Connecting white and grey matter damage in multiple sclerosis using multimodal MRI

Merlin M. Weeda

Connecting white and grey matter damage in multiple sclerosis using multimodal MRI

Merline Maria Weeda

The studies described in this thesis were carried out at the department of Radiology and Nuclear Medicine, the department of Neurology, and the department of Anatomy and Neurosciences of Amsterdam UMC, location VUmc, Amsterdam, the Netherlands. At the Amsterdam UMC, care and research for patients with multiple sclerosis are organized within the MS Center Amsterdam. The participating departments and the MS Center Amsterdam are embedded in Amsterdam Neuroscience.

The work presented in this thesis was funded by the Dutch MS Research Foundation (grant number 14-876).

Printing of this thesis was kindly supported by the Dutch MS Research Foundation.

Design: Merlin M. Weeda

Layout: Hans Schaapherder, www.persoonlijkproefschrift.nl

Printing: Ridderprint, www.ridderprint.nl

ISBN: 978-94-6483-430-7

VRIJE UNIVERSITEIT

Connecting white and grey matter damage in multiple sclerosis using multimodal MRI

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op dinsdag 31 oktober om 9.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Merline Maria Weeda

geboren te Rotterdam

promotor:	prof.dr. F. Barkhof
copromotoren:	dr. P.J.W. Pouwels dr.ir. H. Vrenken
promotiecommissie:	prof.dr. B.M.J. Uitdehaag prof.dr. L. Reneman prof.dr. M. Wattjes dr. M.M. Schoonheim dr. S.D. Roosendaal dr. B.A. de Jong

"Dat is de fout die velen van ons maken. Ze leggen zichzelf grenzen op. Ze bepalen vooraf wat mogelijk en onmogelijk is. En bij het onmogelijke leggen ze zich neer. Maar soms is wat onmogelijk lijkt, alleen maar moeilijk. En dan is het slechts een kwestie van doorzetten."

Stefan Brijs – De Engelenmaker

Table of contents

CHAPTER 1

Introduction

CHAPTER 1.1

General introduction

Merlin M. Weeda

Background

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system (CNS) in which healthy myelin sheaths are targeted, leading to the formation of multiple lesions (Latin: *sclerae*), demyelination, and neurodegeneration[1]. Several clinical phenotypes of MS have been described, the most common type being relapsing-remitting MS (RRMS): a disease course characterized by relapses (attacks) followed by periods of remission, and disability can increase both in the relapse and remission phases. MS can present itself with a large variety of symptoms (**Figure 1**), including visual, motor, sensory, cognitive and autonomic problems[2].

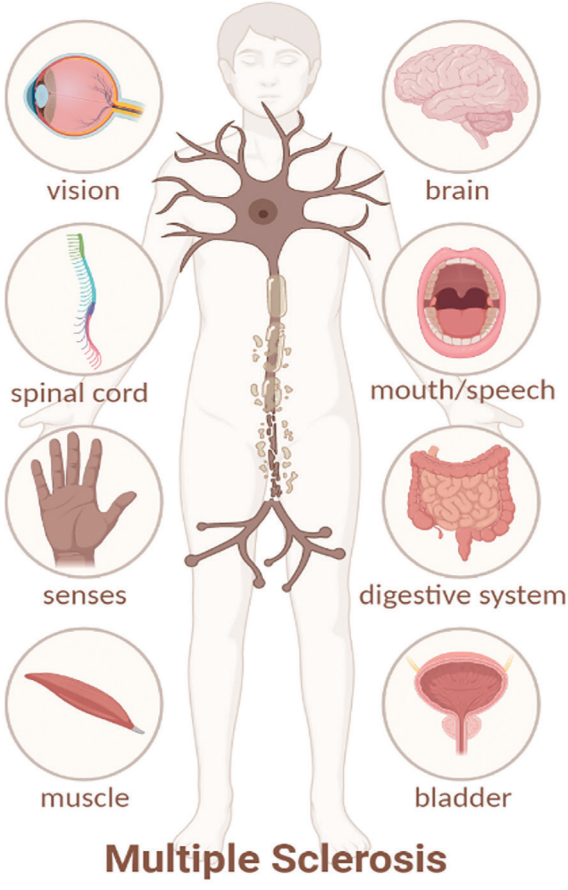

Figure 1 – Illustration of multiple sclerosis neuronal damage to the myelin sheath and the different areas that may be affected by the disturbance of the neuronal signal due to this damage. Image created in BioRender.

Although several disease modifying therapies (DMTs) exist, there is no cure for MS as of yet. One of the main problems is that the exact pathophysiology of the disease still remains unknown. Research has shown the importance of genetics as well as environmental factors in the etiology of MS[3]. Moreover, there is no good correspondence between patients'

symptoms and visualization of the disease with magnetic resonance imaging (MRI), called the clinical-radiological paradox[4]. In particular neurodegeneration, both of the brain and the spinal cord, has been linked to increasing disability[5,6], whereas diagnosis still relies mainly on neuroinflammatory characteristics[7].

