t_i} S_n(t_i))^2]}}$	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Simple moving average $[SMA]$ - unbound	$SMA(t_i, \Delta t) = \frac{1}{\Delta t} \sum_{t_i-\Delta t}^{t_i} S(t_i)$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
SMA standard deviation $[\sigma_{SMA}]$ - unbound	$\sigma_{SMA}(t_i, \Delta t) = \sqrt{\frac{\sum_{t_i-\Delta t}^{t_i} [\mathrm{SMA}(t_i, \Delta t) - \overline{\mathrm{SMA}}(t_i, \Delta t)]^2}{\Delta t}}$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
SMA of signal difference $[DIF]$ - unbound	$DIF(t_i, \Delta t) = \frac{1}{\Delta t} \sum_{t_i-\Delta t}^{t_i} S(t_i) - S(t_{i-1})$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Exponential moving average $[EMA]$ - unbound	$EMA(t_i, \Delta t) = \frac{2S_i(t)}{\Delta t + 1} + EMA(t_{i-1}, \Delta t)\, [1 - \frac{2}{\Delta t + 1}]$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Relative strength index $[RSI]$ – semi-bound	$RSI(t_i, \Delta t) = 100\ - \frac{100}{1\ + \frac{[\frac{1}{\Delta t}\sum_{t_i-\Delta t}^{t_i} \max(0, [S(t_i) - S(t_{i-1})])]}{[\frac{1}{\Delta t}\sum_{t_i-\Delta t}^{t_i} \min(0, [S(t_i) - S(t_{i-1})])]}}$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Commodity channel index $[CCI]$ – semi-bound	$CCI(t_i, \Delta t) = \frac{\mathrm{S}(t_i) - \ SMA(t_i, \Delta t)}{0.015\ \sigma_{SMA}(t_i, \Delta t)}$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Bollinger bands upper border $[BB_{upper}]$ - unbound	$BB_{upper}(t_i, \Delta t) = SMA(t_i, \Delta t) + 2\ \sigma_{SMA}(t_i, \Delta t)$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Bollinger bands lower border $[BB_{lower}]$ - unbound	$BB_{lower}(t_i, \Delta t) = SMA(t_i, \Delta t) - 2\ \sigma_{SMA}(t_i, \Delta t)$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Bollinger bands z-score $[BB_{z-score}]$ – semi-bound	$BB_{z-score}(t_i, \Delta t) = \frac{S(t_i) - BB_{lower}(t_i, \Delta t)}{BB_{upper}(t_i, \Delta t) - BB_{lower}(t_i, \Delta t)}$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Stochastic oscillator $[STOC]$ – semi-bound	$STOC(t_i, \Delta t) = \frac{S(t_i) - \min(S(t_i - \Delta t), \dots, S(t_i))}{\max(S(t_{i-\Delta t}), \dots, S(t_i)) - \min(S(t_{i-\Delta t}), \dots, S(t_{i-\Delta t}))}$ $STOC_s(t_i, \Delta t) = \frac{1}{3}\sum_{t_{i-3}}^{t_i} STOC(t_i, \Delta t)$ $STOC_{ss}(t_i, \Delta t) = \frac{1}{3}\sum_{t_{i-3}}^{t_i} STOC_s(t_i, \Delta t)$	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100

Stochastic oscillator of RSI [$RSI\ STOC$] – semi-bound	RSI: 5, 10, 20, 30, 40 & STOC: 5, 10, 20, 30, 40

$$RSI\ STOC(t_i, \Delta t, \Delta t_{rsi}) = \frac{RSI(t_i) - \min(RSI(t_{i-\Delta t}, \Delta t_{rsi}), \dots, RSI(t_i, \Delta t_{rsi}))}{\max(RSI(t_{i-\Delta t}, \Delta t_{rsi}), \dots, RSI(t_i, \Delta t_{rsi})) - \min(RSI(t_{i-\Delta t}, \Delta t_{rsi}), \dots, RSI(t_{i-\Delta t}, \Delta t_{rsi}))}$$

$$RSI\ STOC_s(t_i, \Delta t, \Delta t_{rsi}) = \frac{1}{3}\sum_{t_{i-3}}^{t_i} RSI\ STOC(t_i, \Delta t, \Delta t_{rsi})$$

$$RSI\ STOC_{ss}(t_i, \Delta t, \Delta t_{rsi}) = \frac{1}{3}\sum_{t_{i-3}}^{t_i} RSI\ STOC_s(t_i, \Delta t, \Delta t_{rsi})$$

Table 1: Overview of used predictors and time windows.

All predictors		All subjects (n = 267)	MS subjects (n = 213)	HC subjects (n = 54)
Full model baseline	r_{mean}	0.326 ± 0.093	0.253 ± 0.044	0.087 ± 0.228
	r_{total}	0.309 (p = 2.55e-07)	0.239 (p = 4.25e-04)	-0.019 (p = 8.90e-01)
	n_{feat}	18200	18200	18200
Full model All predictors	r_{mean}	0.358 ± 0.084	0.379 ± 0.095	0.325 ± 0.268
	r_{total}	0.363 (p = 9.39e-10)	0.369 (p = 2.94e-08)	0.054 (p = 7.00e-01)
	n_{feat}	47320	47320	47320
Minimal model	r_{mean}	0.459 ± 0.139		0.681 ± 0.099
	r_{total}	0.459 (p = 2.39e-15)		0.645 (p = 1.40e-07)
	n_{feat}	2		2
Top 10 predictor model	r_{mean}	0.652 ± 0.119		0.872 ± 0.097
	r_{total}	0.657 (p = 2.59e-34)		0.883 (p = 1.07e-18)
	n_{feat}	10		10

Table 2: Prediction results for all subjects, MS and HC. The baseline model based on classical functional connectivity measures generally performs weakest and no reliable model for HC is generated. Results from the minimal model show that two predictors are sufficient to predict cognitive scores reliably. This performance is further improved when the top ten predictors from recursive feature elimination are used.