Novel composite MRI scale correlates highly with disability in multiple sclerosis patients

Authors: Peter Kosa, Mika Komori, Ryan Waters, Tianxia Wu, Irene Cortese, Joan Ohayon, Kaylan Fenton, Jamie Cherup, Tomas Gedeon, and Bibiana Bielekova

Kosa, Peter, Mika Komori, Ryan Waters, Tianxia Wu, Irene Cortese, Joan Ohayon, Kaylan Fenton, Jamie Cherup, Tomas Gedeon, and Bibiana Bielekova. "Novel composite MRI scale correlates highly with disability in multiple sclerosis patients." Multiple Sclerosis & Related Disorders 4, no. 6 (November 2015): 526-535. DOI: 10.1016/j.msard.2015.08.009.

Novel composite MRI scale correlates highly with disability in multiple sclerosis patients

Peter Kosa [a], Mika Komori [a], Ryan Waters [b], Tianxia Wu [c], Irene Cortese [d], Joan Ohayon [d], Kaylan Fenton [d], Jamie Cherup [a], Tomas Gedeon [b], Bibiana Bielekova [a,n]

[a] Neuroimmunological Diseases Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
[b] Department of Mathematical Sciences, Montana State University, Bozeman, MT, USA
[c] Clinical Neuroscience Program, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
[d] Neuroimmunology Clinic, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

E-mail addresses: kosap@mail.nih.gov (P. Kosa), mika.komori2@nih.gov (M. Komori), waters@math.montana.edu (R. Waters), wuti@mail.nih.gov (T. Wu), corteseir@ninds.nih.gov (I. Cortese), ohayonj@ninds.nih.gov (J. Ohayon), kaylan.fenton@gmail.com (K. Fenton), jamie.cherup@nih.gov (J. Cherup), Gedeon@math.montana.edu (T. Gedeon), Bibi.Bielekova@nih.gov (B. Bielekova).

Abstract

Understanding genotype–phenotype relationships or development/validation of biomarkers requires large multicenter cohorts integrated by universal quantification of crucial phenotypical traits, such as central nervous system (CNS) tissue destruction. We hypothesized that mathematical modeling-guided combination of biologically meaningful, semi-quantitative MRI elements characterized by high signal-to-noise ratio will provide such reliable, universal tool for measuring CNS tissue destruction. We retrospectively graded 15 elements in MRI scans performed in 419 untreated subjects with or without neurological diseases, while being blinded to their prospectively acquired clinical scores. We then used 305 subjects for disability-guided mathematical modeling to select and combine MRI elements that had non-redundant contributions to clinical disability, resulting in Combinatorial MRI Scale (COMRIS). We validated our model on the remaining 114 independent subjects. COMRIS requires 5–10 min per scan on average to compute and demonstrates highly significant (po0.0001) and validation-consistent Spear-man correlation coefficients (0.75, 0.76, and 0.65) for the expanded disability status scale (EDSS), Scripps neurological rating scale (SNRS), and symbol digit modality test (SDMT) measures of neurological dis-ability, respectively. Because COMRIS is not greatly influenced by MRI scanners or protocols and can be computed even in the presence of some motion artifacts, it does not require censoring out patients and it provides comparable results across different cohorts. As such, it represents a broadly available clinical and research tool that can facilitate multicenter research studies and comparative analyses across patient cohorts and research projects.

Abbreviations: COMRIS-CDO-SDMT, combinatorial MRI scale-cognitive disability optimized by SDMT; COMRIS-CTD-v1, combinatorial MRI scale of CNS tissue de-struction version 1; COMRIS-PDO-EDSS, combinatorial MRI scale-physical disability optimized by EDSS; COMRIS-PDO-SNRS, combinatorial MRI scale-physical disability optimized by SNRS; CIS, clinically isolated syndrome; EDSS, expanded disability status scale; HD, healthy donor; MS, multiple sclerosis; NIND, non-inflammatory neurological disorders; OIND, other inflammatory neurological disorders; PPMS, primary progressive multiple sclerosis; RIS, radiologically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SD, standard deviation; SDMT, symbol digit modalities test; SNRS, Scripps neurological rating scale; SPMS, sec-ondary progressive multiple sclerosis

1. Introduction

Translational research on diseases of the central nervous system (CNS) would be greatly facilitated by the development of a biomarker for CNS tissue destruction that can be utilized across diverse research groups. For example, while disease status itself was sufficient for integrating multinational cohorts in genome-wide association studies, genotype–phenotype analyses or predictive biomarker studies, which also require thousands of patients, cannot be performed without universal measures of disease severity. Clinical scales have played a unifying role in multicenter clinical trials; however, they are not ideal for aforementioned translational research applications for several reasons: 1. Each scale measures only some aspect of clinical disability; 2. Multiple clinical scales are rarely documented in broad clinical practice; and, most importantly, 3. Clinical disability is greatly influenced by the localization of a lesion; so a single strategically located lesion (e.g. in the internal capsule) can have devastating clinical outcome,