White and grey matter in MS

Neurodegeneration of the brain and spinal cord can be seen in MS both in the white matter (WM) and grey matter (GM), of which the latter most strongly associates with clinical dysfunction[8]. Large variability has been reported between patients concerning atrophy of the GM, both cortical and subcortical[9]. Although variability of measured atrophy may in part be attributed to measurement uncertainty and physiological variation[10], it seems likely, given the relation to clinical disability progression and other clinical and cognitive measures, that a relevant proportion of the observed variance reflects real inter-patient variability of GM atrophy rates[11]. It has been hypothesized that this variability of GM atrophy rates arises to a large part because of the different distribution and severity of WM damage between patients[12]. It has previously been shown that WM lesion volumes are related GM atrophy, both globally and locally[13,14]. Furthermore, the average anatomical distribution of GM atrophy displays similarities to the distribution of both existing[15,16] and new[17,18] WM lesions. By studying anatomically connected regions, relations were found between GM atrophy and lesions in the connected WM tracts in several different types of MS[19-21], as well as relations between GM atrophy and integrity of the connected WM tracts[22]. However, due to the cross-sectional design of these studies, it is not possible to draw conclusions about directionality of these associations, i.e. whether WM pathology precedes GM pathology (primary WM damage) or whether GM pathology precedes WM pathology (primary GM damage). This causality is of importance to understand how neurodegeneration is related to neuroinflammation and other pathological changes in MS in order to successfully monitor interventions for neurodegeneration in MS.

MRI techniques

MRI of the brain and spinal cord is needed in order to quantify both GM and WM damage in MS. Over the years, various (multimodal) MRI techniques have been developed, each with its own specific advantages of visualizing MS pathology. The MRI techniques used in this thesis include structural imaging of the brain and spinal cord, diffusion weighted imaging (DWI), myelin water imaging (MWI) and susceptibility weighted imaging (SWI), and these techniques are summarized below.

Structural imaging of the brain

The clinical protocol for brain MRI in MS generally contains a T1 and a T2 fluid attenuated inversion recovery (FLAIR) sequence. From T1 images, cortical thickness, subcortical volume, and general WM and GM volumes can be determined. On T2-FLAIR images, MS lesions have enhanced signal, where the signal of the cerebrospinal fluid (CSF) is attenuated. This

allows a more clear visualization of periventricular MS brain lesions. Using these T2-FLAIR images, lesions can be segmented, either manually or automatically. The (semi-)automated software packages for these segmentations require varying amounts of manual input from the user[23]. Unfortunately, depending on the input images, not all methods will have the same volumetric and spatial accuracy as manual segmentation (**Figure 2A**). Additionally, lesions as seen in MS will complicate segmentation of the brain tissue into GM, WM and CSF, since its characteristics may confuse software into recognizing a lesion in the WM as GM tissue or as CSF[24]. To accommodate tissue segmentation in the presence of these lesions, the lesion segmentation masks from T2-FLAIR images are commonly registered to T1 images to enable artificial filling of the lesion tissue with the signal intensity of normal-appearing tissue[25]. When lesions are not correctly segmented and therefore not properly filled with normal appearing WM, this will not only influence the estimated lesion volume, but also the GM and WM volumes (**Figure 2B** and **2C**).

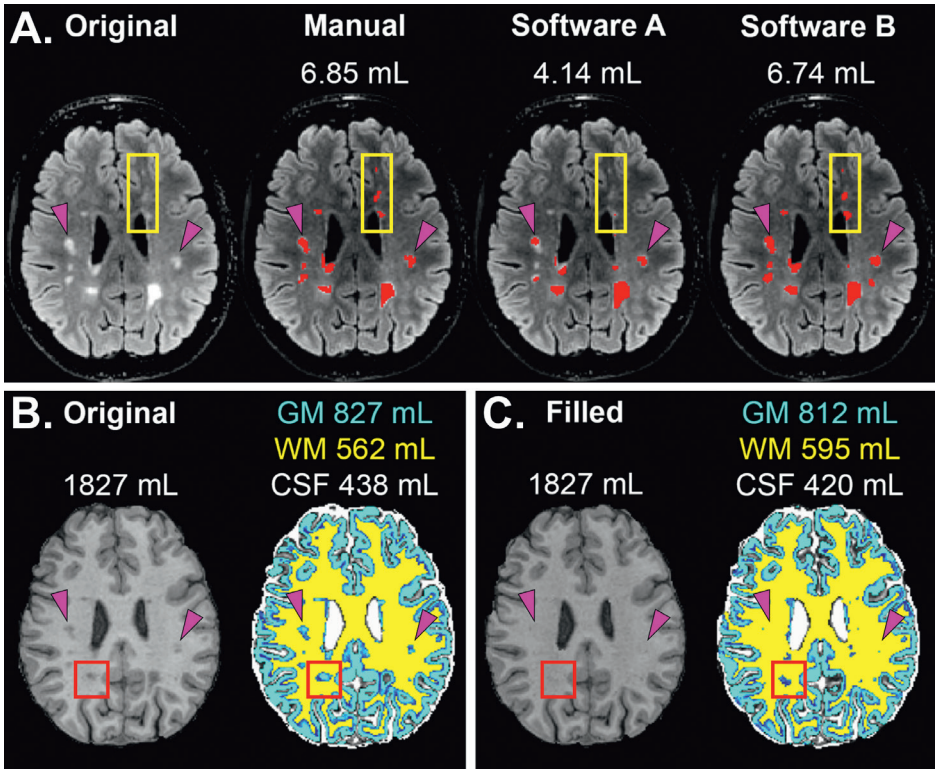

Figure 2 – Example of challenges that occur in the presence of MS lesions in brain MRI. Pink arrows correspond to the same locations in the T2-FLAIR (top row) and T1 (bottom row) images. The original T2-FLAIR image and three different lesion segmentation methods (manual and two different software packages) with their measured lesion volumes are shown in A. The yellow box highlights differences between the methods. Segmentation of the original T1 image with lesions present (B) yields different GM and WM volumes compared to the T1 image in which lesion are filled with the signal intensity of normal-appearing WM (C). The red box highlights that even with lesion filling, lesions may still be recognized as GM when they are not optimally segmented.

