Diplomarbeit

Comparison of brain volume changes and lesion metrics in multiple sclerosis between two MS centers in Graz and Linz

eingereicht von
Anna Damulina

zur Erlangung des akademischen Grades

Doktor der gesamten Heilkunde
(Dr. med. univ.)

an der
Medizinischen Universität Graz
ausgeführt an der
Universitätsklinik für Neurologie

unter der Anleitung von

Assoz. Prof. Priv.-Doz. Dr. med. univ. Christian Enzinger

sowie

Dr. med. univ. Alexander Pichler, PhD

Graz, 11.04.2018

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 11.04.2018 Anna Damulina eh

ACKNOWLEDGEMENT

First and foremost, I would like to thank and express my sincere gratitude to my supervisor Prof. Christian Enzinger for the valuable guidance and advice. Thanks for all his constructive comments and wise suggestions. I have learnt a lot from him.

I am grateful to my second supervisor, Dr. Alexander Pichler. His provided clinical and MRI data have greatly helped me in my work. Many thanks to BSc Florian Borsodi and DI Lukas Pirpamer, who supported me in technical questions and with the technical evaluation of the MR data.

I would like to show gratitude to OA Dr. Michael Guger and his team from the Kepler University Hospital in Linz for the provided MRI and clinical data from their site. Without their data our study could not have been successfully conducted.

I sincerely thank Prof. Franz Fazekas for providing me an opportunity to join his team.

TABLE OF CONTENTS

Abstract

Background and objective:

Cerebral Magnetic Resonance Imaging (MRI) has been proposed for disease monitoring in patients with Multiple Sclerosis (MS). However, outside standardized trials, considerable interscanner and intrascanner variability has to be expected for such measures. So far, the reliability of such brain measures in clinical practice has received relatively little systematic investigation. Our aim thus was to determine the effect of using different MRI protocols at different scanners on the measurement of lesion load and brain volumes metrics in clinical similar cohorts of MS patients of two different centers.

Methods:

174 MS patients from two MS centers (3D MPRAGE MR scans of 111 patients from Graz and 2D SE MR scans of 63 patients from Linz) at baseline and follow-up (after a mean of 3.42±1.58 years) were analyzed by using SIENAX and SIENA. Patients were comparable regarding age, disease duration, Expanded Disability Status Scale, and the proportion of patients on disease-modifying therapy. Variability between 2D and 3D data segmentation was estimated by the difference in brain metrics, percentage brain volume change (PBVC), coefficients of variation and percentage differences in volumes of brain metrics.

Results:

Lesion load estimation showed similar T2 hyperintense lesion load but significantly different T1 hypointense ("black hole") lesion load between the two cohorts (0.35 cm^3 in Graz and 0 cm^3 in Linz). Results from SIENAX varied significantly for all brain metrics except the global brain volume between the Graz and Linz cohorts (white matter volume 768 cm^3 and 854 cm^3, grey matter volume 797 cm^3 and 706 cm^3, respectively). Mean percentage differences showed that despite similar results concerning the global brain volume, the ratios of white and grey matter volumes between the two centers were completely different (-12% in grey and 10% in white matter between the cohorts). SIENA analyses demonstrated that PBVC was significantly lower in Linz cohort than in Graz (-0.19% and -0.39%, respectively).

Conclusions:

Even in cohorts of MS patients with similar clinical characteristics, obtaining MRI data at different scanners (despite locally standardized protocols) yields significant differences in several lesion and brain volume metrics. Given the suggested role of brain volume measurements as a surrogate of atrophy in the concept of "no evidence of disease activity" mirroring the neurodegenerative component of the disease, this study emphasizes the need for staying cautious when interpreting such values both within as well as across cohorts of different centers.

Zusammenfassung

Ziele:

Die zerebrale Magnetresonanztomographie (MRT) spielt eine zunehmend wichtige Rolle für das Krankheits-Monitoring bei PatientInnen mit Multipler Sklerose (MS). Abseits von standardisierten Studien ist jedoch mit einer erheblichen Interscanner- und Intrascanner-Variabilität zu rechnen. Bisher wurde die Zuverlässigkeit solcher Gehirnmessungen in der klinischen Praxis systematisch wenig untersucht. Unser Ziel war daher, den Effekt von verschiedenen MRT-Protokollen und unterschiedlichen Scannern auf die Messung von Läsions- und Hirnvolumina in klinisch ähnlichen Kohorten von MS-PatientInnen aus zwei verschiedenen Zentren zu untersuchen.

Methoden:

174 MS-Patienten aus zwei MS-Zentren (3D-MPRAGE-MR-Scans von 111 PatientInnen aus Graz und 2D-SE-MR-Scans von 63 PatientInnen aus Linz) wurden anhand einer Basisuntersuchung und einer Kontrolluntersuchung (nach durchschnittlich 3.42 + 1.58 Jahren) unter Verwendung der Software SIENAX und SIENA analysiert. Die PatientInnen waren in Bezug auf Alter, Krankheitsdauer, Expanded Disability Status Scale und der immunmodulierenden Therapie vergleichbar. Die Variabilität zwischen der 2D- und 3D-Datensegmentierung wurde anhand des Unterschieds in der Läsionsvolumina und Hirnvolumina, der prozentuellen Veränderung des Gehirnvolumens (PBVC), der Variationskoeffizienten und der prozentuellen Unterschiede der Hirnparameter bestimmt.

Ergebnisse:

Die T2-Läsionsvolumina waren zwischen den 2 Gruppen vergleichbar; es zeigte sich jedoch ein unterschiedliches T1- ("black hole") Läsionsvolumen (0.35 cm^3 in Graz und in Linz 0 cm^3). Die Analyse der verschiedenen Hirnvolumina zeigte eine deutliche Varianz für sämtliche der untersuchten Parameter mit Ausnahme des globalen Hirnvolumens zwischen der Grazer und Linzer Kohorte (weiße Substanz Volumen 768 cm^3 und 854 cm^3, graue Substanz Volumen 797 cm^3 und 706 cm^3 bzw.). Die mittleren prozentuellen Anteile der weißen und grauen Substanzvolumina zwischen den beiden Zentren zeigten große Unterschiede (-12% in grauer Substanz und 10% in weißer Substanz zwischen den Kohorten). Die PBVC in der Kohorte Linz zeigte sich höher im Vergleich zu der Grazer Kohorte (-0.19% bzw. -0.39%).

Schlussfolgerungen:

Diese Studie konnte zeigen, dass bei MS-PatientInnen mit vergleichbaren klinischen Merkmalen bedingt durch Anwendung verschiedener MRT Protokolle und Scanner deutliche Unterschiede in Bezug auf verschiedene Läsions- bzw. Hirnvolumina bestehen.

Angesichts der zunehmenden Bedeutung der Hirnvolumenmessungen als Tool zum Monitoring der Krankheitsaktivität, zeigt diese Studie die Notwendigkeit auf, bei der Interpretation solcher Messungen mit besonderer Sorgfalt vorzugehen.

List of Abbreviations

2D - two-dimensional

3D - three-dimensional

APP - amyloid precursor protein

BVL - brain volume loss

CIS - clinically isolated syndrome

CNS - Central nervous system

CoV - coefficient of variation

CSF - cerebro-spinal fluid

DIS - dissemination in space

DIT - dissemination in time

DMT - disease-modifying therapy

EDSS - Expanded Disability Status Scale

GBV - global brain volume

GE - gradient-echo

GM - grey matter

GMV - grey matter volume

GMV - grey matter volume

IQR - interquartile range

MRI - Magnetic resonance imaging

MS - Multiple sclerosis

MSFC - Multiple Sclerosis Functional Composite

NEDA - no evidence of disease activity

PBVC - percentage brain volume change

PGM - peripheral grey matter

RRMS - relapsing–remitting multiple sclerosis

SD - standard deviation

SE - spin-echo

TR - repetition time

VCSF - ventricular cerebrospinal fluid

WM - white matter

WMV - white matter volume

List of Figures

List of Tables

1. Introduction

Multiple sclerosis (MS) is a chronic inflammatory and degenerative disease of the central nervous system (CNS) characterized by demyelination and axonal damage of the brain. It is among the most common causes of non-traumatic neurological disability in young adults, affecting mostly young women between 20 and 40 years [Kister I. et al. 2013]. Over the last years technical progress in magnetic resonance imaging (MRI) has noticeably advanced and nowadays MRI is the most powerful tool for MS diagnosing and monitoring [Enzinger C. et al., 2015, Watties M. et al., 2015]. MR studies demonstrated an important role of processes of inflammation and neurodegeneration in understanding the mechanisms underlying MS [Zivadinov R. et al., 2005; Horakova D. et al., 2009]. Besides focal lesions, MRI may show volume loss of the grey and white matter in brains of MS patients [Lavorgna L. et al., 2014; Minneboo A. et al., 2008]. Brain atrophy is generally measured from T1-weighted images by means of automatic computational methods [Giorgio A. et al. 2008, Lysandropoulos A. et al., 2016]. With these methods, progressive decreases in global brain volume have been demonstrated in MS patients [De Stefano N. et al., 2003]. Measurement of brain atrophy as an objective marker of MS severity with the potential to monitor treatment efficacy in MS has been suggested as an important parameter in patient monitoring of [Barkhof F. et al., 2009,2012].

In our study, we thus concentrated on the variability of such MRI features in two different cohorts of MS patients with different MRI protocols, reflecting routine clinical practice scenarios. Such finding appear of special interest when it comes to judging whether longitudinal patient MRI examinations may serve to reliably monitoring disease activity and the effect of disease-modifying therapies (DMT) at different centers (i.e., consider the proportion of patients exceeding a threshold of brain volume loss beyond what is expected from normal aging). During the selection of the cohorts, we put an emphasis on the similarity of clinical features and phenotypes of the disease, expecting that clinically similar patients should demonstrate roughly similar brain metrics. For the better understanding of general MS features, this chapter also deals with aspects of MS pathophysiology, clinical symptomatology, diagnostic criteria and measurements of patient's disability, and particularly several MRI measures implicated in the disease process.

1.1 Features of Multiple sclerosis

1.1.1 Epidemiology

The median prevalence of the MS is approximately 33 per 100,000 people, but greater varies depending on geography and ethnicity. MS is more common in the northern hemisphere and there is some genetic susceptibility in the populations of northern Europe or Scandinavia, with prevalences in North America and Europe of 140 and 108 per 100,000 people, respectively. The disease is significantly less disease frequent in black Africans and native populations of the America and Oceania, which have the lowest MS prevalence with 2.2 and 2.1 per 100,000 people, respectively [Belbasis L. et al., 2015; Vidal-Jordana A. et al., 2017]. Females of all races now have incidence rates for MS some three times those of their male counterparts [Wallin M. et al., 2012]. Worldwide, approximately over 2.3 million people suffer from MS. The incidence has increased considerably over the last century, and it is expected to further increase in the future [Weinshenker B., 1998; Hauser S. et al., 2006].

MS is a multifactorial disease, triggered by environmental factors in individuals with complex genetic profiles [Belbasis L. et al., 2015]. MS is influenced by multiple independent or interacting polymorphic genes, with risk alleles common in the population, and each exerting a small or at most a moderate effect to the overall risk. Some epidemiologic risk factors include exposure to the Epstein Barr virus during proposed childhood with manifestations of infectious mononucleosis, deficiency of vitamin D and smoking [Ascherio A. et al., 2007; Hollenbach J. et al., 2015].

The clinical symptomology of MS is heterogeneus, but walking impairment mainly determines disability and disturbs daily activities and quality of life of patients. The average time between disease onset and difficulty in ambulation is eight years [Ma V. et al., 2014].

1.1.2 Pathophysiology

Historically, inflammation of the CNS was considered as the primary pathophysiologic event in MS [Trapp B. et al., 2008]. Early studies of the disease pathogenesis thus more strongly focused on focal white matter demyelination. More recently, studies have refocused on the role of axonal degeneration, unresponsive to immunosuppression, as a major cause of disability in MS patients [Trapp B. et al., 1998; Bitsch A. et al., 2000; Geurts J. et al., 2003].

Matthews et al. reported on heterogeneity of the "pathological" changes underlying MRI lesions. Central brain N-acetyl aspartate, a neuronal-specific marker, measured by magnetic resonance spectroscopy [Birken D. et al., 1989], suggesting axonal loss was found to be reduced relative to normal controls [Matthews P. et al., 1998].

Ferguson B. et al. used an antibody against amyloid precursor protein (APP) in immunocytochemical experiments on paraffin embedded MS lesions of varying ages in order to see at which stage of the disease axonal damage, in addition to demyelination, occurs and may thus contribute to the development of disability in patients [Ferguson B. et al., 1997]. APP is best known as the molecule synthesized in the cell bodies of neurons and conveyed outward to the distal synapses, undergoing anterograde axonal transport. In event of an axonal transection, this transport is disrupted and APP accumulates in the axonal ends. The results showed expression of APP in damaged axons within acute multiple sclerosis lesions, and in the active borders of less acute lesions [Ferguson B. et al., 1997].

Likewise T1-hypointense lesions and the extent of brain atrophy represent regions where irreversible axonal loss, demyelination, and gliosis have occurred and correlate these markers somewhat with the degree of disability [Grimaud J. et al., 1999; Filippi M. et al., 2012]. Degeneration and axonal transection take place in the acute inflammatory [Trapp B. et al. 1998] and chronic [Ganter P. et al. 1999, Bjartmar C. et al. 2000] demyelination. Recent studies [Schultz V. et al., 2017] support a role of axonal degeneration in ongoing demyelination of focal lesions. The major structural features in MS patients are the axonal damage that follow demyelination; in the center of new therapies of the MS treatment remyelination playes key role.

1.1.3 Clinical symptoms

Focal demyelination of the white matter cannot sufficiently explain the variety of MS symptoms, for example cognitive impairment, which indicates a role of axons in the grey matter [Barkhof F. 2002; Geurts J. et al., 2008].

MS symptoms vary related to the location of the affected area. Based on symptoms onset and evolution, in 1996 the clinical phenotypes of MS were defined: relapsing–remitting MS (RRMS), secondary-progressive MS, primary-progressive MS, and relapsing-progressive MS [Lublin F. et al., 1996]. Increased understanding of MS with new imaging and biological correlates has led to re-examination of MS disease phenotypes: relapsing-progressive MS has been removed from the classification and the clinically isolated syndrome (CIS) has been included [Lublin F. et al., 2014]. Classic CIS symptoms or subsequent relapses usually are represented by the affection the optic nerves (20%), the brainstem (10%–20%), or the spinal cord (40%) causing an optic neuritis, a brainstem syndrome, or an incomplete transverse myelitis, respectively [Vidal-Jordana A. et al., 2107] (Table 1).

Table 1. Neurologic symptoms of multiple sclerosis by Vidal-Jordana A. et al., 2017

Relapse neurologic symptoms	
Optic nerve	Mononuclear painful vision loss
Spinal cord	Hemiparesis, mono-/paraparesis Hypoesthesia, dysesthesia, parasthesia Urinary and/or fecal sphincter dysfunction
Brainstem and cerebellum	Diplopia, oscillopsy Vertigo Gait ataxia, dysmetria Intentional / Postural tremor Facial paresis and / or hypoesthesia
Cerebral hemisphere	Facio–brachial–crural hemiparesis Facio–brachial–crural hemihypoesthesia
Other clinical manifestations	Paroxistic symptoms Painful spasms / spasticity Dysarthria / dysphagia Neuropathic pain Sexual dysfunction Spastic gait Ataxic gait Fatigue Cognitive impairment Depression Seizures

1.1.4 Diagnostic criteria

Formulating diagnostic criteria with face validity for such a heterogeneous disease as MS constitutes a difficult task in clinical practice. For the diagnosis of MS different sets of criteria have been proposed, that have been revised over time, depending on the research development.

The first standardized criteria were introduced in 1965 by Schumacher et al. [Schumacher G. et al., 1965]. They were based on clinical findings confirming the concept of dissemination in space (DIS) and dissemination in time (DIT), what also became key features of later imaging MS criteria. Development of MS diagnostic tests such as neuroimaging, evoked potentials and cerebro-spinal fluid (CSF) evaluations were considered in the Poser criteria in 1983 [Poser C. et al., 1983], the first criteria, that included MRI. Paraclinical investigations show abnormalities that indicate the distribution of inflammatory lesions and axonal loss (MRI); interference of conduction in previously myelinated pathways (evoked electrophysiological potentials); and intrathecal synthesis of oligoclonal antibodies (examination by lumbar puncture of the CSF) [Compston A. et al., 2008].