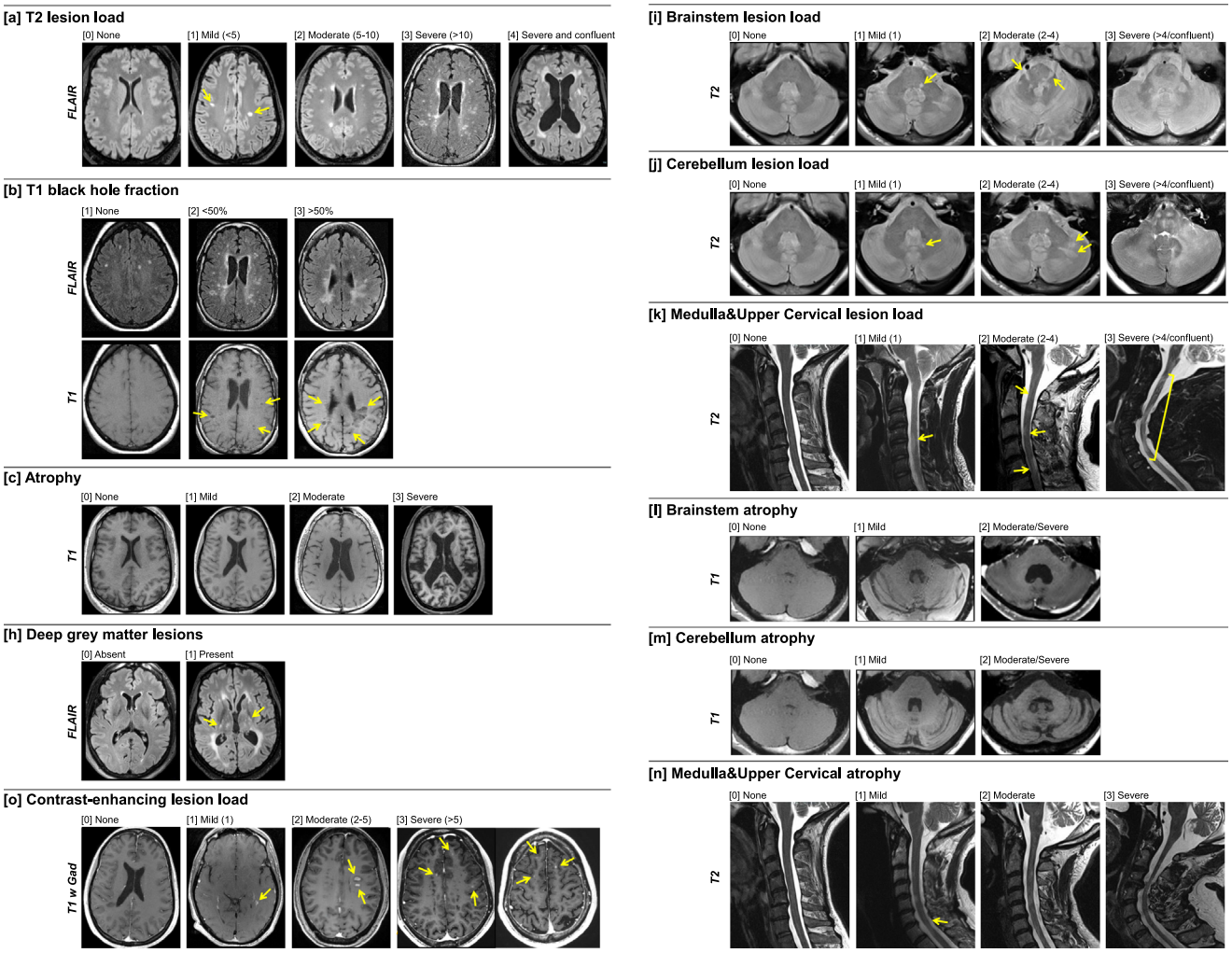

Fig. 1. MRI grading procedure. Representative MRI images demonstrate different grades for fifteen MRI variables (a–o). T2LL [a] was defined as number of lesions (larger than 3 mm in diameter) showing high signal on T2WI or FLAIR images in the supratentorial brain. The number of T2LL was categorized as: 0=None, 1=1–4 lesions, 2=5–10 lesions, 3=more than 10 lesions, and 4=more than 10 and confluent lesions. BHFr [b] was defined as the percentage of T2 lesions that have appearance of black holes on T1WI (i.e. show comparable intensity as CSF). 1=None (no black holes present), 2=less than 50% of lesions are black holes, and 3=50% or more lesions are black holes. To assess cerebral atrophy [c], we considered both ventricular size and width of cortical sulci, best appreciated on axial cuts through insular cortex or vertex. Presented examples focus on difference in ventricular size. 0=None (no visible atrophy of ventricles or sulci), 1=Mild (mild increase in ventricular size or visible atrophy, but often only focal increase in sulci width), 2=Moderate (considerable enlargement of lateral ventricles, definite enlargement of the third ventricle or more wide-spread broadening of the sulci) and 3=Severe (concomitant severe atrophy of ventricles and sulci that leads to diffuse reduction of visible brain tissue). Presence of deep gray matter lesions (GML) [h] was evaluated in thalamus, lenticulate nuclei and caudate nuclei (both head of the cause and tail) by using T2WI/FLAIR as well as T1WI/MP-RAGE, although only FLAIR images are depicted. 0=Absent, 1=Present. Contrast-enhancing lesions (CEL; [o]) were identified on post-contrast T1WI in reference to pre-contrast T1- and T2WI/FLAIR and quantified as 0=None, 0.5=1 CEL, 0.75=2–5 CELs, 1=more than 5 CELs. Evaluation of infratentorial lesion load for brainstem [i], cerebellum [j], and medulla and upper cervical spinal cord [k] were defined on T2WI/proton density images and T2WI/STIR images (for CS lesions). Lesions were verified on T1WI/MP-RAGE. 0=None, 1=1 lesion, 2=2−4 lesions, 3=more than 4 lesions or large/confluent lesions. Level of atrophy of brainstem [l] and cerebellum [m], were evaluated on axial and sagittal of T1WI/MP-RAGE images. For brainstem rating: 0=None, 1=Mild (mild flattening of the pons with visible enlargement of the 4th ventricle or widening of the interpeduncular cistern with mild flattening of cerebral peduncles), 2=Moderate/Severe (severe atrophy of the mesencephalon with small cerebral peduncles and large interpeduncular cistern and definite flattening of pons with enlarged cisterna pontis). For cerebellar rating: 0=None, 1=Mild (mild widening of cerebellar sulci and narrowing of middle cerebellar peduncles), 2=Moderate/Severe (clear widening of cerebellar sulci, prominent atrophy of the 4th ventricle and middle cerebellar peduncles). Atrophy of the medulla and upper cervical spinal cord [n] was evaluated on sagittal and axial T2WI and T1WI/MP-RAGE and rated as 0=None, 1=Mild (segmental atrophy at a single level) 2=Moderate (mild diffuse or 1–3 focal atrophy segments visible by naked eye, but with $>50\%$ of thickness preserved compared to non-affected area), 3=Severe (diffusely atrophied spinal cord or >3 focal atrophy segments or $<50\%$ of SC thickness preserved). Abbreviations: FLAIR=fluid-attenuation inversion recovery, MP-RAGE=Magnetization Prepared Rapid Gradient Echo, STIR=Short TI Inversion Recovery, T1WI=T1-weighted image, T2WI=T2-weighted image, T1 w Gad=T1-weighted image after IV injection of contrast agent, yellow arrows=examples of lesions or focal atrophy.

while many lesions, equally destructive, when located in subcortical white matter may have little impact on disability. In contrast to localization-dependency of clinical outcomes, a biomarker of CNS tissue destruction should reflect damage to CNS cells irrespective of their localization. Consequently, this undesired feature of clinical scales makes alternative measures of CNS tissue destruction, especially those obtained by magnetic resonance imaging (MRI) highly attractive. Indeed, MRI of the brain is routinely performed by neurologists caring for patients with chronic CNS diseases such as multiple sclerosis (MS), making quantitative MRI (qMRI) measures of CNS tissue destruction, such as brain atrophy, favorite candidate(s). Generation of qMRI data of high quality is time- and skill-demanding process that often requires filtering out considerable proportion of scans due to poor technical quality. Additionally, qMRI data are critically influenced by hardware, sequence parameters and post-processing methods. Consequently, data generated by different groups are not readily comparable (Chard et al., 2002; Jones et al., 2013).

Table 1
Semi-quantitative rating scores of MRIs and coefficients for linear equations determined by mathematical modeling.

MRI category	Verbal grade	Numerical code	Mathematically-optimized COMRIS models (coefficient "weights" for linear equation)		
			COMRIS-PDO-EDSS	*COMRIS-PDO-SNRS*	*COMRIS-CDO-SDMT*
[a] *T2 lesion load (T2LL)*	None Mild (< 5 small lesions) Moderate (5–10 lesions) Severe (> 10 lesions) Severe and confluent	0 1 2 3 4	0	0	**−1.94**
[b] *T1 black hole fraction*	None < 50% > 50%	1 2 3	0	0	0
[c] *Atrophy*	None Mild Moderate Severe	0 1 2 3	**0.22**	**−1.95**	**−3.83**
[d] *Periventricular (PV) lesions*	Absent Present	0 1	0	0	0
[e] *Juxtacortical/cortical (JC) lesions*	Absent Present	0 1	0	0	0
[f] *Corpus Callosum (CC) involvement*	Absent Present	0 1	0	0	0
[g] *Mostly deep white matter (WM) lesions*	Absent Present	0 1	0	0	0
[h] *Deep grey matter (GM) lesions*	Absent Present	0 1	0	0	**−1.34**
[i] *Brainstem lesion load (LL)*	None Mild (1 lesion) Moderate (2–4 lesions) Severe (> 4 lesions/ confluent)	0 1 2 3	**0.39**	**−4.29**	**−2.31**
[j] *Cerebellum lesion load*	None Mild (1 lesion) Moderate (2–4 lesions) Severe (> 4 lesions/ confluent)	0 1 2 3	0	0	**−2.18**
[k] *Medulla and Upper Cervical lesion load*	None Mild (1 lesion) Moderate (2–4 lesions) Severe (> 4 lesions/ confluent)	0 1 2 3	**0.84**	**−5.29**	0
[l] *Brainstem atrophy*	None Mild Moderate/Severe	0 1 2	0	0	0
[m] *Cerebellum atrophy*	None Mild Moderate/Severe	0 1 2	**0.32**	**−1.27**	0
[n] *Medulla and Upper Cervical atrophy*	None Mild Moderate Severe	0 1 2 3	**0.69**	**−5.71**	0
[o] *Contrast-enhancing lesion load*	None	0	0	0	0