Structural imaging of the spinal cord

Both whole brain 3D T1 and cervical 2D T2 images can be used for analysis of the spinal cord (**Figure 3**). From these, measures such as the upper cervical cord area (UCCA) can be determined, which is known to be an important predictor of disease progression and prognosis[26]. As with the brain, several segmentation methods exist, but these are often optimized for specific MRI sequences and may not be robust or reliable in clinical settings[27]. In addition, the influence of lesions in the spinal cord for (semi-)automatic spinal cord segmentation methods is not known. Similar as for brain segmentation, this is of importance in MS MRI research.

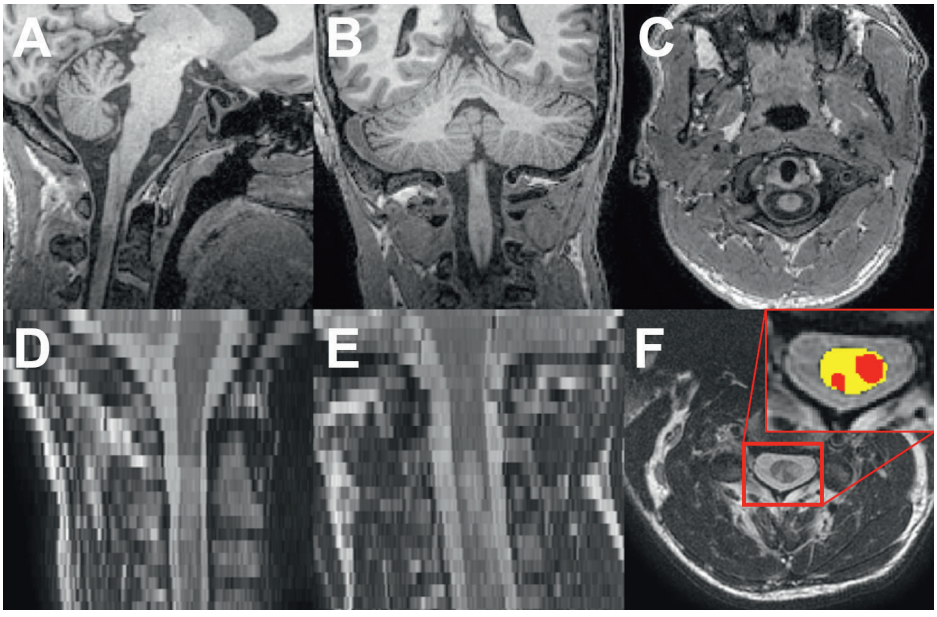

Figure 3 – Example of upper cervical cord imaging in a 3D T1 image and 2D T2 image. Top row: 3D T1 image with 1.0 mm isotropic resolution in sagittal (A), coronal (B) and axial (C) view. Bottom row: 2D T2 image with 0.4x0.4x4.0 mm resolution in sagittal (D), coronal (E) and axial (F) view. Insert shows a magnified example of spinal cord (yellow) and lesion (red) segmentation in F.

Diffusion weighted imaging

Diffusion imaging makes use of the fact that water molecules will not diffuse freely when they are in the vicinity of an anatomical WM tract[28]. From diffusion weighted images (**Figure 4**), the diffusion tensor (DTI) can be calculated, resulting in the quantitative measures fractional anisotropy (FA), and mean, axial and radial diffusivity (MD, AD, RD). When the WM integrity is diminished, for example when damage to the myelin sheath has occurred in MS, this will result in lower FA and higher MD values. In addition to yielding the diffusion tensor, DWI can also be used for tractography[29]: a way of fiber tracking to reconstruct WM structural connectivity in the CNS. Tractography benefits from multi-shell DWI techniques, and in addition these diffusion data can be analyzed with multi-compartment models such as neurite orientation dispersion and density imaging (NODDI)[30], providing additional information on the tissue microstructure such as neurite density index (NDI) and orientation dispersion index (ODI).