In the following years, the aim of research was to identify specific lesion characteristics that predicted disease progression. In the beginning of the 21 century introduction of MRI in the diagnostic process prompted to revise the Poser criteria.

In 2001 an International Panel on the Diagnosis of Multiple Sclerosis formulated the full utility of MRI to confirm both dissemination in space and time. The McDonald criteria were revised two times in 2005 and in 2010 to simplify the approach, while maintaining sufficient sensitivity and specificity [Polman H. et al., 2005; 2011]. The McDonald criteria 2010 are based on clinical assessment and emphasize the need to demonstrate DIS and DIT (Table 3), allowing use of MRI to demonstrate these features. Additionally, the diagnostic criteria require excluding alternative diagnoses. The concept underlying the diagnostic criteria remained unchanged, but over the course of time, the way has changed in which certain requirements may be fulfilled.

The revised McDonald diagnostic criteria 2010 use an interpretation of imaging criteria for DIS and DIT as articulated by the European multicenter collaborative research network (MAGNetic Resonance In Multiple Sclerosis, or MAGNIMS) [Montalban X. et al., 2010]. This group demonstrated that in applying these less restrictive criteria, diagnostic

sensitivity was increased without compromising specificity, with fewer required MRI examinations. As was shown by MAGNIMS, even a single MRI study with both enhancing and non-enhancing lesions on the T1 sequence can be used in confirming DIT (Table 2).

Table 2. McDonald revisions from 2001 by Milo R. et al., 2014

	2001	2005	2010
MRI DIS	≥3 of: • ≥9 T2 lesions or ≥1 enhancing lesion • ≥3 periventricular lesions • ≥1 juxtacortical lesion • ≥1 infratentorial lesion 1 cord lesion can replace I brain lesion	≥3 of: • ≥9 T2 lesions or ≥1 enhancing lesion • ≥3 periventricular lesions • ≥1 juxtacortical lesion • ≥1 infratentorial lesion Any number of cord lesions can be included	≥1 T2 lesions in ≥2 of the following areas: • Periventricular • Juxtacortical • Infratentorial • Spinal cord
MRI DIT	≥1 enhancing asymptomatic lesion ≥3 months after CIS onset ≥1 new T2 lesion on a scan obtained ≥3 months after CIS	≥1 enhancing asymptomatic lesion ≥3 months after CIS onset ≥1 new T2 lesion on a scan obtained ≥30 days after CIS	Asymptomatic enhancing and nonenhancing lesions simultaneously present at any time ≥1 new T2 or enhancing lesion on follow-up MRI at any time
CSF to support?	Yes	Yes — in RRMS	No

The current 2017 McDonald diagnostic criteria for MS revised the 2010 McDonald criteria regarding the role of CSF-specific oligoclonal bands in the MS diagnostic procedure, reconsidered the importance of symptomatic lesions in demonstration of dissemination in space or time in patients with supratentorial, infratentorial, or spinal cord syndromes, and emphasized the role of cortical lesions for the dissemination in space [Thompson A. et al., 2018].

Future criteria revisions may also incorporate MRI features like the 'central vein sign', and the susceptibility signal within lesions that can be incorporated in the diagnostic work-up of MS (at standard field strength) [Rovira À. et al., 2015]. High-priority areas for further research reported by Thompson A. et al. should focus on optic nerve involvement, validation in diverse populations, and incorporation of advanced imaging, neurophysiological, and body fluid markers [Thompson A. et al., 2018]. The currently available evidence is not enough to support the use of advanced MRI to establish the initial diagnosis or differential diagnosis of MS. Advanced quantitative MRI techniques and high-resolution MPRAGE sequences may increase the accuracy of MRI detection of more specific lesions and might have distinct advantages in further studies [Enzinger C. et al., 2015].

Table 3. The 2010 McDonald Criteria for Diagnosis of MS by Polman C. et al., 2010

Clinical Presentation	Additional Data Needed for MS Diagnosis
2 attacks; objective clinical evidence of 2 lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack	None
2 attacks; objective clinical evidence of 1 lesion	Dissemination in space, demonstrated by: 1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord); or Await a further clinical attack implicating a different CNS site
1 attack; objective clinical evidence of 2 lesions	Dissemination in time, demonstrated by: Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or Await a second clinical attack
1 attack; objective clinical evidence of 1 lesion	Dissemination in space and time, demonstrated by: For DIS:

(clinically isolated syndrome)	1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord); or Await a second clinical attack implicating a different CNS site; and For DIT: Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or Await a second clinical attack
Insidious neurological progression suggestive of MS (PPMS)	1 year of disease progression (retrospectively or prospectively determined) plus 2 of 3 of the following criteria: 1. Evidence for DIS in the brain based on at least 1 T2 lesion in the MS-characteristic (periventricular, juxtacortical, or infratentorial) regions 2. Evidence for DIS in the spinal cord based on 2 T2 lesions in the cord 3. Positive CSF (isoelectric focusing evidence of oligoclonal bands and/or elevated IgG index)

1.2 Disability in MS

1.2.1 Clinical measures of disability

MS clinical features are heterogeneous and may change over time, leading to significant and irreversible disability in about 50% of untreated patients after a mean time of 8.6 years [Confavreux C. et al., 2006; Lavorgna L. et al., 2014].

In clinical practice, progression of disability is most frequently evaluated with the Expanded Disability Status Scale (EDSS) [Kurtzke J.,1977] and the Multiple Sclerosis Functional Composite (MSFC) [Cutter G. et al., 1999]. EDSS is the one disability measure presently accepted by regulators as a primary trial endpoint. It is an ordinal scale ranging from 0 (normal neurologic examination) to 10 (death owing to MS) and takes into consideration an extensive range of neurological functions related to MS. The EDSS mostly relies on motor function and important milestones are requiring unilateral assistance for walking 100 m (EDSS score of 6.0), requiring bilateral assistance for walking 20 m (EDSS score of 6.5), or requiring a wheelchair most parts of the day (EDSS score of 8.0) [Vidal-Jordana A. et al., 2017] (Fig. 1).

The MSFC, unlike the EDSS, which is based on the neurological examination, consists of three objective quantitative tests of neurological and cognitive functions [van Munster C., et al., 2017]. The MSFC represents a composite measure encompassing the major clinical dimensions of arm, leg and cognitive function. Within the MSFC, there are scores for each of the three individual measures as well as a composite score, it enables direct comparison within a study, and correlates with the EDSS [van Munster C., et al., 2017, Cohen J. et al., 2012].

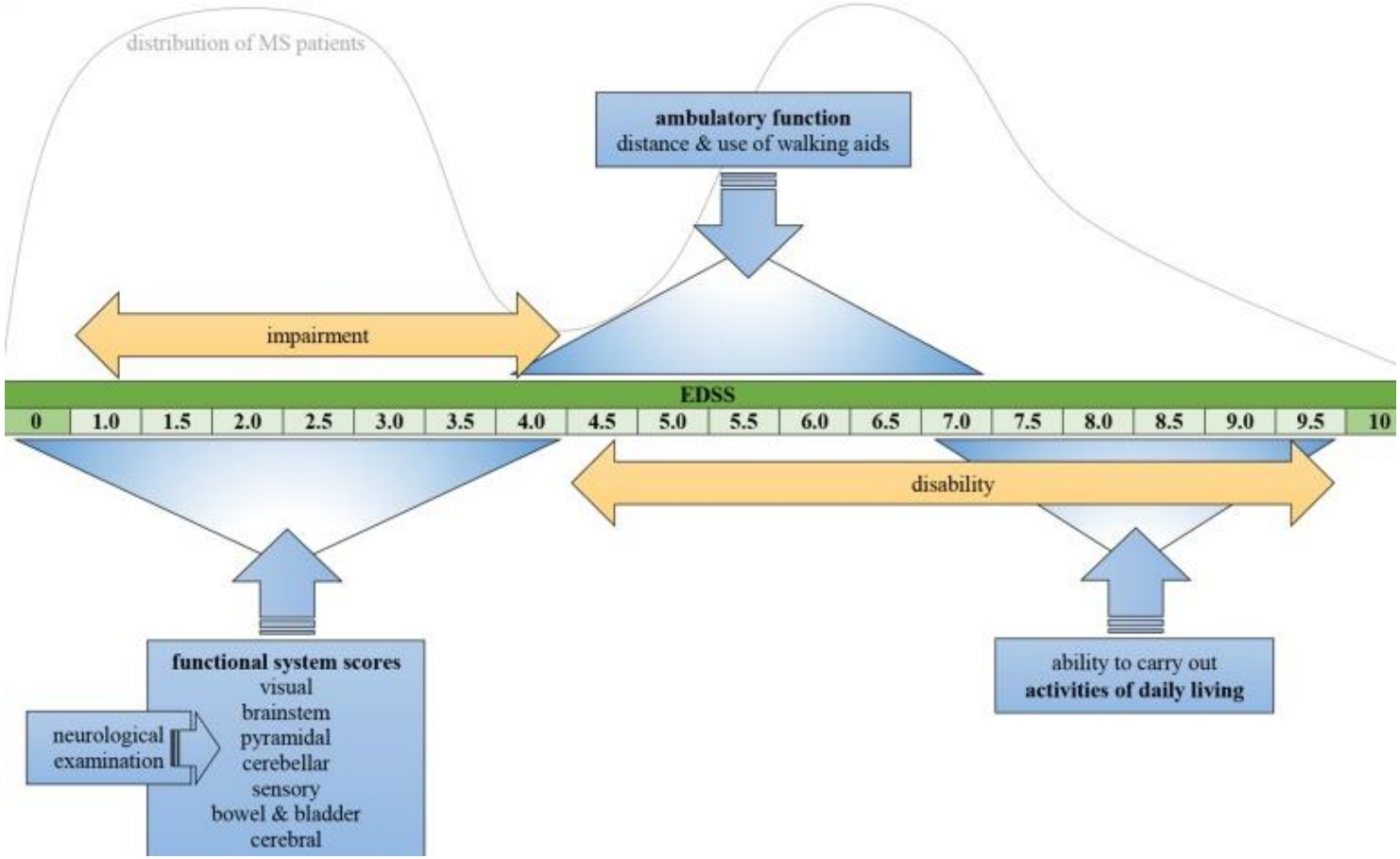


Figure 1. Schematic representation of the Expanded Disability Status Scale (EDSS) depicting the factors that determine the overall score. The graph shows the distribution of patients over the EDSS taken from van Munster C., et al., 2017.

1.2.2 Correlations of brain MRI parameters and disability

MRI is the most powerful noninvasive tool used in the MS diagnostic process and monitoring of the disease progression and also in assessment of treatment efficacy in clinical trials [Gajofatto A. et al., 2013; Wattjes M. et. Al., 2015]

The activity of disease is more often detected on MRI than by clinical relapses and in some cases MRI lesion progression occurs earlier than clinical progression [van Munster C., et al., 2017]. MS studies confirm a strong correlation between treatment effects on MRI features and frequency of relapses [Sormani M. et al., 2009]. MRI-markers are showing a great potential in predicting of the long-term disease evolution and in optimizing patient management [Fisniku L. et al., 2008]. Furthermore, demonstrating treatment efficacy on MRI lesions is important in the development and testing of immunomodulatory medications [Sormani M. et al., 2009].. However, appearance of new T2 lesions and T1 gadolinium-enhancing lesions occurs often subclinically and thus more frequent than clinical relapses [McDonald W. et al., 1994; Barkhof F. et al., 2002]. Also MS studies showed that the disease affects grey matter [Horakova D. et al., 2009; Lavorgna L. et al., 2014], what constitutes a challenge to define which type of brain damage is the most probably marker of disease progression and to monitor of the treatment efficiency of disease-modifying drugs.

Accumulation of focal white matter lesions is variable among MS patients. Decreasing of the inflammation because of the therapy, remyelination or edema resolution leads to the lesion volume reduction [Giorgio A. et al., 2014]. But usually focal lesions after relapses remain in the brain, with the disease progression lesions can merge and form confluent lesions [Giorgio A. et al., 2014]. There is an accrual of focal pathology in MS brains, with increasing of T2 lesions on average by 5%–10% per year [Paty D., et al., 1994; Giorgio A. et al., 2014]. Active (new or enlarging) T2 lesions can be a clinically relevant measure of disease activity [Wattjes M. et al., 2015; Filippi M. et al., 2016]. Contrast-enhanced T1-weighted sequences are recommended to detect acute inflammation [Wattjes M. et al., 2015].

Other expressions of MS focal neurodegeneration related with inflammation, besides T1-weighted contrast-enhancing and T2 lesions, are hypointense T1-weighted lesions "black

holes" (around 30–40% of T2 lesions) [Giorgio A. et al., 2014; Filippi M. et al., 2016]. They have MRI signal intensity equal to or lower than grey matter on T1-weighted images and persist longer than 6 months [Barkhof, F. et al., 2009; Nagtegaal G. et al., 2014; Mitjana R. et al., 2014; Giorgio A. et al., 2014; Wattjes M. et al., 2015]. They represent regions where irreversible axonal loss, demyelination, and gliosis have occurred [Filippi M. et al., 2012]. T1 lesions increase during the course of the disease. Nevertheless, results of the studies, which estimate the correlations between T1 lesion burden and disability, provided conflicting results; some found such a correlation stronger than for T2 lesions, and while others studies did not [Filippi M. et al., 2016].

Demyelinating white matter lesions became a very important MRI criterion for its diagnosis [Fisniku L. et al., 2008]. However, the accrual of focal pathology in the brain does not directly related to long-term disability progression [Gajofatto A. et al., 2013]. Appearance of new focal lesions has a prognostic value in disease progression only in patients at MS onset [Gajofatto A. et al., 2013]. Brex A. et al. [Brex A. et al., 2002] studied the relation between early lesion volume, changes in lesion volume, and long-term disability in 71 CIS patients over a period of 14.1 years. Results showed the development of MS in 88% of patients. The increases in the volume of the lesions seen on MRI in the first five years correlated with the degree of long-term disability. But this relation was only moderate, so the authors concluded, that the volume of the lesions alone may not be an adequate basis for decisions about the use of disease-modifying treatment [Brex A. et al., 2002]. And in a more advanced stage of the disease, correlation between conventional MRI measures and following disability progression is a quite weak [Sormani M. et al., 2009]. A study of Li et al. showed a plateauing relationship between T2 burden of disease and disability for EDSS values above 4.5 [Li D. et al., 2006].

Filippi M. et al. [Filippi M. et al., 1995] obtained T2-weighted scans over an interval of 24 to 36 months in 281 MS patients. Changes in disability between two examinations correlated weakly (Spearman's rank correlation coefficient = 0.13; 0.18, respectively) with the number of new and enlarging MRI lesions.

Kappos L. et al. [Kappos L. et al., 1999] investigated the prognostic value of gadolinium-enhanced MRI in a longitudinal study. They collected data from five natural-course studies and four placebo groups of clinical trials completed between 1992 and 1995. These were included a total of 307 MS patients. The relapse rate in the first year was predicted

moderate ability by the mean number of gadolinium-enhancing lesions in monthly scans during the first 6 months. The mean of gadolinium-enhancing-lesion counts in the first six monthly scans was weakly predictive of EDSS change after 1 year and 2 years and did not strong predict the development of cumulative impairment or disability [Kappos L. et al., 1999].

Enzinger C. et al. [Enzinger C. et al., 2011] investigated the long-term clinical MS course of 99 RRMS patients and the predictive value of MRI metrics after a mean of 10.8 years. Result failed to confirm a clear independent contribution for outcome of cross-sectional and short-term follow-up MRI data for the prediction of long-term disability.

Generally, MRI-visible focal both T1 and T2 lesions showed only modest correlations with clinical outcomes and disability [Chard D. et al., 2017]. The presence of this mismatch has been termed "clinico-radiological paradox" [Barkhof F., 1999; Chard D. et al., 2017]. The insufficient specificity of focal lesion burden pathology which does not distinguish edema and inflammation from irreversible demyelination and axonal loss may be the reason of the clinico-radiological paradox [Filippi M., et al. 2016]. Also it might be a consequence of the fact that many white matter lesions are clinically silent [Goodin D., 2006]. The clinico-radiological paradox emphasizes the need for the search of more sensitive biomarkers of disability progression. In the search for explanations for the clinico-radiological paradox non-conventional quantitative MRI techniques have been employed in study normal-appearing brain tissue [Enzinger C. et al., 2015]. Studies have focused on brain atrophy and showed its relevant clinical impact not only in the diagnostic phase [Grothe M. et al., 2016].