Table 1 (*continued*)

MRI category	Verbal grade	Numerical code	Mathematically-optimized COMRIS models (coefficient "weights" for linear equation)		
			COMRIS-PDO-EDSS	*COMRIS-PDO-SNRS*	*COMRIS-CDO-SDMT*
	Mild (1 small lesion)	0.5			
	Moderate (1–5 lesions)	0.75			
	Severe (> 5 lesions)	1			
Age			**0.04**	**−0.28**	**−0.26**
Intersection at [0, y]			**−0.54**	**107.29**	**71.08**

The table summarizes grading scores (verbal and numerical) of 15 MRI scores (a–o) utilized for calculation of different COMRIS scores and coefficients for linear equation for variables selected by mathematical modeling using three different models (EDSS model, SNRS model and SDMT model). EDSS = expanded disability status scale, SNRS = Scripps neurological rating scale, SDMT = symbol digit modalities test, COMRIS-PDO-EDSS = combinatorial MRI scale – physical disability optimized by EDSS, COMRIS-PDO-SNRS = combinatorial MRI scale – physical disability optimized by SNRS, COMRIS-CDO-SDMT = combinatorial MRI scale – cognitive disability optimized by SDMT.

We hypothesized that it may be possible to devise a more universal (i.e. applicable across different hardware/sequence combinations), reproducible and clinically relevant scale of CNS tissue destruction based on disability-guided combination of multiple semi-quantitative assessments of conventional clinical brain MRIs. We report the construction of such an MRI scale that we named Combinatorial MRI Scale (COMRIS) and its optimization/validation process.

2. Material and methods

2.1. Participants

The study was approved by the National Institutes of Health Institutional Review Board and all participants provided written informed consent. Two patient cohorts (Supplemental Table 1) were prospectively acquired between 5/2007 and 9/2014. Diagnostic work-up included MRI of the brain and upper cervical spinal cord (SC). The diagnoses of relapsing-remitting MS (RRMS), primary progressive MS (PPMS) and secondary progressive MS (SPMS) were based on published diagnostic criteria (Polman et al., 2011). Remaining subjects were classified as either other inflammatory neurological disorders (OIND) or non-inflammatory neurological disorders (NIND) based on the evidence of intrathecal inflammation. Subjects were not treated by disease-modifying therapies.

2.2. Clinical scales

The following clinical scales were prospectively collected. The expanded disability status scale (EDSS) (Kurtzke, 1983), which represents the "gold standard" in MS. The more universal Scripps neurological rating scale (SNRS) (Sipe et al., 1984), which offers a greater range of scores (100 compared to 20 for EDSS) with more linear behavior. Symbol digit modalities test (SDMT), which has been proposed as a reliable scale of cognitive impairment in MS patients (Parmenter et al., 2007).

2.3. MRI acquisition

MRIs were performed on 1.5 T and 3 T Signa units (General Electric, Milwaukee, WI) and 3 T Skyra (Siemens, Malvern PA) equipped with standard clinical head and spinal cord imaging coils. MRI pulse sequences used for grading included FLAIR, PD/T2 weighted images, pre- and post-contrast (gadopentetate dimeglumine at 0.1 mmol/kg) T1-weighted image for brain scans, T1-weighted scan in the sagittal plane, T2- weighted scan in both sagittal and axial planes, and a short-tau inversion recovery scan in the sagittal plane for the cervical SC. Sequence details are in the Supplemental Methods.

2.4. Grading of MRI scans

Scoring of telencephalon focused on T2 lesion load (T2LL; [a]), T1 black hole fraction (BHFr; [b]; a proportion of T2 lesions that have appearance of black holes on post-contrast T1-weighted image), cerebral atrophy [c] and qualitative identification of periventricular (PV) lesions [d], juxtacortical/cortical (JC) lesions [e], involvement of corpus callosum (CC) [f], predominance of deep white matter (WM) lesions [g], and presence of deep gray matter (GM) lesions [h]. Evaluation of infratentorium comprised grading of lesion load (LL) and level of atrophy for brainstem [i,l], cerebellum [j,m], and medulla and upper cervical SC [k,n]. Contrast-enhancing lesions (CELs; [o]) were quantified in reference to pre-contrast T1- and T2-weighted scans (more details in Fig. 1 legend).

2.5. Mathematical modeling

Optimization of the combinatorial scale was performed by the least square methodology described in detail in the Supplemental methods.

2.6. Statistical analyses

For variables that did not follow the normal distribution, Box-Cox transformation was applied. To evaluate the association between different MRI/clinical measures and diagnostic groups, one-way analysis of variance (ANOVA) was performed, with Tukey's correction for pair-wise multiple comparisons, using SAS software version 9.2. The correlation between different variables was evaluated by Spearman correlation methods (GraphPad Prism 6). Due to multiple-comparison, we considered $p < 0.01$ as statistically significant.

3. Results

3.1. Construction of COMRIS

First, we selected those aspects of MRI imaging that reflect CNS tissue destruction based on publication knowledge (Bielekova et al., 2005; Bieniek et al., 2006; Filippi et al., 1995; Fisher et al., 2000; Fisniku et al., 2008; Truyen et al., 1996; Zivadinov and Bakshi, 2004) (Table 1). Because we expected that lesion location will have decisive influence on disability, we rated lesion load and atrophy separately for four structures that may differ in their involvement between individual patients: 1. Supratentorial brain; 2. Brainstem (pons, mesencephalon); 3. Cerebellum (including superior, middle and inferior cerebellar peduncles); and 4. Medulla