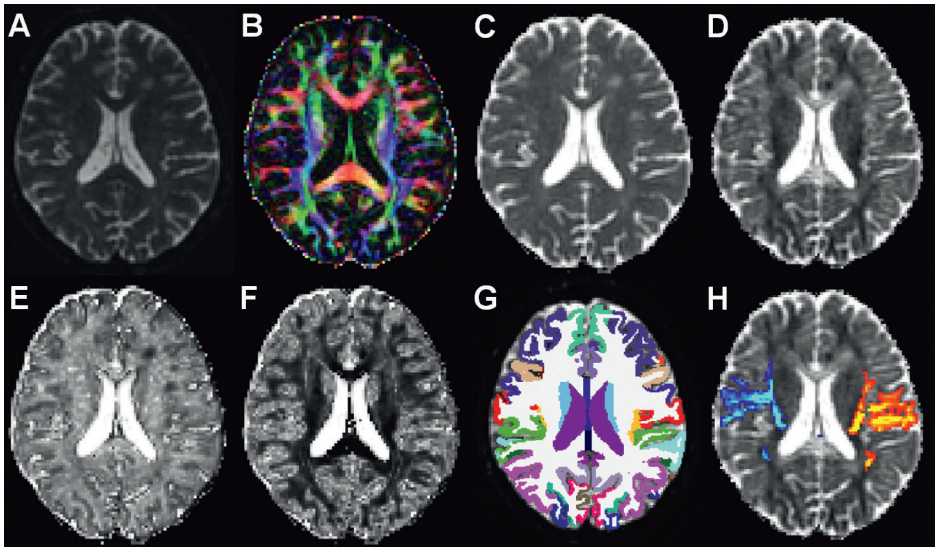

Figure 4 – Diffusion weighted imaging; with (a) non-diffusion weighted image; (b) color-coded FA map; (c) MD map; (d) registered tensor image; (e) NDI map; (f) ODI map; (g) segmentation of (sub)cortical structures; (h) tractography of the WM tracts connected to the left (red-yellow) and right (blue-light blue) postcentral cortex.

Myelin water imaging

Myelin imaging acquired by multicomponent driven equilibrium single pulse observation of T1 and T2 (mcDESPOT)[31] is a method to estimate the fractions of water that are present between the myelin sheaths, in intracellular water, and in free water. The first one, myelin water fraction (MWF), may be considered a measure of integrity of the WM, since diminished MWF indicates damage to the myelin sheath.

Susceptibility weighted imaging

Susceptibility imaging is a technique that is sensitive to disturbances of the magnetic field. This magnetic susceptibility can be quantified with respect to the overall susceptibility of the tissue with quantitative susceptibility mapping (QSM)[32]. QSM values will be higher in the vicinity of (paramagnetic) iron, which is present in MS lesions due to macrophage involvement, but also upon the loss of (diamagnetic) myelin. It is important to note that the effect of iron on QSM values is much stronger than the effect of myelin. Therefore, QSM results need to be interpreted with caution, and the region or location from which the values have been obtained (i.e. lesion, peri-lesional tissue, normal-appearing tissue) needs to be taken into account[33].

Aim of this thesis

The main aim of this thesis was to study the temporal and spatial relationships between WM damage and GM atrophy in early stages of MS using multimodal MRI techniques, and to study how these relationships may affect disability. In particular, we aimed to answer the chicken-

or-egg question: does WM damage precede GM damage (**Figure 5**), or does GM damage precede WM damage in MS?
For this, we recruited 40 subjects with early RRMS (i.e. short disease duration, low disability, and without second-line treatment) and 15 healthy control subjects and followed them for two years with three exams with clinical, imaging and neurocognitive measurements. We also used data from existing (inter)national MS cohorts. In addition, we investigated some methodological issues that arise when analyzing MRI in the MS brain and spinal cord, such as accurate tissue segmentation in the presence of MS lesions, and applied this knowledge to our research.

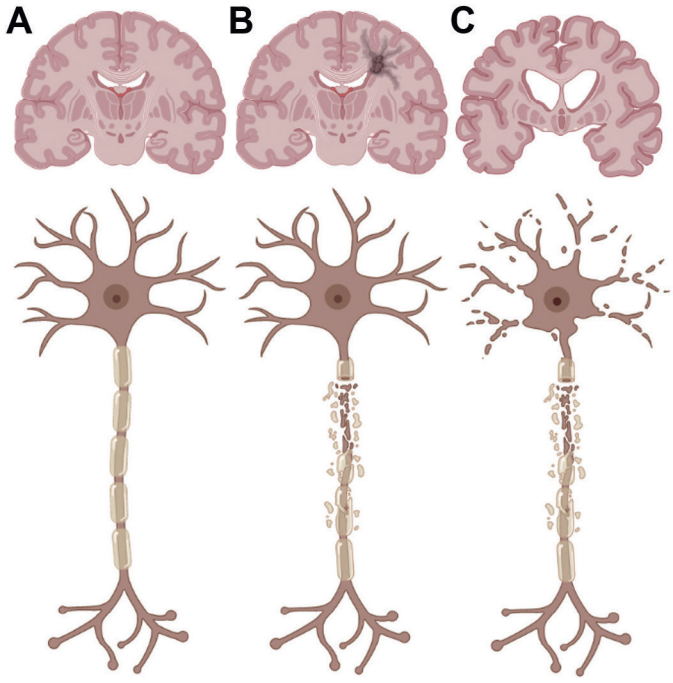

Figure 5 – Hypothesis that WM damage precedes GM atrophy in MS; starting with a healthy situation (A); followed by neuroinflammation: decreasing integrity of the WM due to MS lesions and microstructural damage (B); followed by neurodegeneration: atrophy in GM region connected to the damaged WM tract (C). Image created in BioRender.