Estimation of grey matter (GM) atrophy is possible in vivo by MRI [Grothe M. et al., 2016]. Evidently, that histopathology is more robust evaluation method [Grothe M. et al., 2016]. Gilmore C. et al. [Gilmore C. et al., 2009] examined postmortem the extent and pattern of grey and white matter demyelination in the motor cortex, cingulate gyrus, cerebellum, thalamus and spinal cord in 14 MS cases and three controls. Overall, 28.8% of the grey matter was demyelinated compared with 15.6% of the white matter (WM), with demyelination being greater in the GM than in the white matter at each of the anatomical sites [Gilmore C. et al., 2009]. There was substantial variation in the extent of demyelination between the different CNS regions. GM demyelination was most extensive

in the spinal cord and cerebellum while WM demyelination was most prominent in the spinal cord [Gilmore C. et al., 2009].

However, histopathology studies are always limited in sample size and cannot generate information about the changes that occur in the early disease stages [Grothe M. et al., 2016]. Appearance of advanced MRI technologies offers the opportunity to examine grey matter atrophy in vivo. Horakova D. et al. [Horakova D. et al., 2009] assessed the relationship between grey matter and white matter atrophy and clinical status in a group of 181 RRMS patients over 5 years. Decline in percentage brain volume change (PBVC) and grey matter volume (GMV) were predictive markers of disability deterioration. The change in EDSS over 0–24 months was best predicted by PBVC 0–24 months (R2=0.069 p=0.002) and GMV 0–24 months (R2=0.028 p=0.01). For the 5-year EDSS change, the best predictors were PBVC 0–24 months (R2=0.489 p=0.001), PBVC 0–60 months (R2=0.223 p=0.002). Correlation of T2-lesion volume with clinical status was weaker and decreased over time (for T2-lesion volume 0–24 months R2=0.207 (p=0.01) and for T2-lesion volume 0–60 months R2=0.105 (p=0.036))[Horakova D. et al., 2009].

The aim of the nine-year follow-up study by Lavorgna L. et al. [Lavorgna L. et al., 2014] was to identify clinical or MRI predictors of long-term clinical progression in 241 MS patients. Results showed that baseline (BL) GMV and EDSS were the best long-term predictors of disease progression in RRMS patients with a relatively long and mild disease (respectively: (OR 2.88; CI 1.9–4.36), (OR 2.7; CI 1.7–4.2), (HR 3.86; CI 1.94–7.70)).

At the same time Jacobsen C. et al. [Jacobsen C. et al., 2014] also tried to identify MRI biomarkers of long-term disability progression in 81 MS patients after 5 and 10 years of follow-up. The study showed that whole brain, cortical and putamen atrophy occurs throughout the 10-year follow-up of this MS cohort and is more pronounced in the group that showed disability progression at 5 years (whole brain (-3.8% vs -2.0%, $p<0.001$), cortical (-3.4% vs -1.8%, p=0.009) and putamen volume changes (-10.6% vs -3.8%, p=0.003)), but not at 10 years of follow-up. Overall, GM atrophy showed better association with disease progression than WM atrophy over 5-year and 10-year follow-up [Jacobsen C. et al., 2014].

Advanced MRI techniques, which are used to explain the different changes of MS, may help determine the clinical evolution of the disease.

1.2.3 Measures of disease activity

Advances in treatment options have led to redefinition of treatment goals. In the course of post-hoc analyses of data from the AFFIRM study [Havrdova E. et al., 2009] the concept "disease activity-free status" has been developed which afterwards was re-termed "no evidence of disease activity" (NEDA). NEDA originally has been defined by three measures of disease activity: 1) no relapses, 2) no clinically significant increase in EDSS, and 3) no radiological activity (new or enlarging T2 lesions or gadolinium enhanced lesions) [Havrdova E. et al., 2010]. NEDA-3 has become an important outcome measure in clinical trials in MS since then.

Rotstein et al. [Rotstein D. et al., 2015] reported data from a cohort study on the long-term persistence of NEDA-3 in 219 patients with a diagnosis of a clinically isolated syndrome or RRMS, who had at least 7 years of follow-up. 46.0% of patients had NEDA for clinical and MRI measures at 1 year, but only 7.9% maintained NEDA status after 7 years, despite disease-modifying treatment (DMT). NEDA-3 at 2 years had a positive predictive value of 78.3% for no progression (EDSS score change 0.5) at 7 years.

De Stefano et al. [De Stefano N. et al., 2014] showed that the mean annual rate of brain volume loss (BVL) in MS patients is typically 0.5%–1.35% per year and it is faster than in comparable age-adjusted healthy individuals (0.1%–0.3%). Kappos L. et al. [Kappos L. et al., 2015] proposed that NEDA-3 is weighted towards inflammatory activity, while neurodegeneration mechanisms of the disease are captured to a lesser extent by this concept. For a more comprehensive assessment of disease activity and progression he introduced a new four-parameter measure, termed NEDA-4, considering also the absence of BVL exceeding an annual rate of 0,4%. He and co-workers analyzed data from two DMTs trials in 700 RRMS patients and assessed NEDA-4 using different annual BVL mean rate thresholds (0.2%–1.2%). At 2 years, 31.0% of patients receiving DMT achieved NEDA-3, adding BVL (threshold of 0.4%); the respective proportion of patients achieving NEDA-4 was 19.7%.

In summary, recent studies proposed to include BVL into NEDA to add significance in the evaluation of the disease prognosis.

1.3 Goals of the study and hypothesis

Focal white matter lesions, “black holes” and brain atrophy are the classic morphologic MRI hallmarks of MS [Filippi M. et al., 2016]. There are different methodologies used to assess these measures. Overall, white matter lesions and “black holes” are rated using manual, semi-automatical or automatical computational methods. Brain atrophy is generally measured on two-dimensional (2D) / three-dimensional (3D) T1-weighted images and it is analyzed using cross-sectional methods comparing patients to controls (e.g., FreeSurfer, SIENAX, voxel-based morphometry) as well as longitudinal methods (e.g., SIENA, MSmetrix) [Giorgio A. et al. 2008; Lysandropoulos A. et al., 2016]. However, in routine clinical practice, especially for the longitudinal assessment of the disease activity and response to immunomodulatory treatment, it is not always possible to obtain MRI examinations using entirely standardized MR protocol, since in clinical practice both 1.5T and 3T MRI systems are used with 2D and 3D MRI data.

Nevertheless, the reliability of the brain volume measures between 2D and 3D datasets has received relatively little systematic investigation. Despite a sufficient number of studies comparing brain metrics from different scanners, we found only two systematic comparisons between 2D and 3D segmentation methods. First work from Amann M. et al. [Amann M. et al., 2015] evaluated the possibility of segmenting deep grey matter structures using standard 2D T1-weighted MRI. In a cohort of 70 MS patients, both 2D and 3D T1-weighted data were acquired. The accuracy and reliability of the 2D data segmentation were compared with the respective results of 3D segmentations using volume difference, volume overlap and intra-class correlation coefficients. Results demonstrated that subcortical segmentation of 2D data is feasible. The larger subcortical GM structures can be segmented with high consistency [Amann M. et al., 2015].

The second investigation was done by Vidal-Jordana A. et al. [Vidal-Jordana A. et al., 2017]. Ninety-one GM structures (70 cortical and 21 subcortical) of patients with RRMS were obtained and used for 2D spin-echo (SE) versus 3D gradient-echo (GE) reliability assessments. Cortical thickness was estimated using FreeSurfer software and subcortical structures in all 2D-SE and 3D-GE sequences by FIRST software. Subcortical volumes obtained with both FreeSurfer and FIRST showed good agreement between 2D-SE and 3D-GE images. Cortical thickness estimates with FreeSurfer on 2D-SE images were inaccurate. Also SIENAX software was used to calculate the normalized brain volume from 2D-SE and 3D-GE sequences in a small subset sample (n=15). The intraclass

correlation coefficient value of the normalized brain volume estimate between 2D-SE and 3D-GE T1-weighted images was 0.85 (almost perfect agreement) [Vidal-Jordana A. et al., 2017].

In the study of Vidal-Jordana A. et al. [Vidal-Jordana A. et al., 2017] subcortical volumes and normalized brain volume were estimated, while Amann M. et al. [Amann M. et al., 2015] evaluated deep grey matter structures. However, we were more interested in the possibility of data comparison irrespective of deep brain structures, but regarding global brain volume, grey and white matter volume, not only cross-sectionally, but also longitudinally.

To date, the performance of the most recent versions of SIENA and SIENAX software in MS datasets with different MRI acquisition techniques (such as different field strengths, 2 and 3D segmentation) is not well studied. We hypothesized that protocol standardization is a necessary requirement of reliable brain volume metrics measures cross sectionally as well as longitudinally in MS studies.

In this study we thus collected datasets from two MS centers (Linz and Graz) equipped with MR scanners with the different MRI acquisition techniques (1,5T and 3T field strengths with 2D and 3D segmentation, respectively), but from cohorts with similar clinical characteristics.

The goals of the present study are:

1. to test the reliability of lesion load and brain volume metrics cross-sectionally using SIENAX and longitudinally using SIENA software on the two real-world MS datasets from two centers with protocols not coordinated in clinically similar cohorts of MS patients;

2. to investigate the influence of the differences in scanner parameters of 2D and 3D MR-images on the brain volume characteristics by MS cross-sectionally and longitudinally;

3. to demonstrate the potential use in clinical practice of not standardized protocols regarding the measurement of lesion load and brain volumes metrics in MS patients with a particular focus on brain volume changes over time.

2. Materials and Methods

2.1 Subjects and clinical data

Data came from 174 MS patients (112 female and 62 male; mean age at the baseline 33,12 $\pm$ 9,28 years). The data were acquired at two MS centers (Medical Universities of Graz and Linz) between 2000 and 2014. For all patients disease severity at the baseline and follow-up was measured using EDSS at the baseline and follow-up [Kurtzke J., 1983]. Additional data included the age at the disease onset, disease duration, MS phenotype and treatment status (use of disease-modifying therapy [DMT] at the time of the study). Patients were categorized as RRMS (130 patients) and CIS (44 patients), fulfilling the revised McDonalds criteria 2005. This study was approved by the ethics committee of the Medical Universities of Graz and Linz.

2.2 MRI

2.2.1 MRI protocol

MRI scans of the brains from the Graz cohort were acquired at a 3.0 Tesla scanner (Siemens Tim Trio, Siemens Healthcare, Erlangen, Germany) using the following scan parameters: a fluid-attenuated inversion recovery (FLAIR) sequence (where the repetition time (TR) / echo time (TE) / inversion time (TI) = 9000 ms / 69 ms / 2500 ms; with slice thickness of 3 mm); and a T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE) sequence, with 1-mm isotropic resolution (where TR / TE / TI / flip angle = 1900 ms / 2.19 ms / 900 ms / 9°).

MRI scans of the Linz Cohort were performed at two 1.5 Tesla scanners (Avanto and Symphony, Siemens Healthcare, Erlangen, Germany) with the following scan parameters: T2-weighted 2D turbo spin echo (TSE) sequence (where TR / TE / flip angle = 3620 ms / 80 ms / 180°) and a 2D T1-weighted spin echo (SE) sequence (where TR / TE / flip angle = 472 ms / 14 ms / 90°); with slice thickness of 6 mm.

An overview of the MRI protocols used is presented in Table 4.

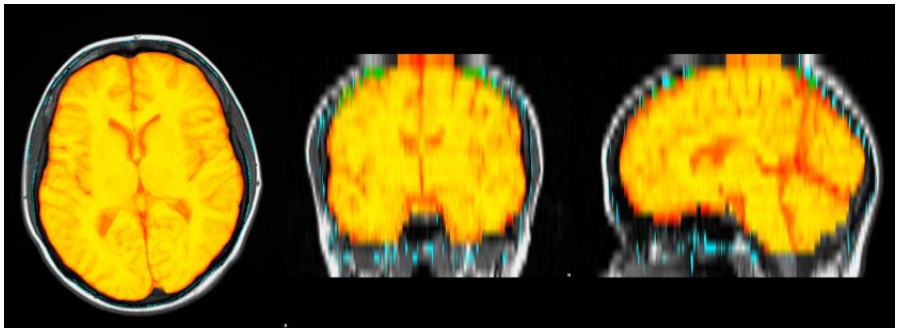
Table 4. MRI Acquisition

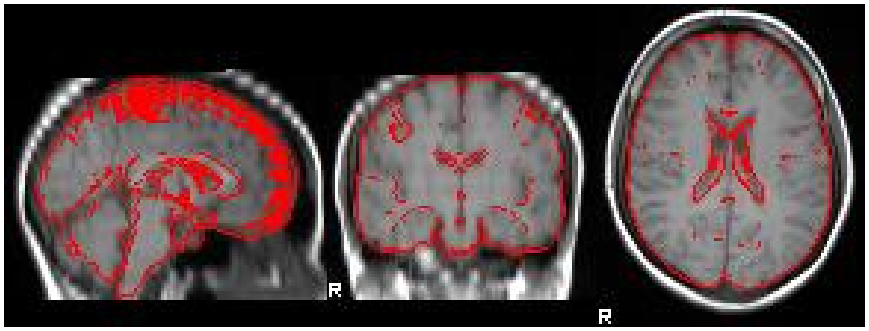
Scanner	Sequence	Number of scans		TR	TE	flip angle	slice thickness
		BL	FU				
Graz cohort Siemens Tim Trio 3T	MPRAGE	111	111	1900	2.19	9	1
	FLAIR	107	104	9000	69		3
Linz cohort Avanto 1,5T	T2 2D TSE	34	36	3620	80	180	6
	T1 2D SE tra	37	37	472	14	90	6
Linz cohort Symphony 1,5T	T2 2D TSE	23	26	3620	80	180	6
	T1 2D SE tra	26	26	491	14	90	6

2.2.2 Image analysis

T1 and T2-lesion loads (T2-FLAIR images for patients from Graz and T2 images for patients from Linz) were evaluated using a custom-written Interactive Data Language program (DispImage; Exelis Visual Information Solutions, Boulder, Colorado) [Plummer D., 1992]. T1-weighted lesions ("black holes") were defined as non-enhancing lesions that appear hypointense on T1-weighted images with a signal intensity below cortex which were concordant with hyperintense lesions on a T2-weighted image [Thaler C. et al., 2015].

Global brain volume (GBV) (including separate estimates of white matter volume (WMV), grey matter volume (GMV), peripheral grey matter (PGM) and ventricular cerebrospinal fluid (VCSF)) were analyzed for each patient using the T1-w images with the fully automated Structural Image Evaluation, using Normalization, of Atrophy/ cross-sectional (SIENAX) (Image Analysis Group, Oxford, UK) [Smith S. 2002], part of FSL [Smith S. et al., 2004]. Brain volume was normalized for subject head size. Longitudinal changes of the GBV between BL and FU MRI scans were measured using SIENA [Smith S. et al., 2001, 2002], part of FSL [Smith S. et al., 2004], method to obtain the percentage brain volume change (PBVC) (Fig.2,3).


A


B

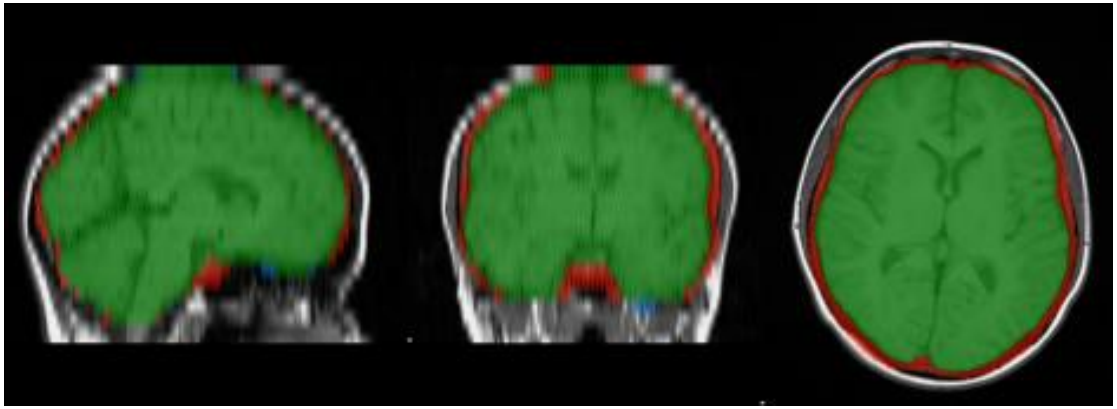

C

Figure 2. SIENA/SIENAX.

BET brain extraction results (A);

FLIRT standard space registration results (B);

Field-of-view and standard space masking. Red shows the standard-space-based brain mask combined with the field-of-view mask (if used). Blue shows the original BET-derived brain mask. Green shows the intersection of the two (C).