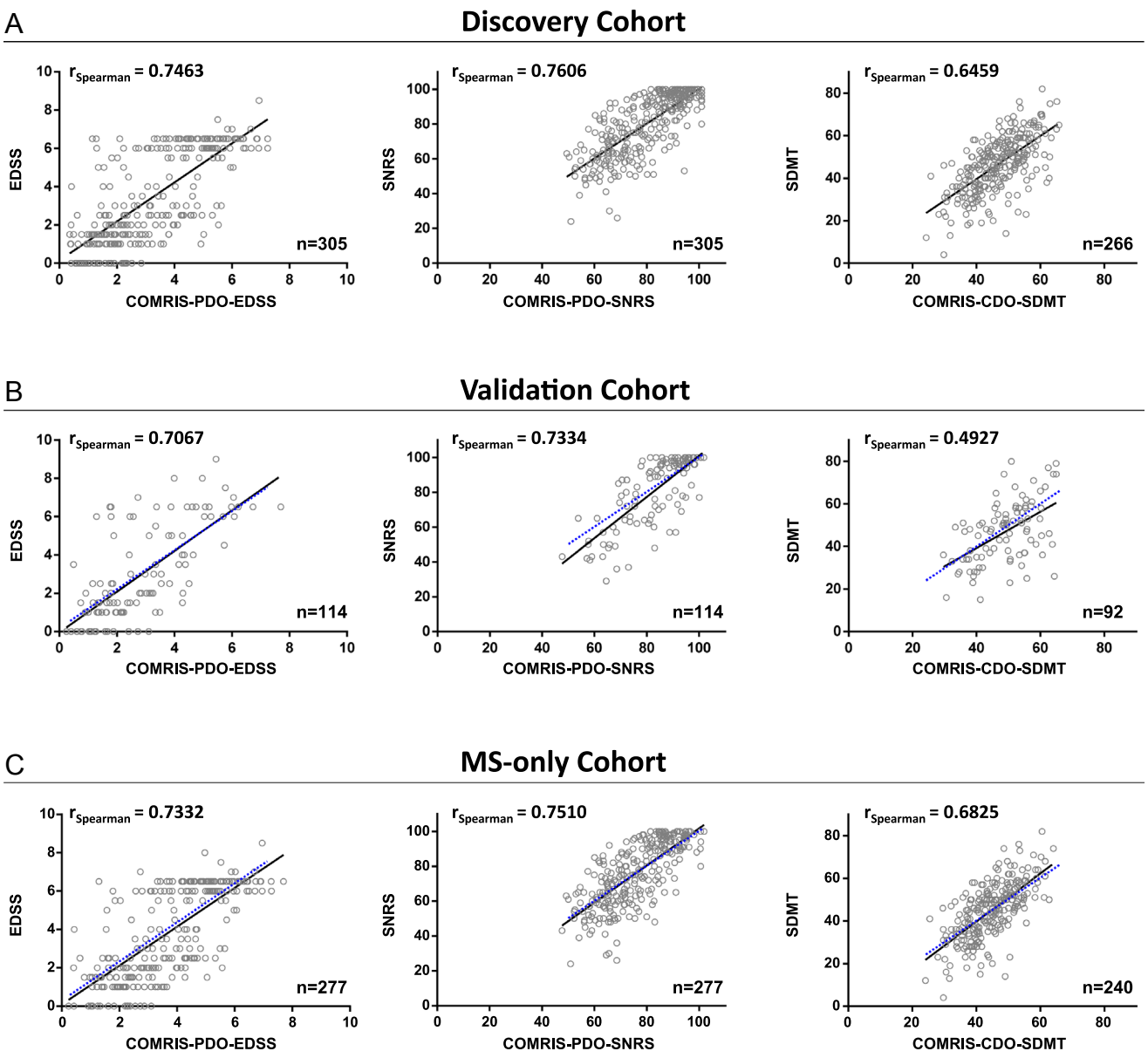

Fig. 2. Correlations of mathematically optimized COMRIS scales with clinical disability in different patient cohorts. Correlation of COMRIS-PDO-EDSS with clinically measured EDSS, COMRIS-PDO-SNRS with clinically measured SNRS, and COMRIS-CDO-SDMT with clinically measured SDMT in the (A) discovery cohort, (B) validation cohort, and (C) MS-only cohort. (A–C) Black lines represent the best linear fit to plotted values, (B–C) blue dotted lines represent the best linear fit from the respective discovery cohort correlation, *n*=number of patients analyzed. Abbreviations: EDSS=expanded disability status scale, SNRS=Scripps neurological rating scale, SDMT=symbol digit modalities test, COMRIS-PDO-EDSS=combinatorial MRI scale – physical disability optimized by EDSS, COMRIS-PDO-SNRS= combinatorial MRI scale – physical disability optimized by SNRS, COMRIS-CDO-SDMT=combinatorial MRI scale – cognitive disability optimized by SDMT.

and upper CS. In the supratentorial brain, we also rated BHFr as a measure of lesion-destructiveness (Bielekova et al., 2005) and added qualitative description of the presence or absence of periventricular, juxtacortical, corpus callosum, deep WM and deep GM lesions. Next, for each selected MRI element, we devised semi-quantitative rating steps (with analogous numerical codes) that seemed logical (e.g. none, mild, moderate and severe atrophy) and time-efficient (i.e. for calculation of T2LL). Consequently, the quantification of T2LL focused on lower number of lesions (i.e. up to ten in the supratentorial brain and up to four in the remaining structures) with the goal to provide discriminatory value in patients in early stages of disease process (Brex et al., 2002), when atrophy is not yet visible with the naked eye.

We then retrospectively rated MRIs from 419 untreated subjects while being blinded to their prospectively acquired clinical scores. All MRI elements were found to positively correlate with disability (EDSS, SNRS, and SDMT), with the exception of CELs and qualitative description of predominance of deep WM lesions in supratentorial brain (Supplemental Fig. 1). Analysis of the rating steps demonstrated that we were not able to reproducibly differentiate subjects with moderate versus severe atrophy of the brainstem and cerebellum and therefore these two categories were collapsed into a single rating step (Moderate/Severe; Fig. 1).

3.2. *Mathematical modeling*

We employed mathematical modeling to select/validate those MRI “building blocks” (i.e. MRI elements and their rating) of the combinatorial scale of CNS tissue destruction that are biologically meaningful and non-redundant. If selected MRI characteristics are

capturing CNS tissue destructions, then mathematical models can "assemble" these non-redundant MRI elements into models that will have high correlations with clinical scales of physical (i.e. EDSS and SNRS) versus cognitive (i.e. SDMT) disability.

Indeed, the least-square methodology combined with leave-one-out-cross-validation revealed redundancy in MRI parameters; the mean square error in cross-validation was the lowest when only five to seven different variables comprised the model (Supplemental Fig. 2A). We therefore selected six variables per model.

Intriguingly, age ranked as the top variable in every optimized model (Supplemental Fig. 2B), consistent with epidemiological observations that MS-related disability is strongly influenced by age (Confavreux and Vukusic, 2006). Out of five remaining critical variables in EDSS and SNRS models (which overlapped significantly) infratentorial MRI parameters were the strongest determinants of physical disability, representing four out of five top MRI elements, followed by brain atrophy (Table 1). The main difference between EDSS and SNRS resided in selecting cerebellum atrophy [m] in the EDSS model, whereas deep GM lesions [h] was selected in the SNRS model (Supplemental Fig. 2B). Because of the significant correlation between EDSS and SNRS (Spearman $r=0.93$, $p<0.0001$) and the historical preference for EDSS, we chose the six highest ranking variables selected by the EDSS model (Table 1) to calculate two physical disability scales: COMRIS-PDO-EDSS (Physical Disability Optimized by EDSS) and COMRIS-PDO-SNRS (Physical Disability Optimized by SNRS). The COMRIS-predicted EDSS and SNRS show statistically highly significant ($p<0.0001$) Spearman correlations (0.75 and 0.76) with clinically measured EDSS and SNRS, respectively (Fig. 2A)

In contrast, different MRI parameters were selected by the model optimized for SDMT, namely brain atrophy [c], supratentorial T2LL [a], brainstem LL [i], cerebellum LL [j], and supratentorial deep GM lesions [h] (Table 1; Supplemental Fig. 2B). Thus, we named this model COMRIS-CDO-SDMT (Cognitive Disability Optimized by SDMT) and it shows significant ($p<0.0001$) Spearman correlation (0.65) with clinically measured SDMT (Fig. 2A).

3.3. Validation of disability-optimized COMRIS models

We validated physical- or cognitive-disability optimized versions of COMRIS in an independent cohort of 114 untreated subjects (MRI graded, but not utilized for mathematical modeling). All three MRI scores show highly statistically significant ($p<0.0001$) Spearman correlation coefficients (0.71, 0.73, and 0.49 for EDSS, SNRS, and SDMT, respectively) comparable to those calculated for the discovery cohort (Fig. 2B).

For the research groups only interested in MS, we assessed performance of COMRIS in an MS sub-cohort: we observed comparable, significant ($p<0.0001$) correlations of clinical measures with the predicted scores (Spearman correlations of 0.73, 0.75, and 0.68 for EDSS, SNRS, and SDMT, respectively) as in the discovery cohort (Fig. 2C).

To address potential influence of scanner/field/sequences, we re-analyzed correlations between COMRIS-predicted scores and clinical disability scales for the subgroup of subjects that were acquired on the 1.5 T scanner (3–5 mm cuts, $n=91$) versus 3 T scanners (1 mm^3 resolution, $n=105$). In both sub-cohorts, we validated statistical significance of correlations between COMRIS-predicted and clinical scales, even though the strength of correlations was higher for 3 T cohort (Spearman coefficients between predicted and measured clinical scales for EDSS/SNRS/SDMT $r=0.58/0.58/0.44$ for 1.5 T cohort and $r=0.82/0.85/0.67$ for 3 T cohort).