Overview of this thesis

With a literature review (**Chapter 1**) we gained insights in the current existing literature on the cross-sectional and longitudinal relationship between WM lesions and GM atrophy in different types of MS. We focused on methodological issues regarding lesions in MRI of the MS brain in **Chapter 2**. We compared several methods for lesion segmentation with various amounts of training or supervision in Chapter 2.1, and developed freely available lesion simulation software to improve brain segmentation software protocols in the presence of MS lesions in Chapter 2.2. Since GM atrophy rates may also be related to pathology in the spinal cord, we focused in **Chapter 3** on the MS spinal cord, specifically the upper cervical cord. First, we compared

different methods for measuring cord atrophy in Chapter 3.1 and next we studied the longitudinal relationship between WM lesions and GM atrophy in the upper cervical cord and their relations to disability in Chapter 3.2. In **Chapter 4**, we investigated the longitudinal relationship between WM damage in the connected tracts and GM atrophy in two early RRMS cohorts. In Chapter 4.1, we studied WM damage of the thalamocortical tracts with single-shell DWI data and the relations with thalamus atrophy based on data from an international cohort. In Chapter 4.2, we studied the relationship between WM damage – as measured with multi-shell DWI, MWF and QSM data – and both deep and cortical GM atrophy from our own cohort, and how this relationship may relate to clinical disability.

References

1 Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B. & Ciccarelli, O. Multiple sclerosis. Lancet 391, 1622-1636, doi:10.1016/S0140-6736(18)30481-1 (2018).

2 Karussis, D. The diagnosis of multiple sclerosis and the various related demyelinating syndromes: a critical review. J Autoimmun 48-49, 134-142, doi:10.1016/j.jaut.2014.01.022 (2014).

3 Ward, M. & Goldman, M. D. Epidemiology and Pathophysiology of Multiple Sclerosis. Continuum (Minneap Minn) 28, 988-1005, doi:10.1212/CON.0000000000001136 (2022).

4 Thompson, A. J. et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 17, 162-173, doi:10.1016/S1474-4422(17)30470-2 (2018).

5 Song, X. et al. Correlation between EDSS scores and cervical spinal cord atrophy at 3T MRI in multiple sclerosis: A systematic review and meta-analysis. Mult Scler Relat Disord 37, 101426, doi:10.1016/j.msard.2019.101426 (2020).

6 Eshaghi, A. et al. Progression of regional grey matter atrophy in multiple sclerosis. Brain 141, 1665-1677, doi:10.1093/brain/awy088 (2018).

7 Brownlee, W. J., Hardy, T. A., Fazekas, F. & Miller, D. H. Diagnosis of multiple sclerosis: progress and challenges. Lancet 389, 1336-1346, doi:10.1016/S0140-6736(16)30959-X (2017).

8 Horakova, D. et al. Gray matter atrophy and disability progression in patients with early relapsing-remitting multiple sclerosis: a 5-year longitudinal study. J Neurol Sci 282, 112-119, doi:10.1016/j.jns.2008.12.005 (2009).

9 Bergsland, N. et al. Gray matter atrophy patterns in multiple sclerosis: A 10-year source-based morphometry study. Neuroimage Clin 17, 444-451, doi:10.1016/j.nicl.2017.11.002 (2018).

10 Amiri, H. et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin 19, 466-475, doi:10.1016/j.nicl.2018.04.023 (2018).

11 Steenwijk, M. D. et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain 139, 115-126, doi:10.1093/brain/awv337 (2016).

12 Filli, L. et al. Spatiotemporal distribution of white matter lesions in relapsing-remitting and secondary progressive multiple sclerosis. Mult Scler 18, 1577-1584, doi:10.1177/1352458512442756 (2012).

13 Roosendaal, S. D. et al. Grey matter volume in a large cohort of MS patients: relation to MRI parameters and disability. Mult Scler 17, 1098-1106, doi:10.1177/1352458511404916 (2011).

14 Sailer, M. et al. Focal thinning of the cerebral cortex in multiple sclerosis. Brain 126, 1734-1744, doi:10.1093/brain/awg175 (2003).

15 Riccitelli, G. et al. Mapping regional grey and white matter atrophy in relapsing-remitting multiple sclerosis. Mult Scler 18, 1027-1037, doi:10.1177/1352458512439239 (2012).

16 Muhlau, M. et al. White-matter lesions drive deep gray-matter atrophy in early multiple sclerosis: support from structural MRI. Mult Scler 19, 1485-1492, doi:10.1177/1352458513478673 (2013).

17 Battaglini, M. et al. Voxel-wise assessment of progression of regional brain atrophy in relapsing-remitting multiple sclerosis. J Neurol Sci 282, 55-60, doi:10.1016/j.jns.2009.02.322 (2009).

18 Bendfeldt, K. et al. Spatiotemporal distribution pattern of white matter lesion volumes and their association with regional grey matter volume reductions in relapsing-remitting multiple sclerosis. Hum Brain Mapp 31, 1542-1555, doi:10.1002/hbm.20951 (2010).

19 Henry, R. G. et al. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. J Neurol Sci 282, 61-66, doi:10.1016/j.jns.2009.02.379 (2009).

20 Jehna, M. et al. An exploratory study on the spatial relationship between regional cortical volume changes and white matter integrity in multiple sclerosis. Brain Connect 3, 255-264, doi:10.1089/brain.2012.0108 (2013).