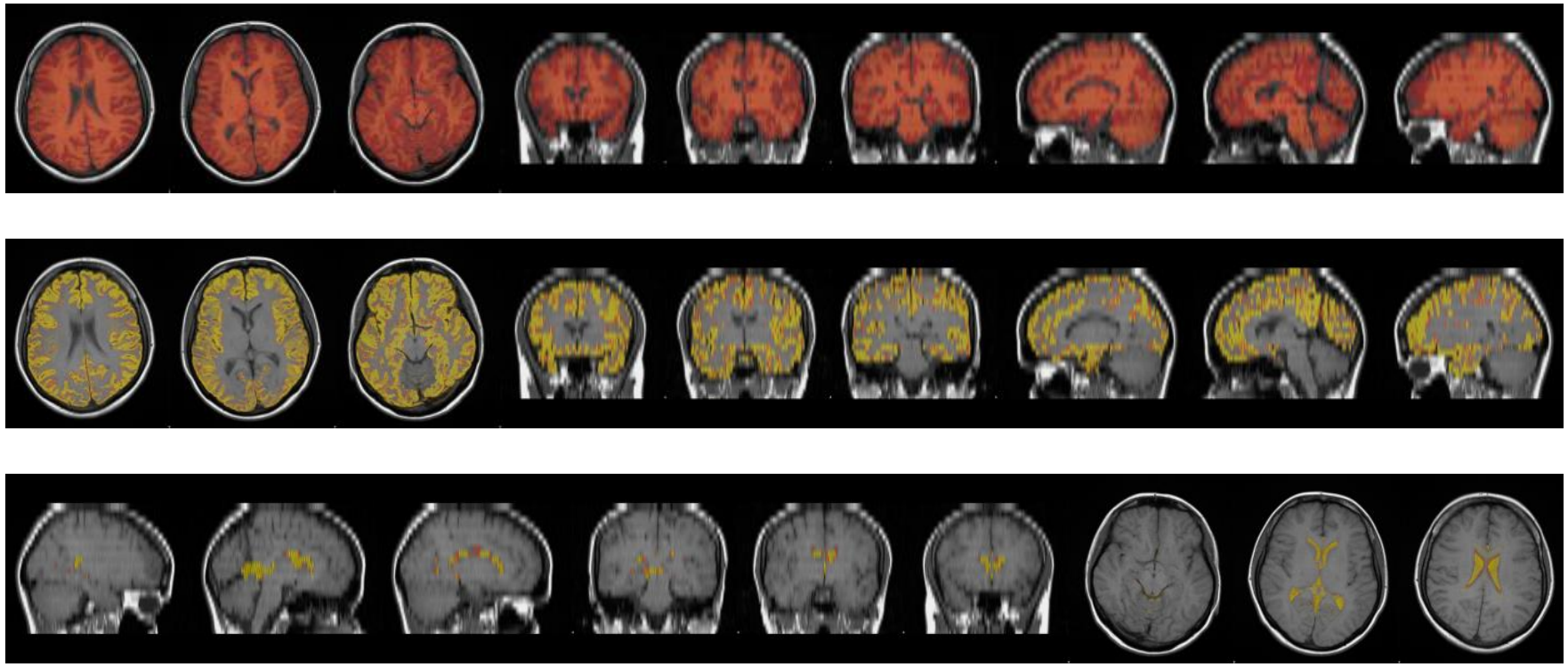


Figure 3. Final SIENAX segmentation results.

Whole-brain segmentation (D);

Peripheral cortex masked segmentation (E);

Ventricle masked segmentation (F).

2.3 Measurement of Variability

For the measure of the brain volumes cohort's variability was used coefficient of variation (CoV), which was defined as ratio of the standard deviation to the mean, multiplied by 100 in case of the Graz cohort. CoV of the Linz cohort, because of the using different scanners, was estimated as ratio of the pooled standard deviation (defined as the square root of average variance) and the overall grand mean measured by the two scanners and was expressed in percentages. Low CoV between scanners would be considered to have a high similarity of volume measurements obtained at each scanner.

Also for the variability measure we used the mean absolute percentage difference. It is a measure of statistical dispersion equal to the average absolute difference of two independent values drawn from a probability distribution. A related statistic is the relative mean absolute difference, which is the mean absolute difference divided by the arithmetic mean. It is calculated as the difference between two values divided by the average of the two values shown as a percentage.

Mean absolute percentage differences in volume between the two cohorts were calculated for each brain volume metric in all patients according to:

$$\Delta V = 2 * (V(2D) - V(3D)) / V(2D) + V(3D)$$

V(2D): mean volume in the 2D data from Linz cohort;

V(3D): mean volume in the 3D data from Graz cohort.

With this definition, ΔV is the difference between the two volumes normalized to the mean volume of each method [Amann M. et al., 2015].

2.4 Statistical analysis

All statistical analyses were performed using SPSS (Statistical Package for the Social Sciences) for Windows, Version 20.0 (SPSS Inc, Chicago, IL, USA). Kolmogorov-Smirnov test was used to assess normal distribution of the variables. To describe the normally distributed variables mean and standard deviation were used, for the skewed distributed variables medians and interquartile range (IQR) were used. For normally distributed variables the Student t-test was applied, for skewed distribution variables we used the Mann–Whitney U-test and for categorical variables the Pearson's $\chi 2$ test.

3. Results

3.1 Demographics and clinical characteristics

Demographics of patients in each participating center are summarized in Table 5. For the whole group, patients had a mean age of 33,12±9,28 years; there were 112 females and 62 males, of whom 130 had RRMS and 44 CIS, with a mean baseline disease duration of 3,79±9,38 years, median EDSS at the baseline of 1,5 (range 0 – 4,5) and the mean interval between baseline and follow-up 3,42±1,58. EDSS score is missing for five patients; one patient did not have information about DMT at the baseline and two at follow-up. Subjects in our study were similar concerning disease disability (EDSS at the baseline and follow-up), disease duration, MS phenotype, DMT and age (Table 5, Fig.4).

Table 5. Demographics and clinical characteristics at baseline and follow-up

Clinical characteristics	Graz	Linz	p-value
Number of patients	111	63	
Baseline age, y, mean ± SD	32,74±8,53	33,79±10,51	0,476[c]
Disease duration at the baseline, y, mean ± SD	3,97±5,65	3,46±13,75	0,734[c]
Interval between baseline MRI and follow-up MRI, y, mean ± SD	3,62±1,28	3,07±1,89	0,025[c]
Interval between first and last EDSS, y, mean ± SD	3,62±1,28	3,07+1,89	0,025[c]
RRMS, n (%)	81 (73,0)	49 (77,8)	0,067[a]
CIS, n (%)	30 (27,0)	14 (22,2)	0,067[a]
First EDSS *	1,5 [1;2]	1 [1;2] N=57	0,379[b]
Last EDSS *	1,5 [0;2]	1,5 [1;1,75] N=57	0,504[b]
DMT at baseline, n (%)	48 (43,2)	42 (67,7) N=62	0,138[a]
DMT at follow-up, n (%)	42 (37,8)	34 (55,7) N=61	0,351[a]

*Data reported as median [IQR]; a= Chi-square-test; b= Mann–Whitney U-test; c= Student t-test; DMT, disease-modifying treatment; EDSS, Expanded Disability Status Scale; RRMS, relapsing–remitting multiple sclerosis; CIS: clinically-isolated syndrome.

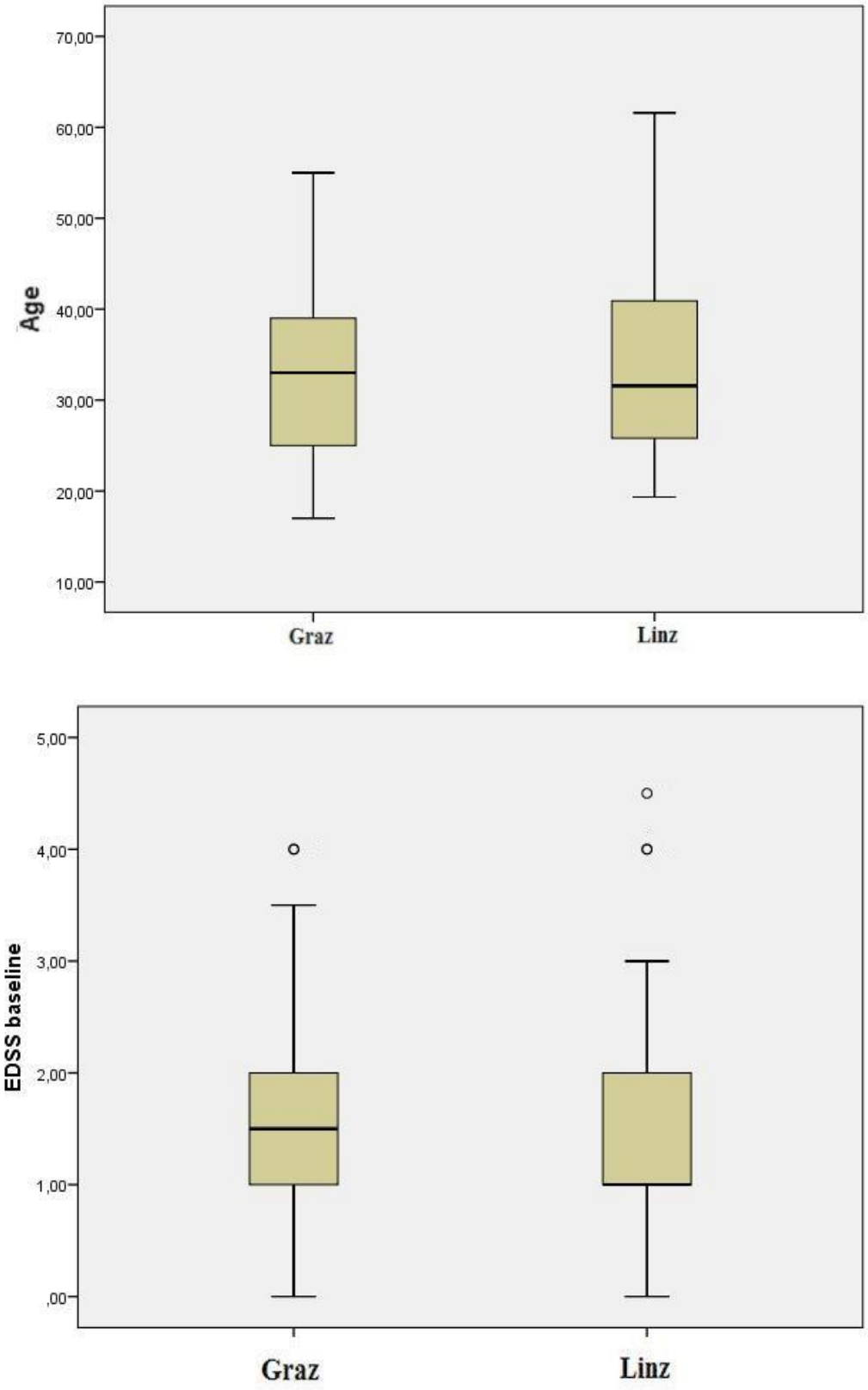


Figure 4. Age and EDSS at the baseline in two cohorts. Box plots show no significant difference between two cohorts in age, p=0,476 (top) and disease disability, p= 0,379 (bottom).

3.2 MRI characteristics

Results of lesion load and brain metrics are presented in Tables 6 and 7. 10 patients did not have T2 scans at the baseline, 8 at the follow-up. In spite of similar T2 lesion load in two cohorts, T1 lesion load was different (Table 6). This fact also could be explained by the different slice sickness of MRI scans between Graz and Linz cohorts. Results from SIENAX (Table 7) showed that GBV was similar in two cohorts, while the other brain metrics (WMV, GMV; PGM, VCSF) varied significantly. Such a large variation in brain volume characteristics by patients with similar disease duration, disability, clinical features, lesion load and GBV indicated the importance of standardized MRI protocols.

Table 6. Baseline and follow-up lesion load characteristics

MRI characteristics	**Graz**	**Linz**	p-value**
T2 lesion load at baseline, ccm*	1,61 [0,59;6,38] N=107	1,73 [0,73;4,85] N=57	0,816
T2 lesion load at follow-up, ccm*	1,74 [0,73;5,21] N=104	1,93 [0,87;4,89] N=62	0,881
Annualized change T2 lesion load**	-0,003	0,050	0,034
T1 lesion load at baseline, ccm *	**0,35[0,16;1,23]**	**0 [0;0,09]**	<0,001
BHR at baseline, % *	**24,24 [12,65;39,16]**	**0 [0;4,15]**	<0,001

*Data reported as median [IQR] ;**Mann–Whitney U-test; BHR- black hole ratio
**follow-up interval for Graz = 3,62+1,28 years, for Linz = 3,07+1,89 years

Table 7. Baseline and follow-up brain volume characteristics

MRI characteristics*	Graz	Linz	p-value**
GBV at baseline	1562661,32 [1513094,01;1615575,29]	1550696,48 [1498364,90; 1591766,49]	0,339
GBV at follow-up	1547312,88 [1489477,98;1600353,15]	1521628,32 [1485716,85; 1585175,02]	0,245
Annualized change GBV	-5172,25	-4054,02	0,965
WMV at baseline	**768348,61 [740866,72;793179,44]**	**854483,96 [823782,66; 872029,73]**	<0,001
WMV at follow-up	**762447,46 [730925,09; 792440,16]**	**839879,89 [823085,07; 859087,45]**	<0,001
Annualized change WMV	-1727,33	-1864,85	0,980
GMV at baseline	**797321,25 [760390,22; 820951,39]**	**705944,05 [673880,35; 728773,89]**	<0,001
GMV at follow-up	**783621,25 [752014,22; 809688,74]**	**685926,74 [657648,29; 718953,33]**	<0,001
Annualized change GMV	-4562,06	-4027,62	0,952
PGM at baseline	**650634,37 [622981,03; 676933,28]**	**564963,28 [535561,63; 587097,76]**	<0,001
PGM at follow-up	**640136,98 [610502,36; 664308,22]**	**552685,59 [526331,33; 578640,72]**	<0,001
Annualized change PGM	-3789,59	-3419,36	0,967
VCSF at baseline	**34076,61 [26154,54; 42638,32]**	**29345,36 [23046,63; 38415,16]**	0,037
VCSF at follow-up	**37624,01 [30999,65; 47121,58]**	**30956,84 [24299,31; 40481,73]**	0,003
Annualized change VCSF	**852,36**	**324,62**	0,023

* Data reported as median [IQR]; GBV- global brain volume; WMV- white matter volume; GMV- grey matter volume; PGM- peripheral grey matter; VCSF- ventricular cerebrospinal fluid;
**Mann–Whitney U-test

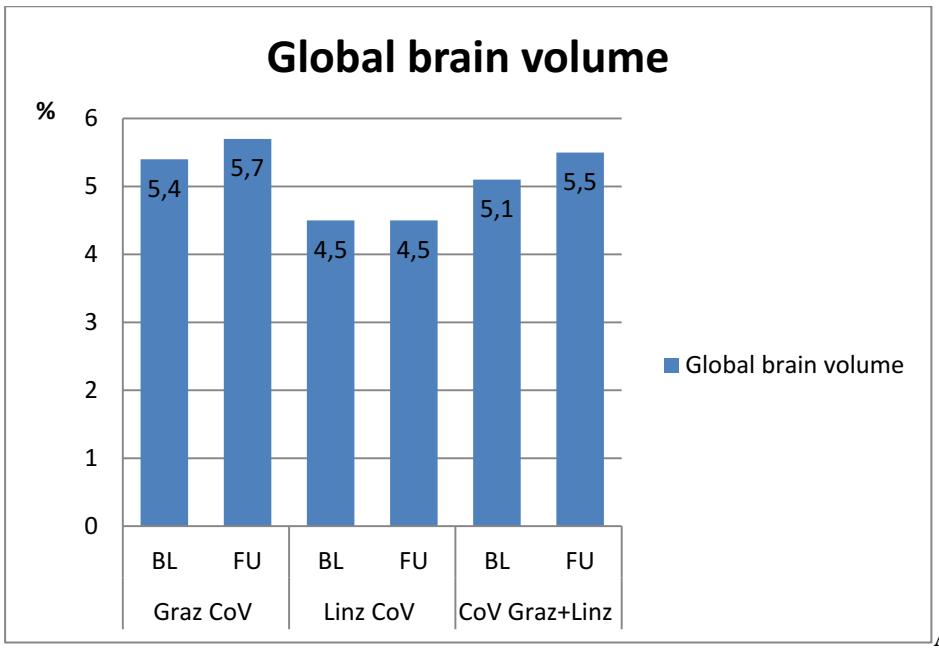
3.3 Variability

Table 8 gives an overview of the CoV in Graz and Linz cohorts and in two centers together. As can be seen, the calculated CoV within every of two cohorts was very similar. CoV of two cohorts together was slightly higher in grey, white and peripheral grey matter volume and considerably higher for ventricular cerebrospinal fluid for the all cohorts, but the variation was lowest in global brain volume (Fig. 5).