Finally, to address reproducibility of the ratings, we calculated intra- and inter-rater variability for 20 selected subjects that were tested independently by each rater, at two separate time points. The mean intra-rater variability for all COMRIS-predicted clinical scales was less than 5% (0.5–4.8% for EDSS, 0.6–1.7% for SNRS and 0.5–1% for SDMT). The intra-rater variability was larger, but still below 10% for all COMRIS-predicted clinical scales (6.9% for EDSS, 5.0% for SNRS and 2.4% for SDMT).

3.4. Universal COMRIS scale for translational studies

Mathematical modeling confirmed intuitive understanding that utilized clinical scales reflect only some aspects of clinical disability. Furthermore, as postulated in the introduction (and validated by mathematical models), lesion location has decisive influence on physical versus cognitive disability. In contrast, a biomarker of CNS tissue destruction should not be influenced by lesion location.

Therefore, while mathematical modeling selected subgroup of tested MRI elements as biologically meaningful and non-redundant, these endorsed MRI "building blocks" need to be re-assembled differently to reflect CNS tissue destruction globally. However, there is no "gold standard" measure of CNS tissue destruction that can be used for mathematical modeling. Consequently, we can device only a conceptual model based on current knowledge, with the understanding that such model inevitably represents only the initial step in subsequent re-iterative process of experimentation based model refinement, according to the principles of systems biology (Bielekova et al., 2014). To emphasize that following model represents only this initial step, we will designate it COMRIS-CTD-v1 (Combinatorial MRI Scale of CNS Tissue Destruction version 1).

Volumetrically, supratentorial brain represents approximately 70% of the CNS, whereas brainstem, cerebellum and spinal cord the remaining 30%. Therefore, we need to re-assemble validated MRI elements into comparable absolute parts (in terms of maximal achievable score), which will then be multiplied by 0.7 for the supratentorial brain and 0.3 for the infratentorial brain. Based on the correlations with disability, atrophy of the supratentorial brain reflects more strongly CNS tissue destruction than T2LL (Rudick et al., 2000; Barkhof, 1999 and current study). Additionally T1 hypo-intensities with T1 values similar to CSF (i.e. black holes) have been validated in the context of MRI-pathological correlations as a marker of axonal damage (Barkhof et al., 2000), even though corresponding MRI element (i.e. BHFr) was not selected by current mathematical models. We consider two plausible explanations for this discrepancy: First, it is a possibility that *a priori* selection of the 50% cut-off is not optimal; second possibility, which we deem more relevant, is the fact that BHFr has to be considered within the context of the T2LL, because by definition it represents proportion of T2LL that has an appearance of black holes. Because mathematical models did not consider partial quadratic terms, the biologically most relevant combinatorial marker (i.e. T2LL*BHFr) was not analyzed. Finally, the last supratentorial MRI element selected as biologically-meaningful by modeling is the presence of deep GM lesions; and the models assigned this parameter similar importance (weight) as T2LL. One of the possible explanations for this observation is that presence of deep GM lesions is a surrogate of more wide-spread GM pathology because it's been demonstrated that cortical and deep GM lesions almost always co-exist (Kutzelnigg et al., 2005). This has an important implication, because deep GM represents less than 5% of supratentorial brain volume, whereas all GM is close to 50%. Consequently, the assumption of surrogacy of deep GM lesions for the general GM pathology enhances the proportional "weight" of this contributing element 10-fold in comparison to its pure volumetric consideration, which is consistent with its weight assigned by the mathematical models.

Due to technical difficulties with obtaining reliable volumetric MRI data in the infratentorial CNS compartment, there is lesser knowledge about relationships between T2LL and atrophy in the remaining three sections. Mathematical models assigned comparable weights to T2LL and atrophy in cerebellum and medulla/CS, but actually did not select brainstem atrophy in any of the models. We believe that this reflects inability of the raters to reliably assess brainstem atrophy by naked eye, rather than the lack of its biological relevance.

Synthesizing all of this knowledge and conceptual reasoning, we developed the following formula for the COMRIS-CTD-v1:

COMRIS-CTD-v1 = 0.7*(supratentorial MRI elements; max score 35) + 0.3*(infratentorial MRI elements; max score 35)

COMRIS-CTD-v1 = 0.7*([a]*[b] + 6*[c] + 5*[h]) + 0.3*(2.5*([i] + [j] + [k] + [m] + [n]))

The "weights" (i.e. multiplication coefficients) in this formula are selected so that the atrophy of the supratentorial brain [c] gives higher absolute number (i.e. 6*3 = 18) than the highest possible score obtained by supratentorial lesions (i.e. T2LL*BHFr = 4*3 = 12) and the deep GM lesions [h] contribute approximately ten times higher values than what would be expected from their volumetric proportion. In the infratentorial brain, all five elements selected as biologically-meaningful in the mathematical models are weighted equally and multiplication coefficient is simply selected to give equal maximal absolute value of 35 for the infratentorial CNS, as for the supratentorial brain.

While validation/refinement of COMRIS-CTD-v1 has to come from follow-up studies (e.g. by modeling based on CSF biomarkers of CNS tissue destruction), there are few logical assumptions that one can make about biomarker of CNS tissue destruction: the disease duration should have stronger influence on the biomarker than age and while it should correlate with disability, the correlations should be lower in comparison to disability-optimized COMRIS versions. We tested these expectations and observed that they were all fulfilled (Table 2).

3.5. *COMRIS-predicted clinical scales provide comparable differences between disease cohorts as identified by measured clinical scales*

Because historical cohorts with available biological samples may not have matched clinical data (especially SNRS and SDMT) it is important to assess how well COMRIS models can re-construct these missing clinical data from available MRI scans.

Therefore, we compared discriminatory power of COMRIS-predicted clinical scores versus measured clinical scores in detecting significant differences between diagnostic subgroups. We observed that COMRIS-predicted EDSS, SNRS and SDMT differentiated diagnostic groups with equivalent power as prospectively-acquired clinical scales (Fig. 3A; analogous results were observed for discovery and validations cohorts, so only combined analysis is presented).

In a search for the parameters that may identify or forecast development of progressive MS, we observed an intriguing dichotomy between COMRIS-predicted and clinically-measured physical disability: the difference between COMRIS-projected and measured EDSS and SNRS scores were consistently higher in RRMS in comparison to progressive MS subjects (Fig. 3B). In other words, RRMS patients had generally lower disability in comparison to the amount of visible CNS tissue destruction on MRI, whereas progressive MS subjects had consistently higher disability in comparison to visible CNS tissue damage, suggesting that lack of compensatory (functional) recovery is the decisive factor that determines onset of progressive MS.

4. Discussion

Technological advances in biomedical research, such as genotyping and RNA sequencing, proteomics, lipidomics and metabolomics allow measurements of multimodality data in large patient cohorts. This produces "catalogue" of the elements that constitute a complex system, such as human body affected by MS (Bielekova et al., 2014). However, to reach predictive understanding of the system, systems biology needs to evolve from cataloguing system's parts to understanding relationships between them, as these are more important determinants of the system's behavior (Mesarovic et al., 2004). In concrete terms, this means understanding how genotypes interact with environmental factors to mediate development of phenotypes or which molecular or cellular mechanisms contribute to the destruction of target organ. These questions cannot be answered using traditional reductionists research methods and instead require application of systems biology approaches to large datasets assembled through international consortia.

The presented study originated from the need to develop a universal measure of CNS tissue destruction that can integrate diverse historical cohorts with banked biological samples for these new types of translational studies. While MRI of the CNS tissue seemed to be the logical vehicle for this purpose, use of different scanners and sequences represented a formidable challenge. Based

Table 2
The degree of correlation between COMRIS scales, age, disease duration and clinical disability.