21 Bodini, B. et al. White and gray matter damage in primary progressive MS: The chicken or the egg? Neurology 86, 170-176, doi:10.1212/WNL.0000000000002237 (2016).

22 Steenwijk, M. D. et al. Unraveling the relationship between regional gray matter atrophy and pathology in connected white matter tracts in long-standing multiple sclerosis. Hum Brain Mapp 36, 1796-1807, doi:10.1002/hbm.22738 (2015).

23 Garcia-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L. & Collins, D. L. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med Image Anal 17, 1-18, doi:10.1016/j.media.2012.09.004 (2013).

24 Guo, C. J., Ferreira, D., Fink, K., Westman, E. & Granberg, T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. Eur Radiol 29, 1355-1364, doi:10.1007/s00330-018-5710-x (2019).

25 Popescu, V. et al. Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. Neuroimage Clin 4, 366-373, doi:10.1016/j.nicl.2014.01.004 (2014).

26 Casserly, C. et al. Spinal Cord Atrophy in Multiple Sclerosis: A Systematic Review and Meta-Analysis. J Neuroimaging 28, 556-586, doi:10.1111/jon.12553 (2018).

27 Lukas, C. et al. Quantification of spinal cord atrophy in MS: which software, which vertebral level, spinal cord or brain MRI? A multi-centric, longitudinal comparison of three different volumetric approaches. Mult Scler J 24, 88-90 (2018).

28 Jones, D. K., Knosche, T. R. & Turner, R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. Neuroimage 73, 239-254, doi:10.1016/j.neuroimage.2012.06.081 (2013).

29 Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F. & Woolrich, M. W. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34, 144-155, doi:10.1016/j.neuroimage.2006.09.018 (2007).

30 Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61, 1000-1016, doi:10.1016/j.neuroimage.2012.03.072 (2012).

31 MacKay, A. L. & Laule, C. Magnetic Resonance of Myelin Water: An in vivo Marker for Myelin. Brain Plast 2, 71-91, doi:10.3233/BPL-160033 (2016).

32 Liu, C. et al. Quantitative Susceptibility Mapping: Contrast Mechanisms and Clinical Applications. Tomography 1, 3-17, doi:10.18383/j.tom.2015.00136 (2015).

33 Zhang, Y. et al. Longitudinal change in magnetic susceptibility of new enhanced multiple sclerosis (MS) lesions measured on serial quantitative susceptibility mapping (QSM). J Magn Reson Imaging 44, 426-432, doi:10.1002/jmri.25144 (2016).

CHAPTER 1.2

The relationship between white matter lesions and gray matter atrophy in multiple sclerosis: a systematic review

Ingrid Anne Lie
Merlin M. Weeda
Rozemarijn M. Mattiesing
Marijke A.E. Mol
Petra J.W. Pouwels
Frederik Barkhof
Øivind Torkildsen
Lars Bø
Kjell-Morten Myhr
Hugo Vrenken

Neurology 98 (2022)

Abstract

Background and objectives

There is currently no consensus about the extent of gray matter (GM) atrophy that can be attributed to secondary changes after white matter (WM) lesions or the temporal and spatial relationships between the two phenomena. Elucidating this interplay will broaden the understanding of the combined inflammatory and neurodegenerative pathophysiology of multiple sclerosis (MS), and separating atrophic changes due to primary and secondary neurodegenerative mechanisms will then be pivotal to properly evaluate treatment effects, especially if these treatments target the different processes individually. To untangle these complex pathologic mechanisms, this systematic review provides an essential first step: an objective and comprehensive overview of the existing in vivo knowledge of the relationship between brain WM lesions and GM atrophy in patients diagnosed with MS. The overall aim was to clarify the extent to which WM lesions are associated with both global and regional GM atrophy and how this may differ in the different disease subtypes.

Methods

We searched MEDLINE (through PubMed) and Embase for reports containing direct associations between brain GM and WM lesion measures obtained by conventional MRI sequences in patients with clinically isolated syndrome and MS. No restriction was applied for publication date. The quality and risk of bias in included studies were evaluated with the Quality Assessment Tool for observational cohort and cross-sectional studies (NIH, Bethesda, MA). Qualitative and descriptive analyses were performed.

Results

A total of 90 articles were included. WM lesion volumes were related mostly to global, cortical and deep GM volumes, and those significant associations were almost without exception negative, indicating that higher WM lesion volumes were associated with lower GM volumes or lower cortical thicknesses. The most consistent relationship between WM lesions and GM atrophy was seen in early (relapsing) disease and less so in progressive MS.

Discussion

The findings suggest that GM neurodegeneration is mostly secondary to damage in the WM during early disease stages while becoming more detached and dominated by other, possibly primary neurodegenerative disease mechanisms in progressive MS.