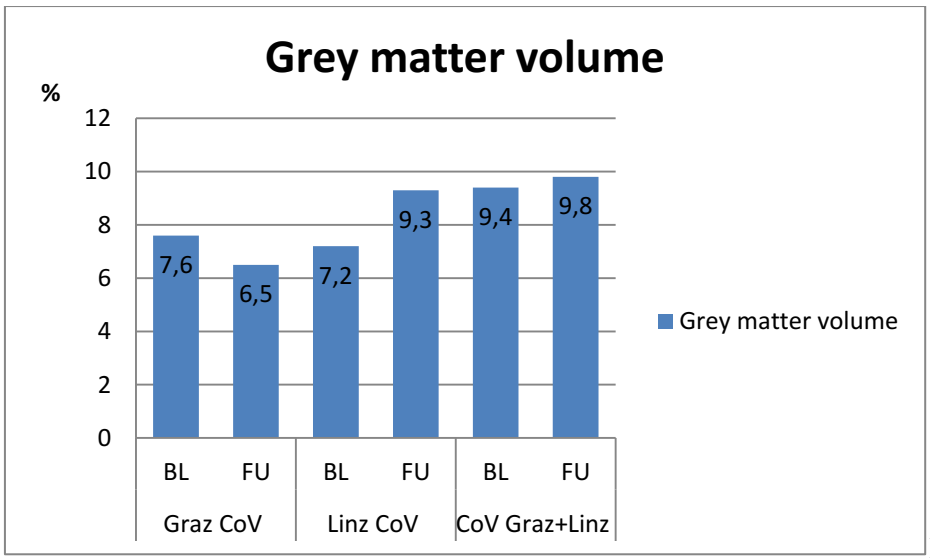
Table 8. Coefficient of variation of brain volume metrics

Brain volume metrics	Graz CoV (%)		Linz CoV (%)		CoV Graz+Linz (%)	
	BL	FU	BL	FU	BL	FU
Number of MRI scans	111	111	63	63	174	174
Global brain volume	5,4	5,7	4,5	4,5	5,1	5,5
Grey matter volume	7,6	6,5	7,2	9,3	9,4	9,8
White matter volume	6,6	6,1	3,8	3,7	7,4	7,3
Peripheral grey matter	8,6	7,2	8,1	10,1	10,6	10,8
Ventricular cerebrospinal fluid	46,5	46,6	39,6	41,7	45,0	47,1

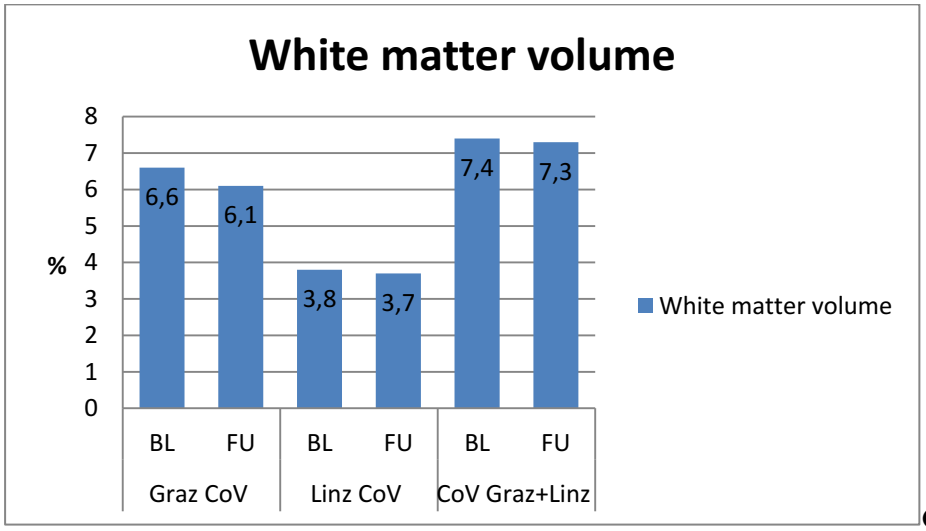
CoV=coefficient of variation


Global brain volume
%
6
5
4
3
2
1
0
5,4
5,7
4,5
4,5
5,1
5,5
BL
FU
BL
FU
BL
FU
Graz CoV
Linz CoV
CoV Graz+Linz
Global brain volume

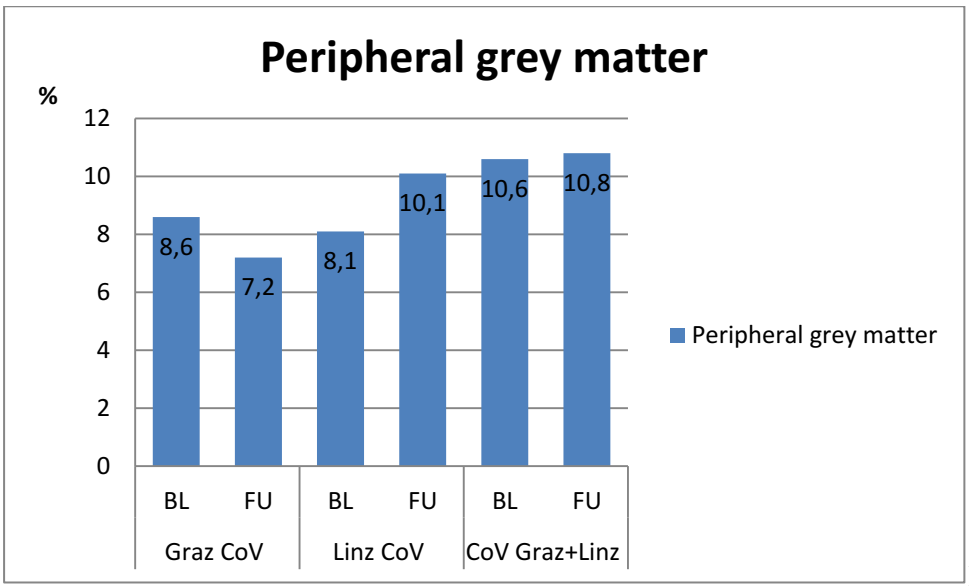
A


Grey matter volume
%
12
10
8
6
4
2
0
7,6
6,5
7,2
9,3
9,4
9,8
BL
FU
BL
FU
BL
FU
Graz CoV
Linz CoV
CoV Graz+Linz
Grey matter volume

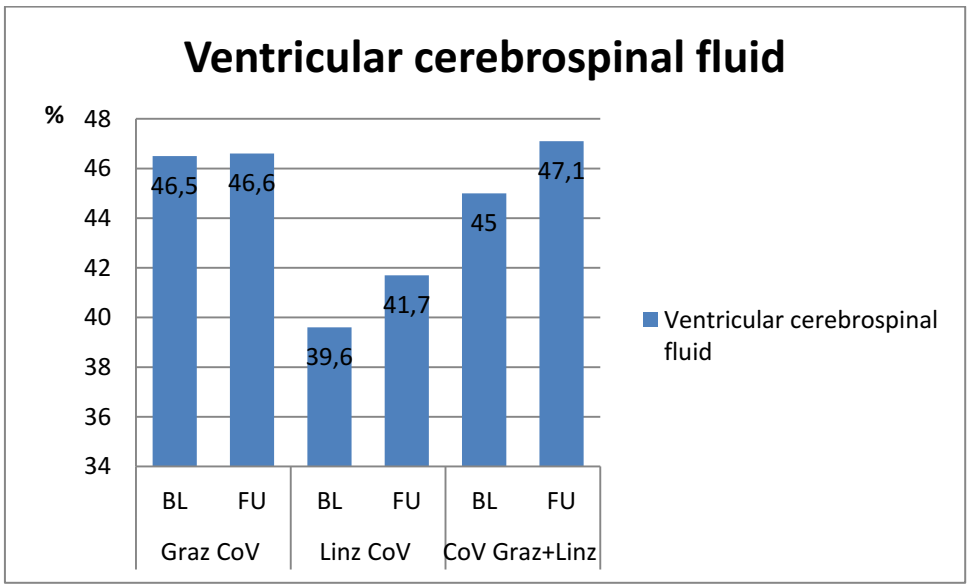
B


White matter volume
8
7
6
5
4
3
2
1
0
%
6,6
6,1
3,8
3,7
7,4
7,3
BL
FU
BL
FU
BL
FU
Graz CoV
Linz CoV
CoV Graz+Linz
White matter volume

C



D



E

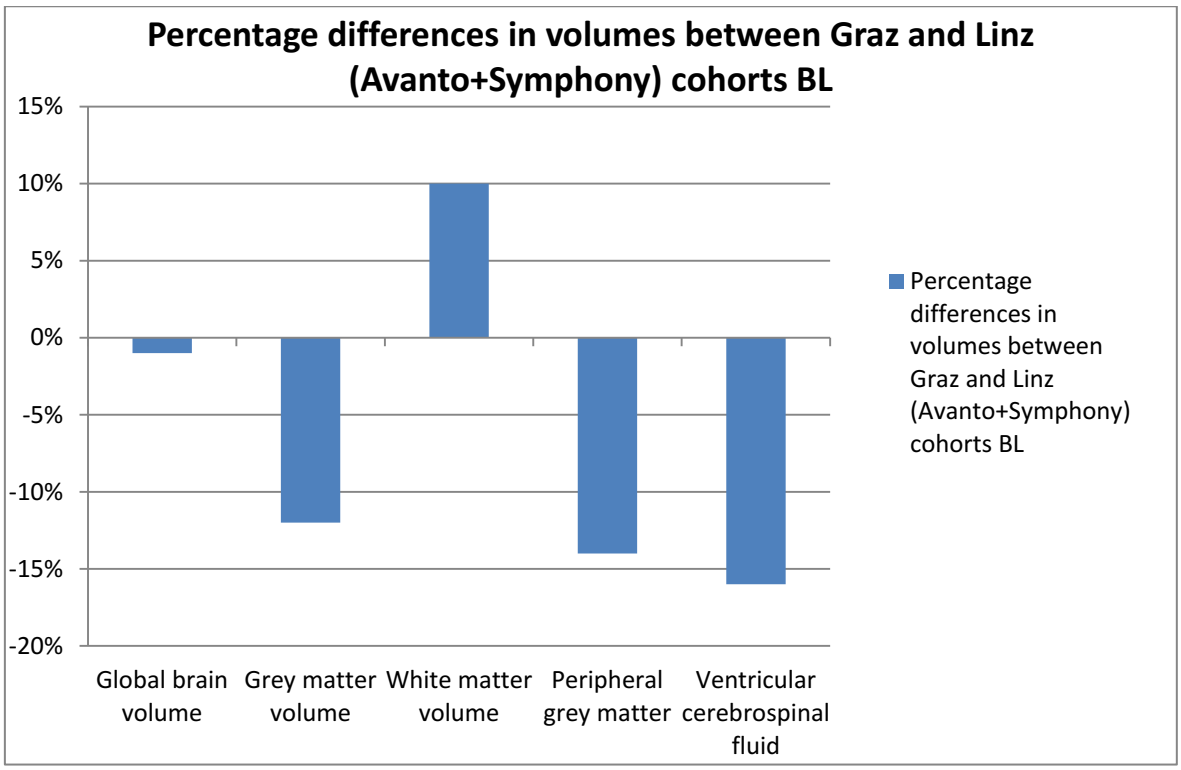
Figure 5. Coefficient of variations (CoV) of brain volume metrics in Graz and Linz cohorts. CoV in both centers was similar within cohort, CoV of two cohorts together was low in global brain volume (A) and higher in grey matter volume (B), white matter volume (C) and peripheral grey matter (D), and considerably higher for ventricular cerebrospinal fluid (E).

Table 9 demonstrates the percentage volume difference for each brain metric at the BL and FU which is necessary for detecting a significant volume difference or change between two cohorts. According to the table, ratio between grey and white matter was completely different (-12% and 10% between centers in grey and white matter, respectively). This observation can also be confirmed from the Fig.6.

Table 9. Percentage differences in volumes of brain volume metrics

Brain volume metrics	Percentage differences in volumes between Avanto and Symphony MR scanners from Linz cohort*		Percentage differences in volumes between Graz and Linz (Avanto+Symphony) cohorts **	
	BL	FU	BL	FU
Global brain volume	1%	1%	-1%	0%
Grey matter volume	-1%	1%	-12%	-12%
White matter volume	2%	0%	10%	10%
Peripheral grey matter	-1%	1%	-14%	-14%
Ventricular cerebrospinal fluid	-6%	11%	-16%	-19%

*Positive difference means that the brain volume metric is larger in Linz Avanto cohort.
**Positive difference means that the brain volume metric is larger in Linz (Avanto+Symphony) cohort.



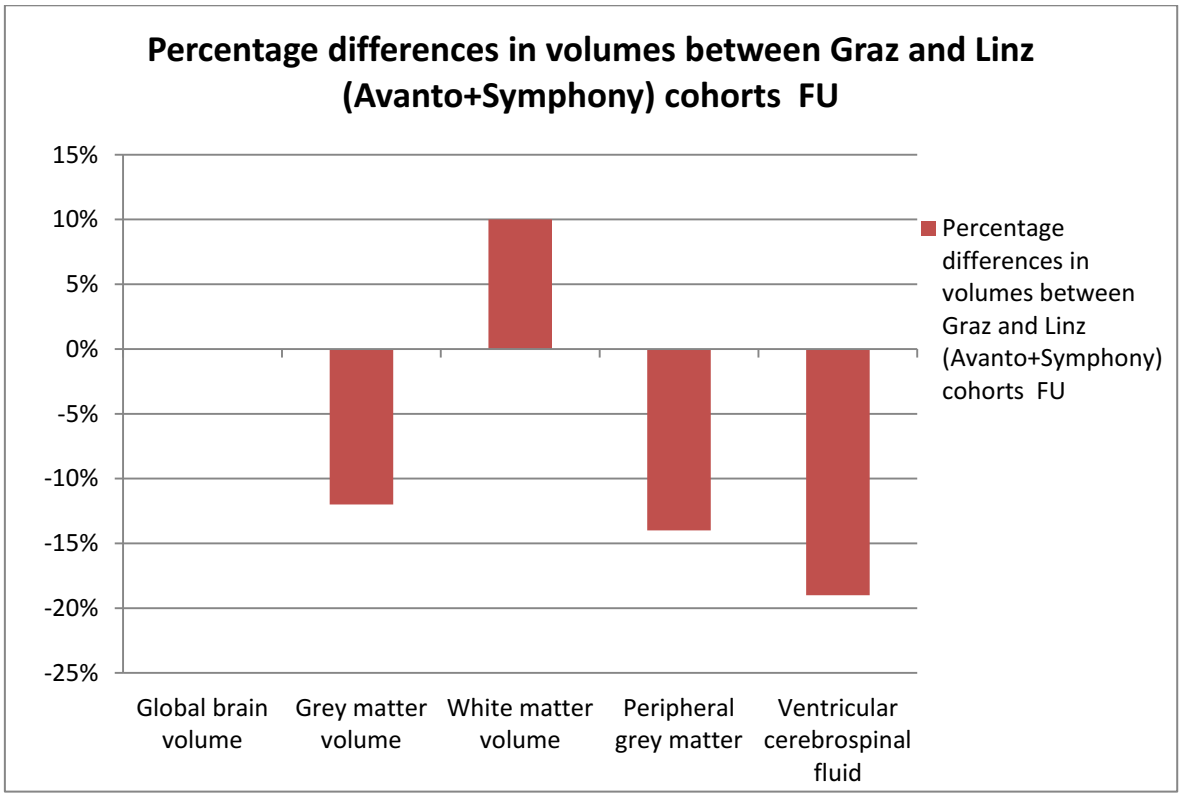


Figure 6. Percentage differences in volumes of brain volume metrics BL (top) and FU (bottom) in Graz and Linz cohorts. White matter volumes at BL and FU highly varied from the grey, peripheral grey matter volumes and ventricular cerebrospinal fluid by similar global brain volumes between centers. Positive difference means that the brain volume metric is larger in Linz (Avanto+Symphony) cohort.