		Age	Disease duration	EDSS	SNRS	SDMT
COMRIS-CTD-v1	$r_{Spearman}$	**0.3735**	**0.4424**	**0.6042**	**−0.6217**	**−0.5356**
	95% CI	0.2854 to 0.4554	0.3553 to 0.5218	0.5375–0.6633	−0.6787 to −0.5572	−0.6076 to −0.4549
COMRIS-PDO-EDSS	$r_{Spearman}$	0.5753	0.6034	**0.7388**	−0.7534	−0.4980
	95% CI	0.5053–0.6377	0.5334–0.6653	0.6904–0.7805	−0.7931 to −0.703	−0.5742 to −0.4132
COMRIS-PDO-SNRS	$r_{Spearman}$	−0.5582	−0.5982	−0.7366	**0.7537**	0.5056
	95% CI	−0.6225 to −0.4863	−0.6607 to −0.5275	−0.7787 to −0.6880	0.7076–0.7933	0.4217–0.5810
COMRIS-CDO-SDMT	$r_{Spearman}$	−0.6148	−0.4925	−0.6206	0.6530	**0.6022**
	95% CI	−0.6726 to −0.5495	−0.5670 to −0.4101	−0.6777 to −0.5560	0.5925–0.7061	0.5295–0.6661
	Number of subjects	419	382	419	419	358

Summary of Spearman correlation coefficients (r) for age, disease duration and three clinical scales (EDSS, SNRS, SDMT). Abbreviations: EDSS = expanded disability status scale, SNRS = Scripps neurological rating scale, SDMT = symbol digit modalities test, COMRIS-CTD-v1 = combinatorial MRI scale of CNS tissue destruction version 1, COMRIS-PDO-EDSS = combinatorial MRI scale – physical disability optimized by EDSS, COMRIS-PDO-SNRS = combinatorial MRI scale – physical disability optimized by SNRS, COMRIS-CDO-SDMT = combinatorial MRI scale – cognitive disability optimized by SDMT.

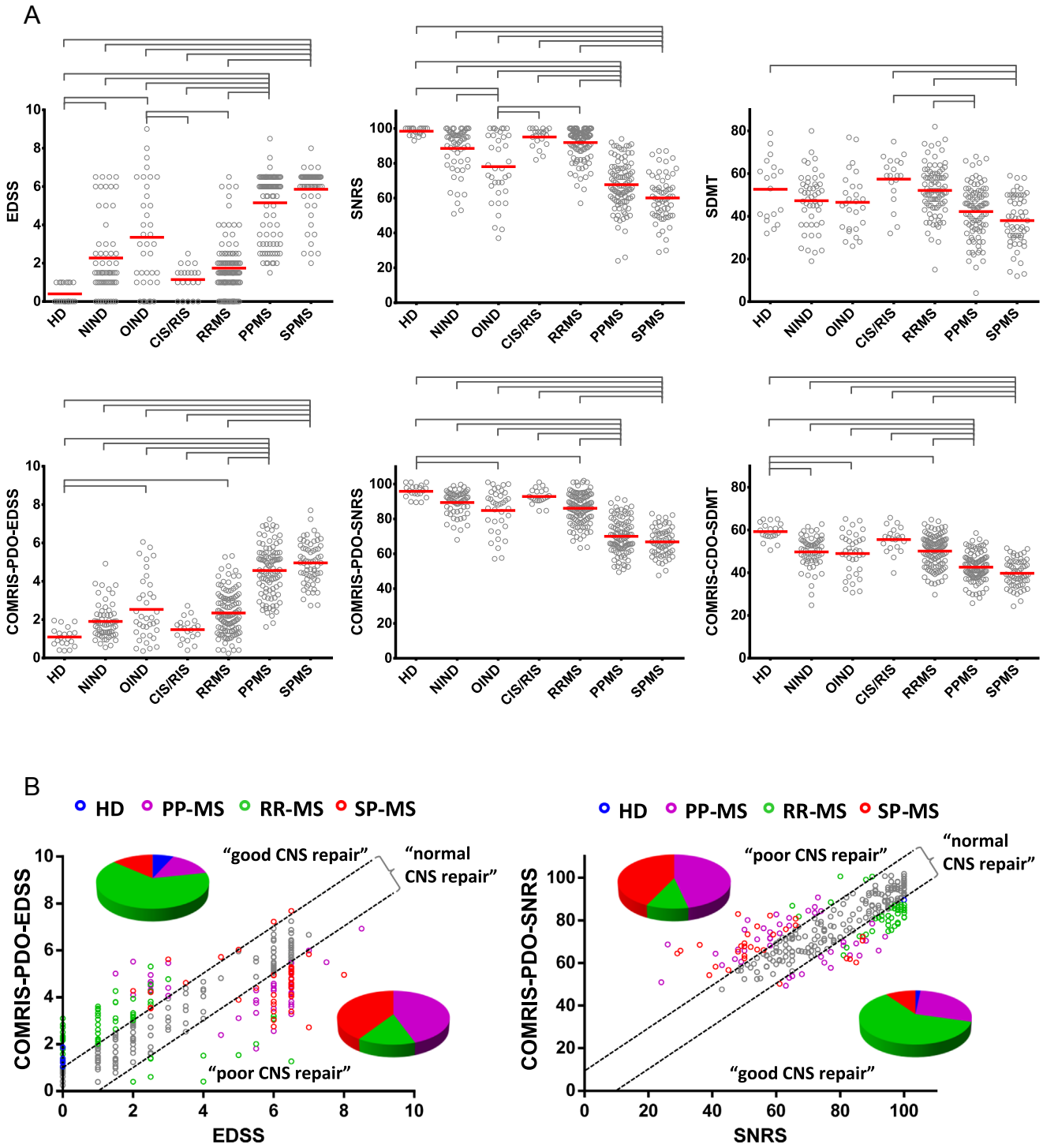

Fig. 3. COMRIS provides novel insight about diagnostic subgroups of neurological diseases. (A) Evaluation of clinical scores and COMRIS-predicted clinical scores (EDSS, SNRS, and SDMT) between diagnostic subgroups. Gray brackets highlight statistically significant differences between seven diagnostic subcategories with adjusted $p < 0.001$. The data (gray circles) are depicted with means (red bars). (B) Correlation between clinically-measured and COMRIS-predicted disability scores (EDSS and SNRS) identifies a group of subjects characterized by lower clinical disability than predicted by the amount of CNS tissue destruction observed on MRI ("good CNS repair") and a group of subjects with higher clinically measured disability than predicted by MRI ("poor CNS repair") for both EDSS and SNRS. The black dashed lines identify subjects (grey circles) with "normal CNS repair" with predicted disability within 10% of clinically measured disability. Subjects with good and poor CNS repair are color-coded based on their diagnosis (blue – HD, magenta – PPMS, green – RRMS, red – SPMS), the pie-charts show distribution of diagnoses for subjects with poor or good CNS repair for both EDSS and SNRS. Abbreviations: HD = healthy donors, NIND = non-inflammatory neurological disorders, OIND = other inflammatory neurological disorder, CIS/RIS = clinically-/radiologically isolated syndrome, RRMS = relapsing-remitting multiple sclerosis, PPMS = primary progressive multiple sclerosis, SPMS = secondary progressive multiple sclerosis, EDSS = expanded disability status scale, SNRS = Scripps neurological rating scale, SDMT = symbol digit modalities test, COMRIS-PDO-EDSS = combinatorial MRI scale – physical disability optimized by EDSS, COMRIS-PDO-SNRS = combinatorial MRI scale – physical disability optimized by SNRS, COMRIS-CDO-SDMT = combinatorial MRI scale – cognitive disability optimized by SDMT.

on our experience with MRI analyses in natural history cohorts (Bielekova et al., 2005; Markovic-Plese et al., 2003) and in clinical trials of MS (Bielekova et al., 2009; Borges et al., 2013), we recognized that semi-quantitative MRI measures, such as number of CELs or T2 lesions, have an excellent signal-to-noise ratio but limited correlation with clinical outcomes. On the other hand, qMRI measures suffer from considerable non-biological variability (Cover et al., 2011) and are affected by hardware (scanner/field strength, coil), sequence parameters (Neacsu et al., 2008) and post-processing methods. At least part of qMRI variability is due to movement, which is inevitable even in the most cooperative subjects due to the long duration of the MRI in comparison to the small changes that qMRI metrics can discern. Even when research groups remove MRIs due to "poor technical quality", quantitative changes of censored MRI data are reliable across several years but not in shorter intervals (Bakshi et al., 2008; Ingle et al., 2003; Kappos et al., 1999; Khaleeli et al., 2008; Rovaris et al., 2006; Stevenson et al., 2004).