Introduction

Gray matter (GM) atrophy occurs in patients with multiple sclerosis (MS)[e1] already in early disease stages.[e2,1] Reflecting axonal loss and irreversible neuronal damage,[2] GM atrophy can be measured noninvasively in vivo from standard MRI. It is considered a marker of neurodegeneration that could help bridge the current gap between measures of clinical disability and traditional inflammatory MRI markers.[3]

Recent work has found that MS pathology affects both GM and white matter (WM) structures throughout the CNS. Therefore, it is unlikely that disability progression and worsening of higher functions such as cognition can be strongly predicted by a single MRI marker.[4] Nevertheless, brain GM atrophy is associated with several clinical outcomes: GM volumes are lower in people with MS than in healthy controls,[5] may predict conversion from clinically isolated syndrome (CIS) to MS,[6,7] and relate to disability progression.[8] Moreover, GM atrophy relates strongly with cognitive dysfunction[9-13] and more so than WM lesion volume (LV).[14]

WM lesions have been the principal imaging marker of disease activity and progression in MS and are incorporated into diagnostic criteria[15] and treatment goals,[16] as well as outcome measures in research trials. These focal areas of demyelination, consisting of inflammation and variable gliosis,[17] can be visualized as hyperintense or hypointense lesions in T2- and T1-weighted MRIs, respectively.[3]

If and how WM lesions and GM atrophy are temporally, spatially, and causally related are insufficiently clear. Elucidating this interplay will not only broaden understanding of the combined inflammatory and neurodegenerative pathophysiology of MS but also provide reliable biomarkers for research and therapeutic purposes. As treatment targets expand from inflammatory lesions to neurodegenerative processes, GM atrophy is a natural choice of outcome measure. Separating atrophic changes due to primary and secondary neurodegenerative mechanisms will then be crucial to properly evaluate treatment effects, especially if these treatments target the different processes individually. While some studies have addressed the relationship between WM lesions and GM atrophy directly, a larger body of literature reports measures of both. In this systematic review, we have therefore aimed to review this existing evidence in its entirety to establish how brain WM lesions and GM atrophy in MS are related.

Methods

This review was conducted and presented according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[18]

Search Strategy

To select studies of relevance to this systematic review, the electronic databases Medline (through PubMed) and Embase were searched. The search strategies were developed in consultation with a medical librarian (M.A.E.M.). Thesaurus terms and free-text words, including synonyms and closely related words, were used for the following concepts: MS, GM atrophy,

and WM lesions. No restrictions were applied for language (at this stage) or publication date, but conference abstracts were excluded. The search strategy is detailed in **eAppendix 1**. The last search was conducted on August 17, 2020.

Eligibility Criteria

Studies were included if they fulfilled all following criteria: (1) controlled trials or observational studies in English and published in a peer-reviewed journal; (2) trials or studies that involved patients diagnosed with CIS or MS; and (3) study abstract containing associations between brain GM and WM lesion measures obtained by conventional MRI sequences. To limit the scope of this review and the possible variability in pathologic substrates and disease mechanisms, we excluded studies of patients diagnosed with pediatric MS or with radiologically isolated syndrome.

Outcome Measures

The primary outcome measures of interest were direct associations made between brain WM lesion and GM atrophy measures, obtained by conventional MRI sequences, in patients diagnosed with CIS or MS.

Selection Process

After excluding duplicate publications, we screened the remaining abstracts on selection criteria by 2 independent raters (H.V., I.A.L.) using Rayyan software,[19] a web-based application designed for systematic reviews.[20] Conflicting selections were discussed until consensus. When eligibility could not be determined from the title and abstract alone, full texts of potentially relevant articles were consulted.

Data Extraction and Quality Assessment

Independent extraction and quality assessment of relevant data from each included article were conducted by at least 2 reviewers (I.A.L., M.M.W., R.M.M., H.V.), according to a customized checklist. The quality and risk of bias in included studies were further evaluated with the Quality Assessment Tool for observational cohort and cross-sectional studies (NIH, Bethesda, MA). A rating scale of yes = 1, no = 0, and not reported = 0 was applied for the 14 questions of the checklist, and the final study quality was rated, in consensus between the raters (I.A.L. and H.V.), as good, fair, or poor on the basis of individual scores and the severity of the risk of bias. To visually illustrate the main results for the different disease phenotypes, composite figures were prepared combining available, clear figures from key studies.

Results

Through the initial search, 3,750 records were identified. After the updated search and removal of duplicates, 2,260 citations were screened on title and abstract, resulting in 106 full-text articles considered, of which 90 articles met the inclusion criteria and were included in this review (**Figure 1**). The 90 studies are listed in the **eReferences** (e1–e90), and the study design

in all included articles is described in **Table 1**. Last, the quality assessment rate for each study is reported in **eTables 1-3**.