Then, concerning the importance in determining the MS progression, it was important for us to understand whether the progression of "atrophy" (i.e. the magnitude of brain volume changes) was different in patients from two centers. By using SIENA to assess the impact of measurement variability on longitudinal volumetric studies we determined the PBVC in two cohorts, which is necessary for detecting a significant volume change between two volume measurements in a period of time. Figure 7 and Table 10 show results from longitudinal SIENA analyses and demonstrated percentage brain volume change (PBVC) between two centers (for 111 patients from Graz and 56 patients from Linz cohort). PBVC varied significantly between the two cohorts.

Table 10. Percentage brain volume change between two centers

	Graz	Linz	p**
N of patients	111	56	
PBVC*	-0,39 [-0,64; -0,18]	-0,19 [-0,40; -0,05]	<0,0001

* Data reported as median [IQR]; **Mann–Whitney U-test; PBVC=percentage brain volume change

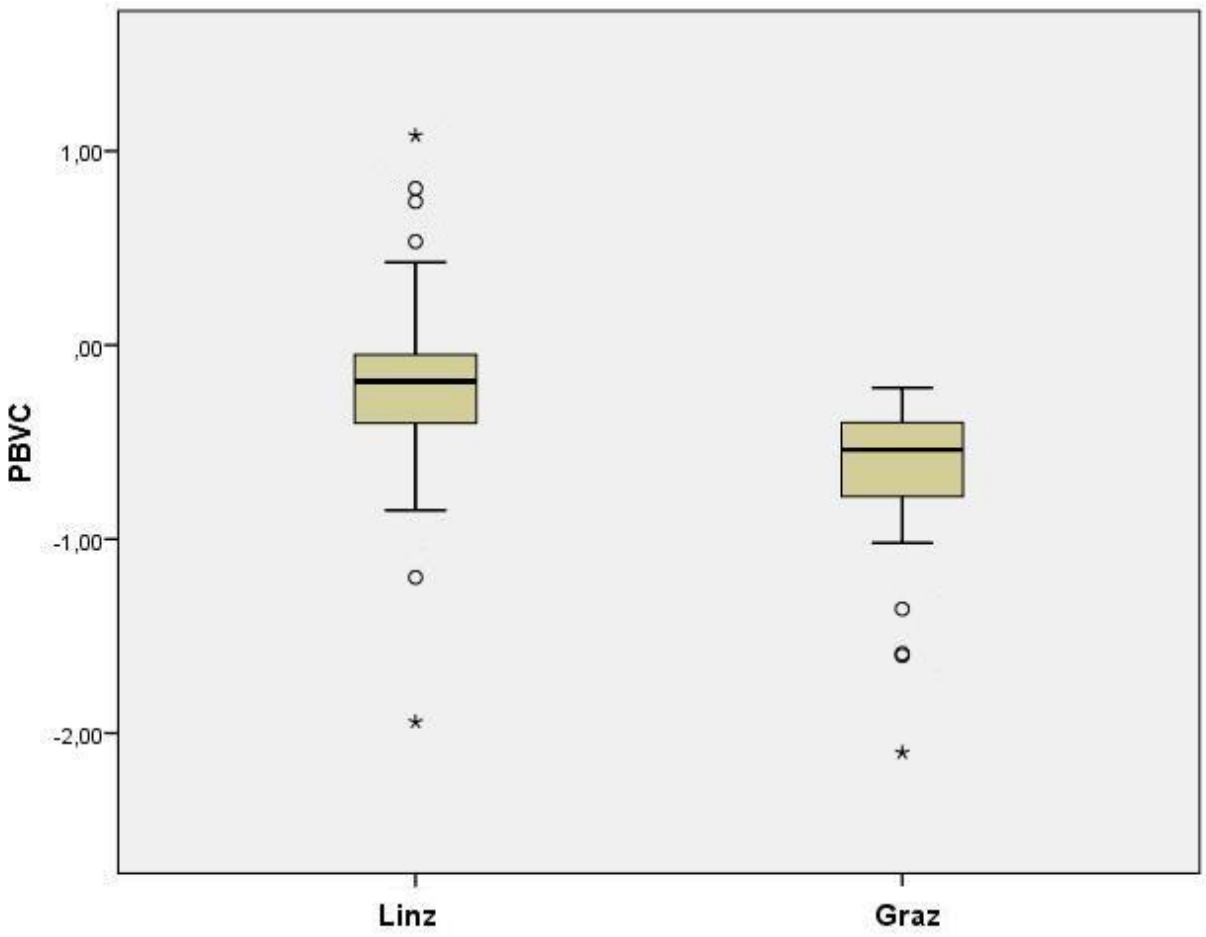


Figure 7. Percentage brain volume change between two centers. The results from SIENA demonstrated significant differences between the centers regarding the percentage brain volume change with p less than 0.0001.

4.Discussion

Treatment options in MS grow rapidly and it has a high importance to have generally accepted and comprehensive outcome measures. Many paraclinical outcome measures are already available and are used in addition to clinical data for obtaining information on treatment efficacy. Meta-analyses show a strong correlation between treatment effects on MRI-derived activity measures and relapses activity [Rommer, P. S. et al. 2014]. Many computerized methods have been proposed recent years for in vivo MRI-defined quantitative measurement of the brain metrics [Wattjes M. et al., 2015]. They are beginning to demonstrate important potential applications in neuroscience. For the treatment of MS patients a high importance has understanding the causes of long-term disability. From the earliest neuroimaging studies atrophy of brain tissue has frequently been observed in MS patients. Later clinical studies showed that brain atrophy can develop already in the early phases of MS and grey matter atrophy is a better indicator of long-term disability progression than white matter atrophy or accumulation of lesion burden [Horakova D. et al.,2009; Filippi M.et al., 2013; Jacobsen C. et al.,2014; Grothe M. et al.,2016]. Brain volume loss as an outcome measure to assess treatment efficacy is increasingly added to NEDA (referred to as NEDA-4). Last Canadian guidelines about role of MRI in optimizing the treatment of multiple sclerosis already lean on NEDA-4 [Arnold D. et al., 2015]. In this respect, measurement of brain atrophy as an objective marker of MS severity with the potential to monitor treatment efficacy in MS is becoming an important parameter in patients' follow-up.

Assessing the influence of cerebral atrophy on the MS progression is more reliable in conducting multicenter studies because of the increased statistical significance. Also large datasets analysis is necessary for understanding of MS pathophysiology and for optimizing of its treatment. Especially it is important in longitudinal MRI studies for examination the disease activity and response to immunomodulatory treatment. Such analysis often needs a pooling of MRI datasets from several MS centers. In routine clinical practice is not always possible to make a MRI study with standardize MR protocol. The lower reproducibility of not optimized MRI protocols is a major challenge in cross-sectional and longitudinal studies as also in clinical trials.

MRI scans in our study were not performed with identical sequence parameters because of the uncoordinated imaging protocols. Whereas that the protocol difference reduced the overall accuracy of the whole segmentation protocol, our aim was to estimate the influence of the differences in scanner parameters on the similarity of brain volume characteristics in MS patients on two distinct datasets from two clinical centers. Our greatest interest was the longitudinal measurement of the brain volume change, such a very important feature of MS monitoring, and especially assessment of the variability of PBVC between the various protocols, which is often found in clinical practice.

Our study consisted of two parts. First we tried to select two cohorts with different scanner parameters but similar to each other regards clinical characteristics. Subjects in our study were similar concerning disease duration and disability, MS phenotype, DMT and age (Table 5). In spite of similar T2 lesion load in the two cohorts, T1 lesion load was different (Table 6). A plausible explanation for this may be the various slice sickness between two cohorts (1 mm for Graz cohort and 6 mm for Linz cohort) with which at 6 mm slices it is not always possible to detect all the hypointense T1-weighted lesions ("black holes"). Results from SIENAX (Table 7) showed that all brain metrics varied significantly between the two cohorts except the global brain volume. We would expect that patients with similar clinical features, lesion load and global brain volume would be also having no differences in other brain volume metrics. Therefore, the large variation in brain volume characteristics between two centers indicated the need for the similar MRI protocols.

In the second part of the study, variability analysis was performed. It needs to be taken into account that protocols of our study were not standardized, our aim was to find out whether patients from different centers could be compared without standardizing the protocols.

By investigating the variability between brain volume metrics we first computed the CoV, which did not vary to a great extent between cohorts. Second the mean percentage differences were evaluated (Table 9). The findings demonstrate that despite the same global brain volume according to SIENAX, the ratio of white and grey matter between the two centers was completely different.

Then, concerning the importance in determining the MS progression, it was important for us to understand whether the progression of atrophy was different in patients from two centers. By using SIENA to assess the impact of measurement variability on longitudinal volumetric studies we determined the PBVC in two cohorts, which is necessary for detecting a significant volume change between two volume measurements in a period of time. The results from SIENA showed significant differences in PBVC with p less than 0.0001 (Table 10).

From these results, it becomes evident that pooling data from 2D and 3D sequences does not result in comparable T1 lesion load and brain volume characteristics, both cross-sectional and longitudinal. It raises the question of the causes of such variability.

Since in clinical practice both 1.5T and 3T MRI systems are used in longitudinal MRI studies, more and more interest is focusing on the questions of datasets variability. There are several sources of variability, which include subject-related physiological factors like brain volume changing depending on hydration status [Duning T. et al., 2005] or diurnal fluctuations in brain volume [Nakamura K. et al., 2015], inflammation [Cotton F. et al., 2003] or steroid therapy [Hoogervorst E. et al.,2002]; instrument-related factors such as field strength, head coil, gradients, and sequence parameters and also factors related to the scanning procedure as subject positioning. Additional variability can be related with the image post-processing stage in differences between the software tools. The definition of lesions and the interpretation of images by several experts may be other sources of variability [De Guio F. et al., 2016] (Fig.8).

The interscanner and intrascanner variability of brain MRI is expectable in multicenter and longitudinal MRI MS studies. But the reliability of the brain measures has received relatively little systematic investigation. In spite of quite a number of such multicenter studies, were conducted only few studies of the intra (differences between the measurements acquired with one scanner) and interscanner variability (differences between scanners) in MS patients (Table 11).

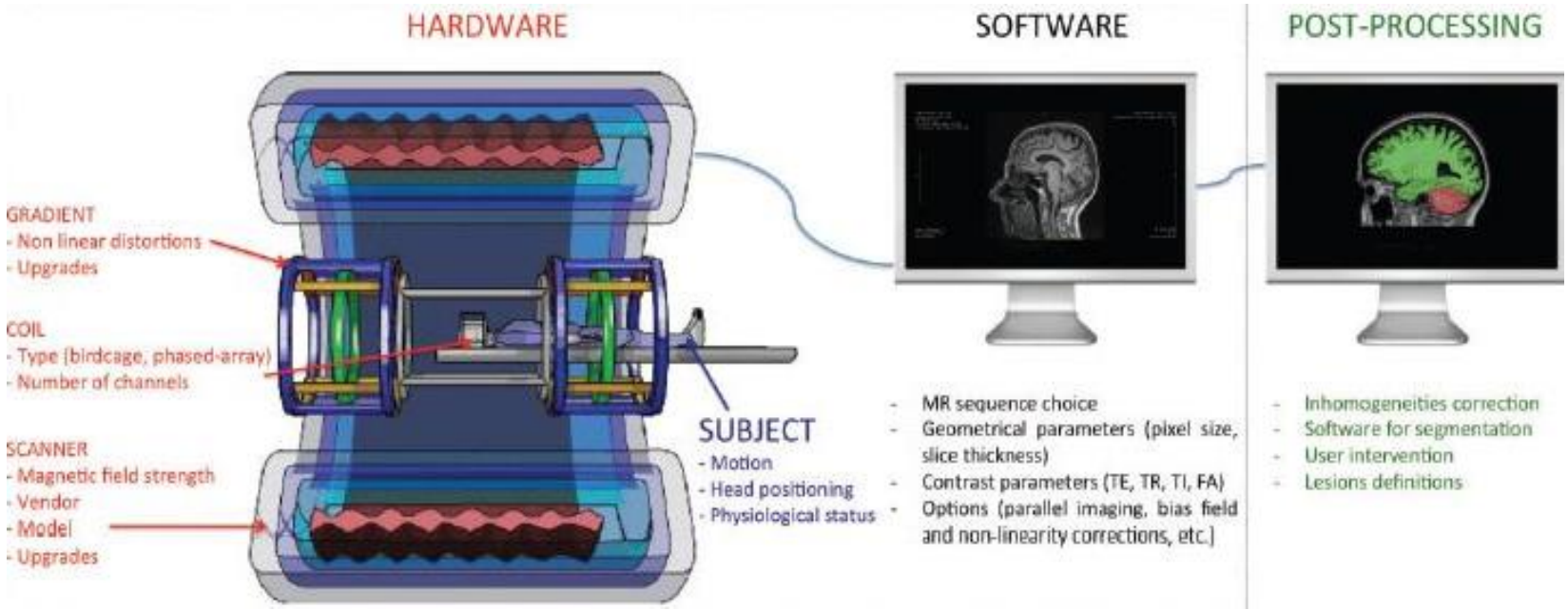


Figure 8. Schematic view of the different sources of variability in a multicenter MRI study (adapted from www.imaios.com) taken from De Guio F. et al., 2016.

Table 11. Summary of MS reproducibility studies

First author	Year published	Marker of interest	Magnetic Field strength	Analysis tool(s)	Nb subjects rescanned	Nb of MRI scanners	Results
V. Biberacher	2016	T2-hyperintense lesions, volumes of grey matter, white matter, whole brain; cortical thickness PBVC	3 T	automated lesion segmentation tool; SPM8/VBM8 SIENAX; SIENA; FreeSurfer	2 RRMS patients	3	Variability between the scanners by all software tools: significant differences in total lesion volume, global brain tissue volumes and cortical thickness measures. PBVC is more reliable to assess by SIENA than by SPM/VBM.
F. Durand-Dubief	2012	Brain volume change	1.5T	segmentation-based	9 MS patients	2	Segmentation-based techniques and SIENAX provided larger and more

				algorithms(Bayesian tissue classificatio n algorithm, FreeSurfer) and registration-based algorithms (BBSI, KN-BSI, SIENA, SIENAX, JI)			heterogeneous values of brain volume changes than registration-based techniques.
A. Lysandropo ulos	2015	GM volume and parenchymal volume	1,5 and 3 T	MSmetrix Siena	18 MS patients	2	MSmetrix showed more robust results on both the 1.5T and 3T systems.

M. Derakhshan	2010	GM atrophy	1.5 T	SPM8b, FIRST, SIENAx, Freesurfer, Medical Imaging NetCDF and Multispectral Bayesian Classifier	3 SPMS patients	3	Overall GM segmentation of the automated techniques is comparable. The automated segmentation of deep GM is much less accurate than that of cortical GM. FIRST and Freesurfer were the only automated methods that produced relatively accurate segmentations of deep GM.
C. Gasperini	2001	Whole brain volume	1.5 T	seed-growing technique based on signal intensity thresholding	9 RRMS patients	2	No significant differences of the measured brain volume between MR scanners.

However, all these variability studies used relatively similar acquisition parameters. The ratification of segmentation protocol between two-dimensional (2D) and three-dimensional (3D) data sets has become especially important since a large amount of such data has been collected in the last time in the majority of clinical MS trials.

Evaluation of the brain volume metrics depends on the tissue segmentation. 3D acquisition schemes offer the advantage of allowing improved through-plane spatial resolution. This leads to improved image registration, and also to smaller interpolation-induced resampling errors, compared to 2D images with thick slices [Vrenken H. et al., 2013]. Such different results of brain volume metrics by clinically similar MS patient cohorts in our study also can be related to poor tissue contrast. It should be noted that the image acquisition parameters were slightly different between scanners from each cohort. A T1-weighted MRI data set is usually used to calculate brain volume metrics. Repetition time (TR) is the length of the relaxation period between two excitation pulses and is therefore crucial for T1 contrast.

In our study 3D MPRAGE sequences from Graz cohort had TR=1900 ms, and TR from 2D SE sequences of two Linz cohort scanners was 491 and 472 ms. If a short repetition time (less than about 600 msec) is selected, image contrast is strongly affected by T1 (TR A in Fig. 9). Tissues with a long T1 appear dark because they do not regain much of their longitudinal magnetization during the TR interval and thus produce a weaker MR signal. If a fairly long repetition time (typically over 1500 msec) is selected, all tissues including those with a long T1 have enough time to return to equilibrium and hence they all give similar signals (TR B in Fig. 9). As a result, there is less T1 weighting because the effect of T1 on image contrast is only small. Tissues with a short T1 appear bright because they regain most of their longitudinal magnetization during the TR interval and thus produce a stronger MR signal [Weishaupt D. et al., 2006]. Thus, the poor tissue contrast in 2D compared with 3D T1-weighted images (Fig. 10) could explain the large percentage differences in volumes of brain volume metrics between two cohorts. From the Figure 10, it is clear that 3D MPRAGE gives better grey/white contrast than 2D SE.