As an alternative solution that mitigates the aforementioned problems, we considered development of a combinatorial scale based on thoughtful combination of several semi-quantitative MRI measures that are validated through mathematical modeling as biologically meaningful because they can be assembled into models that correlate highly with physical and cognitive disability (Filippi et al., 1995; Mainero et al., 2001; Riahi et al., 1998; Sailer et al., 2001). We were actually impressed by the strength and high reproducibility of the developed models and how well the models selected those anatomical structures that underlie physical versus cognitive (Kearney et al., 2014; Lazeron et al., 2005; Preziosa et al., 2014), without any pre-existing knowledge of biology. This validates the MRI elements selected by mathematical models as biologically meaningful. Our comparison of results between 1.5 T and 3 T cohorts indicates that while higher field strength and enhanced sequences increase performance of the COMRIS, results obtained from older 1.5 T cohorts still remain valid; confirming our presupposition that semi-quantitative MRI scale will moderate problems related to unavoidable differences in the collection of MRI data between diverse cohorts. Finally, while the semi-quantitative nature of COMRIS leads to intra- and inter-rater differences, adherence to the provided rating guide (Fig. 1) assures that these technical errors are below 10%. The strength of independent validation (including the precision of regression slopes, Fig. 2B) on a cohort three times smaller than the modeling cohort, authenticates the robustness of COMRIS models and their MRI "building blocks".

COMRIS-CTD-v1 is not a diagnostic scale; on the contrary – as a scale of CNS tissue destruction, it should be applicable to more diseases than just MS. While MS MRI studies (Bielekova et al., 2005; Bieniek et al., 2006; Filippi et al., 1995; Fisher et al., 2000; Fisniku et al., 2008; Truyen et al., 1996; Zivadinov and Bakshi, 2004) guided selection of semi-quantitative elements, we saw no reason to apply limitations of thresholds used by MS diagnostic criteria (Polman et al., 2011; McDonald et al., 2001). Instead, to promote broad utilization, simplicity became important consideration. For example, quantification of supratentorial T2LL beyond ten lesions takes disproportionately more of the rater's effort in relation to benefit provided. In patients with supratentorial T2LL above ten, the magnitude of brain atrophy is a more important determinant of disability than presence of additional lesions, as confirmed by modeling.

The main limitation is *a priori* selection of MRI elements and their rating steps, leading to a possibility that different MRI markers or rating steps may provide a better performance. We expect that current publication will spur further developmental work from interested research groups, which may test a broader combination of MRI elements and potentially improve the scale, analogous to the improvements in MS diagnostic criteria (Polman et al., 2011). Nevertheless, we also believe that current version of COMRIS represents an excellent starting point for unifying diverse historical cohorts to aforementioned multicenter translational studies, based on the outstanding performance of COMRIS models in estimating clinical disability. Because any neurological scale is only an approximation of the true disability and most of the brain tissue is "clinically silent", the correlation between MRI and clinical scales can never achieve rho = 1. As the correlations between COMRIS and disability scales are comparable to those previously reported for qMRI measures, (Fisher et al., 2000; Riahi et al., 1998; Sailer et al., 2001; Nijeholt et al., 1998) including composite qMRI scales (Mainero et al., 2001; Bakshi et al., 2014; Moodie et al., 2012), COMRIS predicts disability exceptionally well.

COMRIS supports the notion that multi-modality MRI imaging and/or development of composite models (Mainero et al., 2001; Bakshi et al., 2014; Moodie et al., 2012; Sormani et al., 2014) that merge imaging parameters reflecting somewhat different biological processes explains higher proportion of disability variance than any single imaging modality. This notion is also evident from a large meta-analysis of MS clinical trials, when Sormani et al. demonstrated that including both T2 lesion load and brain atrophy in mathematical model significantly strengthen correlation between drug-induced changes on MRI and disability (Sormani et al., 2014). Therefore, the emerging question is whether application of analogous systematic modeling used for development of COMRIS to cohorts with fully quantitative MRI measures will lead to composite qMRI scale(s) that outperforms COMRIS in the ability to predict clinical and cognitive disability.

We expect that as a tool that can readily bridge historical patient cohorts to diverse translational research projects, COMRIS will facilitate multicenter studies, such as those being formed under the Progressive MS Alliance. In order to assist such collaborations, we provide a downloadable Excel file (Table S2) with the template of non-redundant MRI elements and macro that calculates all aspects of the COMRIS, including MRI-predicted disability scores.

5. Conclusions

Clinical MRI of the brain routinely performed in vast majority of patients with neurological symptoms represents a unique opportunity to develop a tool for bridging diverse cohorts of patients in multicenter projects. Using a disability-guided mathematical combination of semiquantitative assessments of brain MRI we developed and validated a universal, reproducible and clinically relevant scale of CNS tissue destruction that can be applied across different MRI scanners/coils/sequences. As such this scale is readily available to facilitate multicenter studies and research projects comparing historical patient cohorts.

Funding

This study was funded by the intramural research program of the National Institute of Neurological Disorders and Stroke, National Institutes of Health.

Disclosure

The authors report no disclosures relevant to the manuscript.

Authors' contribution

BB – conceptual study design.

PK, MK, BB – study design, data analysis, data interpretation, figures, writing.

RW, TG – mathematical modeling, writing.

TW – statistical analysis.

IC, JO, KF, JC, BB – collection of clinical data.

Conflict of interest disclosure

The authors report no disclosures relevant to the manuscript.

Acknowledgments

This work was supported by the Intramural research program of the National Institute of Neurological Disorders and Stroke of the National Institutes of Health. We thank Govind Bhagavatheeshwaran, Alison Wichman, and Muktha Natrajan for critical review of the manuscript and clinical team members who did not fulfill co-authorship criteria (Jenifer Dwyer, Alison Wichman and Anne Mayfield) for expert patient care and support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.msard.2015.08.009.

References

Bakshi, R., Neema, M., Healy, B.C., et al., 2008. Predicting clinical progression in multiple sclerosis with the magnetic resonance disease severity scale. Arch. Neurol. 65 (11), 1449–1453.

Bakshi, R., Neema, M., Tauhid, S., et al., 2014. An expanded composite scale of MRI-defined disease severity in multiple sclerosis: MRDSS2. Neuroreport 25, 1156–1161.

Barkhof, F., 1999. MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS). Mult. Scler. 5 (4), 283–286.

Barkhof, F., Karas, G.B., van Walderveen, M.A., 2000. T1 hypointensities and axonal loss. Neuroimaging Clin. N Am. 10 (4), 739–752.

Bielekova, B., Kadom, N., Fisher, E., et al., 2005. MRI as a marker for disease heterogeneity in multiple sclerosis. Neurology 65, 1071–1076.