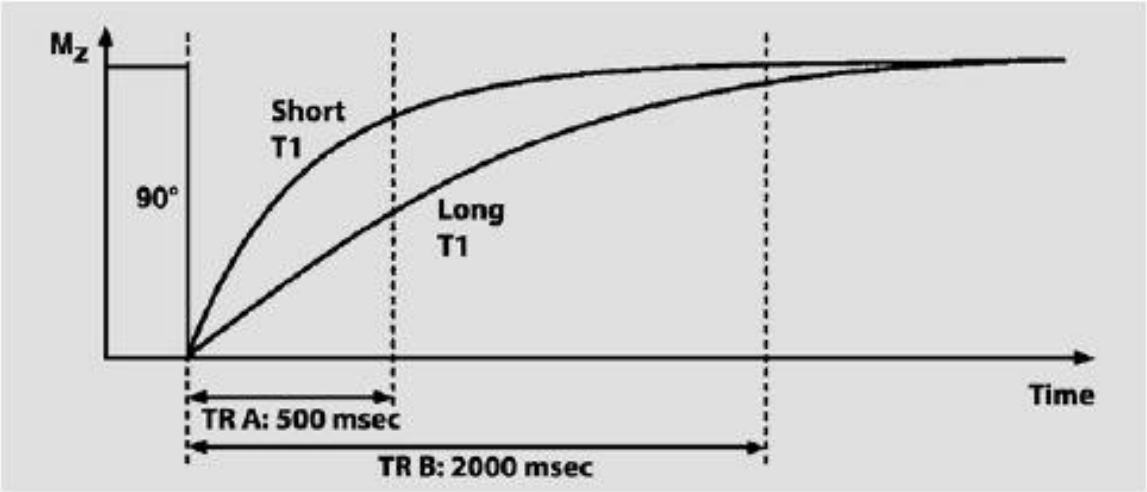


Figure 9. Relationship between TR and T1 contrast based on Weishaupt D. et al., 2006. By short repetition time (less than about 600 msec), image contrast is strongly affected by T1 (TR A). By a fairly long repetition time (typically over 1500 msec), all tissues including those with a long T1 have enough time to return to equilibrium and hence give similar signals (TR B).

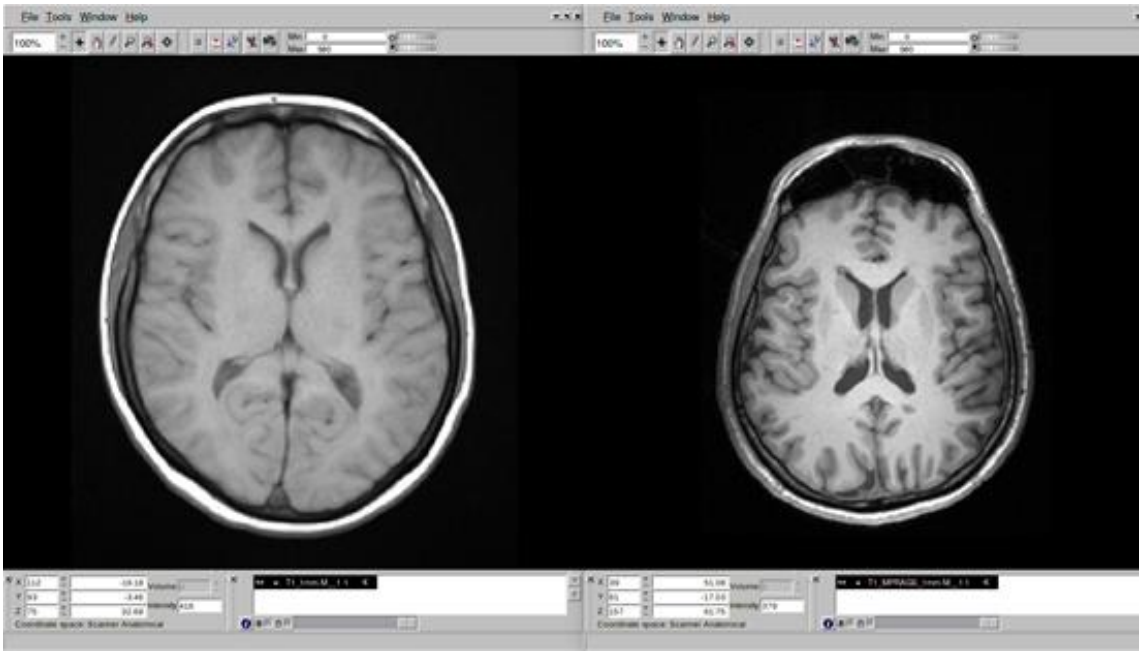


Figure 10. Example of 3D T1-weighted images from Graz (right) and Linz (left) cohorts demonstrated a poor tissue contrast in 2D T1-weighted images.

Variability between centers can be due to contrast sensitive acquisition parameters (such as the flip angle, the echo time, the repetition time, and the inversion time). Furthermore, hardware differences such as inhomogeneities of the different field strengths and different sensitivity profiles of the radiofrequency coils used may yield site-specific contrast differences.

The data in our study consists of MRI scans of 174 patients at the baseline and follow-up from two MS centers with different MRI protocols, including a different sequences and magnetic field strength. The fact that our 2D and 3D MRI scans were performed not on the same patients sample is a major limitation of our study. Nevertheless, this allowed us to use a larger sample size than in other variability studies. Another shortcoming was lack of opportunity to distinguish the different sources of our data variability. However, the main goal of the current study was to determine the possibility using different MRI protocols in clinical practice concerning the measurement of lesion load and brain volumes metrics in clinically similar MS patients.

In conclusion, our finding in atrophy metrics demonstrated that PBVC was significantly lower in Linz cohort than in Graz (-0,19% and -0,39%, respectively). But patients in our study were similar concerning age, disease duration and disability, DMT, lesion load and the global brain volume at the baseline and follow-up. On the assumption of a pathological brain volume loss about 0.4% per year [De Stefano N. et al., 2014] in MS patients, reliability of brain volume change comparison between MRI with different protocols seems not to be possible. T1 lesion load and brain volume metrics as well cross-sectional as longitudinal of 2D SE data were not comparable with the automatic segmenting of 3D MPRAGE data for MS patients as well in clinical routine as in longitudinal and multicenter studies.

REFERENCES

1. Arnold D., Li D., Hohel M., et al. “Evolving role of MRI in optimizing the treatment of multiple sclerosis: Canadian Consensus recommendations”. Mult Scler J. 2015; 1:1-9.
2. Ascherio A., Munger K. “Environmental risk factors for multiple sclerosis. Part I: the role of infection”. Ann Neurol. 2007 Apr;61(4):288-99.
3. Barkhof F, Calabresi PA, Miller DH, Reingold SC Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. 2009 Nat Rev Neurol 5:256–266
4. Barkhof F, Simon JH, Fazekas F, Rovaris M, Kappos L, de Stefano N, Polman CH, Petkau J, Radue EW, Sormani MP, Li DK, O’Connor P, Montalban X, Miller DH, Filippi M MRI monitoring of immunomodulation in relapse–onset multiple sclerosis trials. 2012 Nat Rev Neurol 8:13–21
5. Barkhof F. “MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS)”. Mult Scler. 1999; 5(4): 283–6.
6. Barkhof F. “The clinico-radiological paradox in multiple sclerosis revisited”. Curr Opin Neurol 2002; 15: 239–45.
7. Barkhof, F., Calabresi, P., Miller D., Reingold, S. “Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials”. Nat. Rev. Neurol. 2009. 5, 256–266.
8. Belbasis L., Bellou V., Evangelou E., Ioannidis J., Tzoulaki I. “Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses”. Lancet Neurol. 2015 Mar;14(3):263-73.
9. Biberacher V., Schmidt P., Keshavan A., Boucard C., Righart R., Sämann P., Preibisc C., Fröbel D., Aly L., Hemmer B., Zimmer C., Henry R., Mühlau M. “Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis”. Neuroimage. 2016 Nov 15;142:188-197.
10. Birken D., Oldendorf W. “N-acetyl-L-aspartic acid: a literature review of a compound prominent in 1H-NMR spectroscopic studies of brain”. Neurosci Biobehav Rev. 1989 Spring;13(1):23-31.

11. Bitsch A., Schuchardt J., Bunkowski S., Kuhlmann T., Bruck W. “Acute axonal injury in multiple sclerosis. Correlation with demyelination and inflammation”. Brain. 2000;123:1174–1183.
12. Bjartmar C., Kidd G., Mörk S., Rudick R., Trapp B. “Neurological disability correlates with spinal cord axonal loss and reduced N-acetyl aspartate in chronic multiple sclerosis patients”. Ann Neurol. 2000 Dec;48(6):893-901.
13. Brex P., Ciccarelli O., O’Riordan J., Sailer M., Thompson A., Miller D. “A longitudinal study of abnormalities on MRI and disability from multiple sclerosis”. The New England Journal of Medicine, vol. 346, no. 3, pp. 158–164, 2002.
14. Chard D, Trip SA. Resolving the clinico-radiological paradox in multiple sclerosis. F1000Res. 2017 Oct 12;6:1828.
15. Cohen JA, Reingold SC, Polman CH, Wolinsky JS; International Advisory Committee on Clinical Trials in Multiple Sclerosis. Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects.Lancet Neurol. 2012 May;11(5):467-76.
16. Compston A., Coles A. “Multiple sclerosis”. Lancet. 2008 Oct 25;372(9648):1502-17.
17. Confavreux C.,Vukusic S. “Natural history of multiple sclerosis: A unifying concept”. Brain 2006; 129: 606–616.
18. Cotton F., Weiner H., Jolesz F., Guttmann C. “MRI contrast uptake in new lesions in relapsing-remitting MS followed at weekly intervals”. Neurology. 2003 Feb 25;60(4):640-6.
19. Cutter G., Baier M., Rudick R., Cookfair D., Fischer J., Petkau J., Syndulko K., Weinshenker B., Antel J., Confavreux C., Ellison G., Lublin F., Miller A., Rao S., Reingold S., Thompson A., Willoughby E. “Development of a multiple sclerosis functional composite as a clinical trial outcome measure”. Brain. 1999 May;122 (Pt 5):871-82.
20. Dalton C., Brex P., Miszkiel K., Hickman S., MacManus D., Plant G., et al. “Application of the new McDonald criteria to patients with clinically isolated syndromes suggestive of multiple sclerosis”. Ann Neurol 2002;52:47–53.
21. De Guio F., Jouvent E., Biessels J., et al. “Reproducibility and variability of quantitative magnetic resonance imaging markers in cerebral small vessel disease”. Journal of Cerebral Blood Flow & Metabolism. 2016;36(8):1319-1337.

22. De Stefano N, Matthews PM, Filippi M, Agosta F, De Luca M, Bartolozzi ML, Guidi L, Ghezzi A, Montanari E, Cifelli A, Federico A, Smith SM. Evidence of early cortical atrophy in MS: relevance to white matter changes and disability. Neurology. 2003 Apr 8;60(7):1157-62.
23. De Stefano N., Airas L., Grigoriadis N., Mattle H., O'Riordan J., Oreja-Guevara C., Sellebjerg F., Stankoff B., Walczak A., Wiendl H., Kieseier B. "Clinical relevance of brain volume measures in multiple sclerosis". CNS Drugs. 2014 Feb;28(2):147-56.
24. Derakhshan M., Caramanos Z., Giacomini P., Narayanan S., Maranzano J., Francis J., Arnold D., Collins D. "Evaluation of automated techniques for the quantification of grey matter atrophy in patients with multiple sclerosis". Neuroimage. 2010 Oct 1;52(4):1261-7.
25. Duning T., Kloska S., Steinsträter O., Kugel H., Heindel W., Knecht S. "Dehydration confounds the assessment of brain atrophy". Neurology. 2005 Feb 8;64(3):548-50.
26. Durand-Dubief F., Belaroussi B., Armspach J., Dufour M., Roggerone S., Vukusic S., Hannoun S., Sappey-Marinier D., Confavreux C., Cotton F. "Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques". AJNR Am J Neuroradiol. 2012 Nov;33(10):1918-24.
27. Eksborg S. "Evaluation of method-comparison data". Clin Chem. 1981;27:1311–2.
28. Enzinger C, Barkhof F, Ciccarelli O, Filippi M, Kappos L, Rocca MA, Ropele S, Rovira À, Schneider T, de Stefano N, Vrenken H, Wheeler-Kingshott C, Wuerfel J, Fazekas F; MAGNIMS study group. Nonconventional MRI and microstructural cerebral changes in multiple sclerosis. Nat Rev Neurol. 2015 Dec;11(12):676-86.
29. Enzinger C., Fuchs S., Pichler A., Wallner-Blazek M., Khalil M., Langkammer C., Ropele S., Fazekas F. "Predicting the severity of relapsing-remitting MS: the contribution of cross-sectional and short-term follow-up MRI data". Mult Scler. 2011 Jun;17(6):695-701.
30. Ferguson B., Matyszak M., Esiri M., Perry V. "Axonal damage in acute multiple sclerosis lesions". Brain. 1997 Mar;120 (Pt 3):393-9.
31. Filippi M, Rocca MA, Barkhof F, Brück W, Chen JT, Comi G, DeLuca G, De Stefano N, Erickson BJ, Evangelou N, Fazekas F, Geurts JJ, Lucchinetti C, Miller DH, Pelletier D, Popescu BF, Lassmann H; Attendees of the Correlation between

Pathological MRI findings in MS workshop. Association between pathological and MRI findings in multiple sclerosis. Lancet Neurol. 2012 Apr;11(4):349-60.

32. Filippi M., Paty D., Kappos L. et al. “Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis: a follow-up study”. Neurology, vol. 45, no. 2, pp. 255–260, 1995.
33. Filippi M., Preziosa P., Rocca M. “Multiple sclerosis”. 2016, Handbook of Clinical Neurology, Vol. 135 (3rd series) Neuroimaging, Part I, 399-423
34. Filippi M., Rocca, M., Gasperini, C., et al. “Interscanner variation in brain MR lesion load measurements in multiple sclerosis using conventional spin-echo, rapid relaxation-enhanced, and fast-FLAIR sequences”. Am. J. Neuroradiol. 1999;20, 133–137.
35. Fisniku LK, Brex PA, Altmann DR, Miszkiel KA, Benton CE, Lanyon R, Thompson AJ, Miller DH. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain. 2008 Mar;131(Pt 3):808-17.
36. Ganter P., Prince C., Esiri M. “Spinal cord axonal loss in multiple sclerosis: a post-mortem study”. Neuropathol Appl Neurobiol. 1999 Dec;25(6):459-67.
37. Gasperini C., Rovaris M., Sormani M., Bastianello S., Pozzilli C., Comi G., Filippi M. “Intra-observer, inter-observer and inter-scanner variations in brain MRI volume measurements in multiple sclerosis”. Mult Scler. 2001 Feb;7(1):27-31.
38. Geurts J., Barkhof F. “Grey matter pathology in multiple sclerosis”. Lancet Neurol 2008; 7: 841–51.
39. Geurts J., Wolswijk G., Bo L., van d. V Polman C., Troost D., Aronica E. “Altered expression patterns of group I and II metabotropic glutamate receptors in multiple sclerosis”. Brain. 2003;126:1755–1766.
40. Gilmore C., Donaldson I., Bö L., Owens T., Lowe J., Evangelou N. “Regional variations in the extent and pattern of grey matter demyelination in multiple sclerosis: a comparison between the cerebral cortex, cerebellar cortex, deep grey matter nuclei and the spinal cord”. J Neurol Neurosurg Psychiatry. 2009. Feb;80(2):182-7.
41. Giorgio A, Battaglini M, Smith SM, De Stefano N. Brain atrophy assessment in multiple sclerosis: importance and limitations. Neuroimaging Clin N Am. 2008 Nov;18(4):675-86, xi.
42. Giorgio A., Stromillo M., Bartolozzi M., Rossi F., Battaglini M., De Leucio A., Guidi L., Maritato P., Portaccio E., Sormani M., Amato M., De Stefano N.

"Relevance of hypointense brain MRI lesions for long-term worsening of clinical disability in relapsing multiple sclerosis". Mult Scler. 2014 Feb;20(2):214-9.

43. Goodin D. "Magnetic resonance imaging as a surrogate outcome measure of disability in multiple sclerosis: have we been overly harsh in our assessment?" Ann Neurol. 2006.59: 597–605.
44. Grimaud J., Barker G., Wang L., Lai M., MacManus D., Webb S., Thompson A., McDonald W., Tofts P., Miller D. "Correlation of magnetic resonance imaging parameters with clinical disability in multiple sclerosis: a preliminary study". J Neurol. 1999 Oct;246(10):961-7.
45. Grothe M., Lotze M., Langner S., Dressel A. "The role of global and regional gray matter volume decrease in multiple sclerosis". J Neurol. 2016 Jun;263(6):1137-45.
46. Hauser S., Oksenberg J. "The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration". Neuron. 2006;52:61–76
47. Havrdova E., Galetta S., Stefoski D., Comi G. "Freedom from disease activity in multiple sclerosis". Neurology 2010; 74 Suppl 3: S3-7.
48. Hollenbach J., Oksenberg J. "The Immunogenetics of Multiple Sclerosis: A Comprehensive Review. Journal of autoimmunity". 2015;64:13-25.
49. Hoogervorst E., Polman C., Barkhof F. "Cerebral volume changes in multiple sclerosis patients treated with high-dose intravenous methylprednisolone". Mult Scler 2002;8:415–19.
50. Horakova D., Dwyer M., Havrdova E., Cox J., Dolezal O., Bergsland N., Rimes B., Seidl Z., Vaneckova M., Zivadinov R. "Gray matter atrophy and disability progression in patients with early relapsing-remitting multiple sclerosis: a 5-year longitudinal study". J Neurol Sci. 2009 Jul 15;282(1-2):112-9.
51. Jacobsen C., Hagemeier J., Myhr K., Nyland H., Lode K., Bergsland N., Ramasamy D., Dalaker T., Larsen J., Farbu E., Zivadinov R. "Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study". J Neurol Neurosurg Psychiatry. 2014 Oct;85(10):1109-15.
52. Kappos L., De Stefano N., Freedman M., Cree B., Radue E., Sprenger T., Sormani M., Smith T., Häring D., Meier P., Tomic D. "Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis". Mult Scler. 2016 Sep;22(10):1297-305.
53. Kappos L., Moeri D., Radue E. et al., "Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or

impairment in multiple sclerosis: a meta-analysis". The Lancet, vol. 353, no. 9157, pp.964–969, 1999.

54. Kister I., Chamot E., Salter A., Cutter G., Bacon T., Herbert J. "Disability in multiple sclerosis: a reference for patients and clinicians". Neurology. 2013 Mar 12;80(11):1018-24.
55. Kornek B., Storch M., Weissert R., Wallstroem E., Stefferl A., Olsson T., Linington C., Schmidbauer M., Lassmann H. "Multiple sclerosis and chronic autoimmune encephalomyelitis: a comparative quantitative study of axonal injury in active, inactive, and remyelinated lesions". Am J Pathol. 2000;157:267–276.
56. Kurtzke J. "Geography in multiple sclerosis". J Neurol 1977;215(1):1–26.
57. Kurtzke J. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983 Nov;33(11):1444-52.
58. Lavorgna L., Bonavita S., Ippolito D., Lanzillo R., Salemi G., Patti F., Valentino P., Coniglio G., Buccafusca M., Paolicelli D., d'Ambrosio A., Bresciamorra V., Savettieri G., Zappia M., Alfano B., Gallo A., Simone I., Tedeschi G. "Clinical and magnetic resonance imaging predictors of disease progression in multiple sclerosis: a nine-year follow-up study". Mult Scler. 2014 Feb;20(2):220-6.
59. Li DK, Held U, Petkau J, Daumer M, Barkhof F, Fazekas F, Frank JA, Kappos L, Miller DH, Simon JH, Wolinsky JS, Filippi M; Sylvia Lawry Centre for MS Research. MRI T2 lesion burden in multiple sclerosis: a plateauing relationship with clinical disability. Neurology. 2006 May 9;66(9):1384-9.
60. Lublin F., Reingold S. "Defining the clinical course of multiple sclerosis: results of an international survey". Neurology 1996;46:907–911.
61. Lublin F., Reingold S., Cohen J., et al. "Defining the clinical course of multiple sclerosis: the 2013 revisions". Neurology 2014;83(3):278–86.
62. Lysandropoulos A., Absil J., Metens T., Mavroudakis N., Guisset F., Van Vlierberghe E., Smeets D., David P., Maertens A., Van Hecke W. "Quantifying brain volumes for Multiple Sclerosis patients follow-up in clinical practice - comparison of 1.5 and 3 Tesla magnetic resonance imaging". Brain Behav. 2016 Jan 12;6(2):e00422.
63. Ma V., Chan L., Carruthers K. "The Incidence, Prevalence, Costs and Impact on Disability of Common Conditions Requiring Rehabilitation in the US: Stroke, Spinal Cord Injury, Traumatic Brain Injury, Multiple Sclerosis, Osteoarthritis,

Rheumatoid Arthritis, Limb Loss, and Back Pain." Archives of physical medicine and rehabilitation. 2014;95(5):986-995.

64. Matthews P., De Stefano N., Narayanan S., Francis S., Wolinsky J., Antel J., Arnold D. "Putting magnetic resonance spectroscopy studies in context: axonal damage and disability in multiple sclerosis". Semin Neurol. 1998;18(3):327-36.
65. McDonald W., Miller D., Thompson A. "Are magnetic resonance findings predictive of clinical outcome in therapeutic trials in multiple sclerosis? The dilemma of interferon-beta". Ann Neurol. 1994;36(1):14–8.
66. Miller DH, Simon JH, Wolinsky JS, Filippi M; Sylvia Lawry Centre for MS Research. MRI T2 lesion burden in multiple sclerosis: a plateauing relationship with clinical disability. Neurology. 2006 May 9;66(9):1384-9.
67. Milo R., Miller A. "Revised diagnostic criteria of multiple sclerosis". Autoimmun Rev. 2014 Apr-May;13(4-5):518-24.
68. Minneboo A, Jasperse B, Barkhof F, Uitdehaag BM, Knol DL, de Groot V, Polman CH, Castelijns JA. Predicting short-term disability progression in early multiple sclerosis: added value of MRI parameters. J Neurol Neurosurg Psychiatry. 2008 Aug;79(8):917-23.
69. Mitjana R, Tintoré M, Rocca MA, Auger C, Barkhof F, Filippi M, Polman C, Fazekas F, Huerga E, Montalban X, Rovira A. Diagnostic value of brain chronic black holes on T1-weighted MR images in clinically isolated syndromes. MultScler. 2014 Oct;20(11):1471-7.
70. Montalban X., Tintore M., Swanton J., et al. "MRI criteria for MS in patients with clinically isolated syndromes". Neurology 2010;74(5):427-434.
71. Nagtegaal GJ, Pohl C, Wattjes MP, Hulst HE, Freedman MS, Hartung HP, Miller D, Montalban X, Kappos L, Edan G, Pleimes D, Beckman K, Stemper B, Polman CH, Sandbrink R, Barkhof F. Interferon beta-1b reduces black holes in a randomized trial of clinically isolated syndrome. Mult Scler. 2014 Feb;20(2):234-42.
72. Nakamura K., Brown R., Narayanan S., Collins D., Arnold D. "Diurnal fluctuations in brain volume: Statistical analyses of MRI from large populations". Alzheimer's Disease Neuroimaging Initiative. Neuroimage. 2015 Sep;118:126-32.
73. Paty D., Li D., Oger J., et al. "Magnetic resonance imaging in the evaluation of clinical trials in multiple sclerosis". Ann Neurol 1994; 36 (Suppl): S95–S96.

74. Plummer DL. DispImage: a display and analysis tool for medical images. Revista di Neuroradiologica 1992;5:489–95.
75. Polman C., Reingold S., Banwell B., Clanet M., Cohen J., Filippi M., Fujihara K., Havrdova E., Hutchinson M., Kappos L., Lublin F., Montalban X., O'Connor P., Sandberg-Wollheim M., Thompson A., Waubant E., Weinshenker B., Wolinsky J. "Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria". Ann Neurol. 2011 Feb;69(2):292-302.
76. Polman C., Reingold S., Edan C., Filippy M., Hartung H., Kappos L., et al. "Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol 2005;56:840–6.
77. Poser C., Paty D., Scheinberg L., McDonald W., Davis F., Ebers G., et al. „New diagnostic criteria for multiple sclerosis: guidelines for research protocols". Ann Neurol 1983;13:227–31.
relationship between T2 lesion load and disability in MS. Neurology. 2009 Nov
78. Rommer PS, Zettl UK, Kieseier B, Hartung HP, Menge T, Frohman E, Greenberg BM, Hemmer B, Stüve O. Requirement for safety monitoring for approved multiple sclerosis therapies: an overview. Clin Exp Immunol. 2014 Mar;175(3):397-407.
79. Rotstein D., Healy B., Malik M., Chitnis T., Weiner H. "Evaluation of no evidence of disease activity in a 7-year longitudinal multiple sclerosis cohort". JAMA Neurol. 2015 Feb;72(2):152-8.
80. Rovira À., Wattjes M., Tintoré M., Tur C., Yousry T. Sormani M., De Stefano N., Filippi M., Auger C., Rocca M., Barkhof F., Fazekas F., Kappos L., Polman C., Miller D., Montalban X.; MAGNIMS study group. "Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-clinical implementation in the diagnostic process". Nat Rev Neurol. 2015 Aug;11(8):471-82.
81. Roxburgh R., Seaman S., Masterman T., Hensiek A., Sawcer S., et al."Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity". Neurology. 2005; 64: 1144–1151.
82. Sampat M., Healy, B., Meier, D., et al. "Disease modeling in multiple sclerosis: assessment and quantification of sources of variability in brain parenchymal fraction measurements". NeuroImage. 2010; 52, 1367–1373.

83. Scalfari A., Neuhaus A., Degenhardt A., et al. „The natural history of MS: a geographically based study 10: relapse and long-term disability". Brain 2010; 133: 1914–29.
84. Schultz V., van der Meer F., Wrzos C., Scheidt U., Bahn E., Stadelmann C., Brück W., Junker A. "Acutely damaged axons are remyelinated in multiple sclerosis and experimental models of demyelination". Glia. 2017 Aug;65(8):1350-1360.
85. Schumacher G., Beebe G., Kibler R., Kurland L., Kurtzke J., McDowell F., Nagler B., Sibley W., Tourtellotte W., Willmon T. „Problems of experimental trials of therapy in multiple sclerosis: report by the panel on the evaluation of experimental trials of therapy in multiple sclerosis". Ann N Y Acad Sci 1965;122:552–68.
86. Smith S. "Fast robust automated brain extraction". Human Brain Mapping, 17(3):143-155, November 2002.
87. Smith S., Jenkinson M., Woolrich M., Beckmann C., Behrens T., Johansen-Berg H., Bannister P., De Luca M., Drobnjak I., Flitney D., Niazy R., Saunders J., Vickers J., Zhang Y., De Stefano N., Brady J., Matthews P. "Advances in functional and structural MR image analysis and implementation as FSL". NeuroImage, 23(S1):208-219, 2004.
88. Smith S., Zhang Y., Jenkinson M., Chen J., Matthews P., Federico A., De Stefano N. "Accurate, robust and automated longitudinal and cross-sectional brain change analysis". NeuroImage, 17(1):479-489, 2002.
89. Sormani MP, Rovaris M, Comi G, Filippi M. A reassessment of the plateauing relationship between T2 lesion load and disability in MS. Neurology. 2009 Nov 10;73(19):1538-42.
90. Thaler C, Faizy T, Sedlacik J, Holst B, Stellmann JP, Young KL, Heesen C, Fiehler J, Siemonsen S. T1- Thresholds in Black Holes Increase Clinical-Radiological Correlation in Multiple Sclerosis Patients. PLoS One. 2015 Dec 11;10(12):e0144693.
91. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M, Freedman MS, Fujihara K, Galetta SL, Hartung HP, Kappos L, Lublin FD, Marrie RA, Miller AE, Miller DH, Montalban X, Mowry EM, Sorensen PS, Tintoré M, Traboulsee AL, Trojano M, Uitdehaag BMJ, Vukusic S, Waubant E, Weinshenker BG, Reingold SC, Cohen JA. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018 Feb;17(2):162-173.

92. Trapp B., Nave K.. “Multiple sclerosis: an immune or neurodegenerative disorder?” Annu Rev Neurosci. 2008;31:247-69.
93. Trapp B., Peterson J., Ransohoff R., Rudick R., Mork S., Bo L. “Axonal transection in the lesions of multiple sclerosis”. N. Engl. J. Med. 1998;338:278–285.
94. van Munster C., Bernard M., Uitdehaag J. “Outcome measures in clinical trials for multiple sclerosis”. CNS Drugs 2017;31(3):217-236.
95. Vidal-Jordana A., Montalban X. “Multiple Sclerosis: Epidemiologic, Clinical, and Therapeutic Aspects”. Neuroimaging Clin N Am. 2017 May;27(2):195-204.
96. Vidal-Jordana A., Pareto D., Sastre-Garriga J., Auger C., Ciampi E., Montalban X., Rovira A. “Measurement of Cortical Thickness and Volume of Subcortical Structures in Multiple Sclerosis: Agreement between 2D Spin-Echo and 3D MPRAGE T1-Weighted Images”. AJNR Am J Neuroradiol. 2017 Feb;38(2):250-256.
97. Vrenken H., Jenkinson M., Horsfield M., et al. “Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis”. Journal of Neurology. 2013;260(10):2458-2471.
98. Wallin M., Culpepper W., Coffman P., Pulaski S., Maloni H., Mahan C., Haselkorn J., Kurtzke J. “The Gulf War era multiple sclerosis cohort: age and incidence rates by race, sex and service”. Brain. 2012 Jun;135(Pt 6):1778-85.
99. Wattjes M., Rovira À., Miller D., Yousry T., Sormani M., de Stefano M., Tintoré M., Auger C., Tur C., Filippi M., Rocca M., Fazekas F., Kappos L., Polman C., Barkhof F., Montalban X.; MAGNIMS study group. “Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis--establishing disease prognosis and monitoring patients”. Nat Rev Neurol. 2015 Oct;11(10):597-606.
100. Weinshenker B. “Natural history of multiple sclerosis”. Ann. Neurol. 1998;36:S6–S11.
101. Weishaupt D., Köchli V., Marincek B. “How Does MRI Work? An Introduction to the Physics and Function of Magnetic Resonance Imaging”, 2nd ed New York, NY: Springer-Verlag, 2006.
102. Zivadinov R, Leist TP. Clinical-magnetic resonance imaging correlations in multiple sclerosis. J Neuroimaging. 2005;15(4 Suppl):10S-21S. Review.

Expanded Disability Status Scale (EDSS) by Kurtzke J. et al., 1983

Score	Description
1.0	No disability, minimal signs in one FS
1.5	No disability, minimal signs in more than one FS
2.0	Minimal disability in one FS
2.5	Mild disability in one FS or minimal disability in two FS
3.0	Moderate disability in one FS, or mild disability in three or four FS. No impairment to walking
3.5	Moderate disability in one FS and more than minimal disability in several others. No impairment to walking
4.0	Significant disability but self-sufficient and up and about some 12 hours a day. Able to walk without aid or rest for 500m
4.5	Significant disability but up and about much of the day, able to work a full day, may otherwise have some limitation of full activity or require minimal assistance. Able to walk without aid or rest for 300m
5.0	Disability severe enough to impair full daily activities and ability to work a full day without special provisions. Able to walk without aid or rest for 200m
5.5	Disability severe enough to preclude full daily activities. Able to walk without aid or rest for 100m
6.0	Requires a walking aid - cane, crutch, etc - to walk about 100m with or without resting
6.5	Requires two walking aids - pair of canes, crutches, etc - to walk about 20m without resting
7.0	Unable to walk beyond approximately 5m even with aid. Essentially restricted to wheelchair; though wheels self in standard wheelchair and transfers alone. Up and about in wheelchair some 12 hours a day
7.5	Unable to take more than a few steps. Restricted to wheelchair and may need aid in transfering. Can wheel self but cannot carry on in standard wheelchair for a full day and may require a motorised wheelchair
8.0	Essentially restricted to bed or chair or pushed in wheelchair. May be out of bed

Score	Description
	itself much of the day. Retains many self-care functions. Generally has effective use of arms
8.5	Essentially restricted to bed much of day. Has some effective use of arms retains some self-care functions
9.0	Confined to bed. Can still communicate and eat
9.5	Confined to bed and totally dependent. Unable to communicate effectively or eat/swallow
10.0	Death due to MS