Bielekova, B., Howard, T., Packer, A.N., et al., 2009. Effect of anti-CD25 antibody daclizumab in the inhibition of inflammation and stabilization of disease progression in multiple sclerosis. Arch. Neurol. 66 (4), 483–489.

Bielekova, B., Vodovotz, Y., An, G., Hallenbeck, J., 2014. How implementation of systems biology into clinical trials accelerates understanding of diseases. Front Neurol. 5, 102.

Bieniek, M., Altmann, D.R., Davies, G.R., et al., 2006. Cord atrophy separates early primary progressive and relapsing remitting multiple sclerosis. J. Neurol. Neurosurg. Psychiatry 77 (9), 1036–1039.

Borges, I.T., Shea, C.D., Ohayon, J., et al., 2013. The effect of daclizumab on brain atrophy in relapsing-remitting multiple sclerosis. Mult. Scler. Relat. Disord. 2 (2), 133–140.

Brex, P.A., Ciccarelli, O., O'Riordan, J.I., Sailer, M., Thompson, A.J., Miller, D.H., 2002. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. N. Engl. J. Med. 346 (3), 158–164.

Chard, D.T., Griffin, C.M., Parker, G.J., Kapoor, R., Thompson, A.J., Miller, D.H., 2002. Brain atrophy in clinically early relapsing-remitting multiple sclerosis. Brain 125 (Pt 2), 327–337Feb 125 (Pt 2), 327–337.

Confavreux, C., Vukusic, S., 2006. Age at disability milestones in multiple sclerosis. Brain 129 (Pt 3), 595–605.

Cover, K.S., van Schijndel, R.A., van Dijk, B.W., et al., 2011. Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans. Psychiatry Res. 193 (3), 182–190.

Filippi, M., Paty, D.W., Kappos, L., et al., 1995. Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis: a follow-up study. Neurology 45 (2), 255–260.

Fisher, E., Rudick, R.A., Cutter, G., et al., 2000. Relationship between brain atrophy and disability: an 8-year follow-up study of multiple sclerosis patients. Mult. Scler. 6 (6), 373–377.

Fisniku, L.K., Brex, P.A., Altmann, D.R., et al., 2008. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain 131 (Pt 3), 808–817.

Ingle, G.T., Stevenson, V.L., Miller, D.H., Thompson, A.J., 2003. Primary progressive multiple sclerosis: a 5-year clinical and MR study. Brain 126 (Pt 11), 2528–2536.

Jones, B.C., Nair, G., Shea, C.D., Crainiceanu, C.M., Cortese, I.C., Reich, D.S., 2013. Quantification of multiple-sclerosis-related brain atrophy in two heterogeneous MRI datasets using mixed-effects modeling. Neuroimage Clin. 3, 171–179.

Kappos, L., Moeri, D., Radue, E.W., et al., 1999. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis Gadolinium MRI Meta-analysis Group. Lancet 353 (9157), 964–969.

Kearney, H., Rocca, M.A., Valsasina, P., et al., 2014. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. Mult. Scler. 20 (1), 72–80.

Khaleeli, Z., Ciccarelli, O., Manfredonia, F., et al., 2008. Predicting progression in primary progressive multiple sclerosis: a 10-year multicenter study. Ann. Neuro. 63 (6), 790–793.

Kurtzke, J.F., 1983. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 33 (11), 1444–1452.

Kutzelnigg, A., Lucchinetti, C.F., Stadelmann, C., et al., 2005. Cortical demyelination and diffuse white matter injury in multiple sclerosis. Brain 128 (Pt 11), 2705–2712.

Lazeron, R.H.C., Boringa, J.B., Schouten, M., et al., 2005. Brain atrophy and lesion load as explaining parameters for cognitive impairment in multiple sclerosis. Mult. Scler. 11 (5), 524–531.

Mainero, C., De Stefano, N., Iannucci, G., et al., 2001. Correlates of MS disability assessed in vivo using aggregates of MR quantities. Neurology 56 (10), 1331–1334.

Markovic-Plese, S., Bielekova, B., Kadom, N., et al., 2003. Longitudinal MRI study: the effects of azathioprine in MS patients refractory to interferon beta-1b. Neurology 60 (11), 1849–1851.

McDonald, W.I., Compston, A., Edan, G., et al., 2001. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann. Neurol. 50 (1), 121–127.

Mesarovic, M.D., Sreenath, S.N., Keene, J.D., 2004. Search for organising principles: understanding in systems biology. Syst. Biol. 1 (1), 19–27.

Moodie, J., Healy, B.C., Buckle, G.J., et al., 2012. Magnetic resonance disease severity scale (MRDSS) for patients with multiple sclerosis: a longitudinal study. J. Neurol. Sci. 315 (1–2), 49–54.

Neacsu, V., Jasperse, B., Korteweg, T., et al., 2008. Agreement between different input image types in brain atrophy measurement in multiple sclerosis using SIENAX and SIENA. J. Magn. Reson. Imaging 28 (3), 559–565.

Nijeholt, G.J., van Walderveen, M.A., Castelijns, J.A., et al., 1998. Brain and spinal cord abnormalities in multiple sclerosis. Correlation between MRI parameters, clinical subtypes and symptoms. Brain: J. Neurol. 121, 687–697.

Parmenter, B.A., Weinstock-Guttman, B., Garg, N., Munschauer, F., Benedict, R.H., 2007. Screening for cognitive impairment in multiple sclerosis using the Symbol digit Modalities Test. Mult. Scler. 13 (1), 52–57.

Polman, C.H., Reingold, S.C., Banwell, B., et al., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol. 69 (2), 292–302.

Preziosa, P., Rocca, M.A., Mesaros, S., et al., 2014. Relationship between damage to the cerebellar peduncles and clinical disability in multiple sclerosis. Radiology 271 (3), 822–830.

Riahi, F., Zijdenbos, A., Narayanan, S., et al., 1998. Improved correlation between scores on the expanded disability status scale and cerebral lesion load in relapsing-remitting multiple sclerosis results of the application of new imaging methods. Brain: J. Neurol. 121, 1305–1312.

Rovaris, M., Judica, E., Gallo, A., et al., 2006. Grey matter damage predicts the evolution of primary progressive multiple sclerosis at 5 years. Brain 129 (Pt 10), 2628–2634.

Rudick, R.A., Fisher, E., Lee, J.C., Duda, J.T., Simon, J., 2000. Brain atrophy in relapsing multiple sclerosis: relationship to relapses, EDSS, and treatment with interferon beta-1a. Mult. Scler. 6 (6), 365–372.

Sailer, M., Losseff, N.A., Wang, L., Gawne-Cain, M.L., Thompson, A.J., Miller, D.H., 2001. T1 lesion load and cerebral atrophy as a marker for clinical progression in patients with multiple sclerosis. A prospective 18 months follow-up study. Eur. J. Neurol. 8 (1), 37–42.

Sipe, J.C., Knobler, R.L., Braheny, S.L., Rice, G.P., Panitch, H.S., Oldstone, M.B., 1984. A neurologic rating scale (NRS) for use in multiple sclerosis. Neurology 34 (10), 1368–1372.

Sormani, M.P., Arnold, D.L., De Stefano, N., 2014. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. Ann. Neurol. 75 (1), 43–49.

Stevenson, V.L., Ingle, G.T., Miller, D.H., Thompson, A.J., 2004. Magnetic resonance imaging predictors of disability in primary progressive multiple sclerosis: a 5-year study. Mult. Scler. 10 (4), 398–401.

Truyen, L., van Waesberghe, J.H., van Walderveen, M.A., et al., 1996. Accumulation of hypointense lesions ("black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis. Neurology 47 (6), 1469–1476.

Zivadinov, R., Bakshi, R., 2004. Role of MRI in multiple sclerosis II: brain and spinal cord atrophy. Front. Biosci. 9, 647–664.