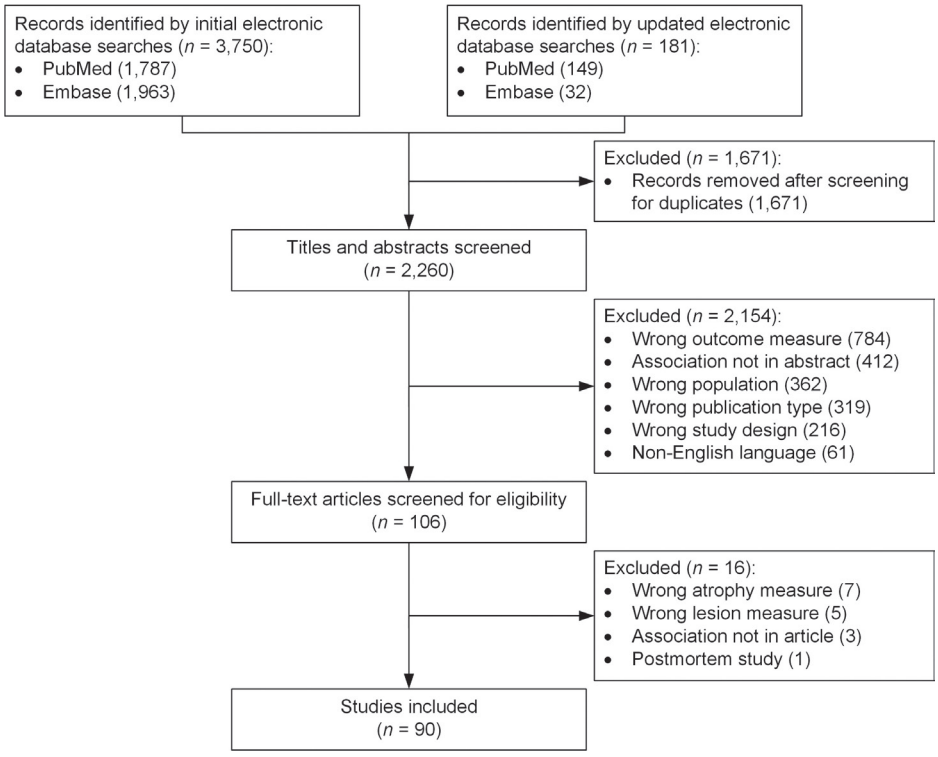

Figure 1 – Flowchart demonstrating the selection process.

Clinically Isolated Syndrome

Eight cross-sectional and 4 longitudinal studies investigated patients diagnosed with CIS. In the longitudinal studies, the follow-up period ranged from 2 to 5.5 years.

The association of lesions with global GM measures were reported in 5 studies, while cortical GM (CGM) and deep GM (DGM) measures were each considered in 7 studies. Four studies reported regional WM lesion measures.

Included studies are described in **eTables 1-3**, and a more detailed discussion of results of each section is in **eAppendix 2**.

Global GM in CIS

In 2 of 3 cross-sectional CIS studies, no significant association was found between global GM volume and either T2[e3] or T1[e4] LV. One study found a significant correlation between T2 LV and global GM volume ($r = -0.56$, $p < 0.020$).[e5]

Table 1 – Study design in included studies.

Study design	*N*
Cross-sectional studies	**64**
Observational case-control	50
Observational cohort	11
Clinical trial	3
Longitudinal studies	**18**
Observational case-control	3
Observational cohort	12
Clinical trial	3
Cross-sectional and longitudinal studies	**8**
Observational case-control	4
Observational cohort	2
Clinical trial	2

The longitudinal relationship of global GM volume with global WM lesion measures was reported in 2 studies (follow-up time ranging from 2–3 years), both observing significant but different associations. In 1 study, change in global GM fraction correlated with WM LV changes (r values ranging from −0.3071 to −0.4280, p values from 0.0032 to 0.0426) but not with baseline lesion measures,[e6] while the other study found associations with baseline lesion measures ($p \leq 0.004$), but not with LV changes.[e7]

CGM in CIS

In cross-sectional CIS studies, lower CGM volume showed variable associations with global WM lesion measures. Two studies observed a significant relationship with the presence ($t = 2.48$, $p = 0.020$)[e8] or volume ($r = -0.49$, $p = 0.045$)[e5] of T2 lesions, while 3 studies did not.[e2,e3,e8] Of 2 studies reporting regional WM lesion measures, 1 study found a significant association between regional cortical thickness and T2 LV ($p \leq 0.0466$),[e9] while the other did not.[e8]

Of the 2 available longitudinal studies, 1 study found no relations,[e10] while the other found significant associations of cortical volume change over 48 months with baseline WM lesion measures ($p \leq 0.004$) and the total cumulative number of new/enlarging T2 lesions ($p = 0.036$), while no associations were observed for LV changes.[e7]

DGM in CIS

In the 5 available cross-sectional studies in patients with CIS, all except 1 study[e3] showed significant associations between global[e2,e4] and regional[e11,e12] WM LV and total[e2] and regional DGM volumes[e2,e4,e11,e12] (p values ranging from <0.0001–0.05). In contrast, no associations with DGM volumes were found for global T2 lesion number or the presence of gadolinium-enhancing lesions.[e2] Of the regional DGM volumes investigated, the most consistent relationships were

found for the thalamus and hippocampus. This pattern was true considering both global[e2,e4] and regional WM LV.[e11,e12]

Longitudinally, 1 of the 3 available studies found that regional DGM atrophy was related to global baseline lesion measures ($p \leq 0.018$),[e7] but there was no relationship with changes in global[e7,e10] or regional[e11] LV (follow-up times between 2 and 5.5 years).

Relapsing-Remitting MS

Overall, 37 cross-sectional and 14 longitudinal studies reported associations between WM lesion measures and GM atrophy in relapsing-remitting MS (RRMS). The follow-up period of the available longitudinal studies ranged from 1 to 5.5 years.

Seventeen publications reported the relationship with global GM, 29 on that with CGM, and 25 on that with DGM measures. Eleven studies considered regional WM lesion measures.

Included studies are described in **eTables 1-3**, and a more detailed discussion of results of each section is given in **eAppendix 2**.

Figure 2 illustrates the main results from this section.

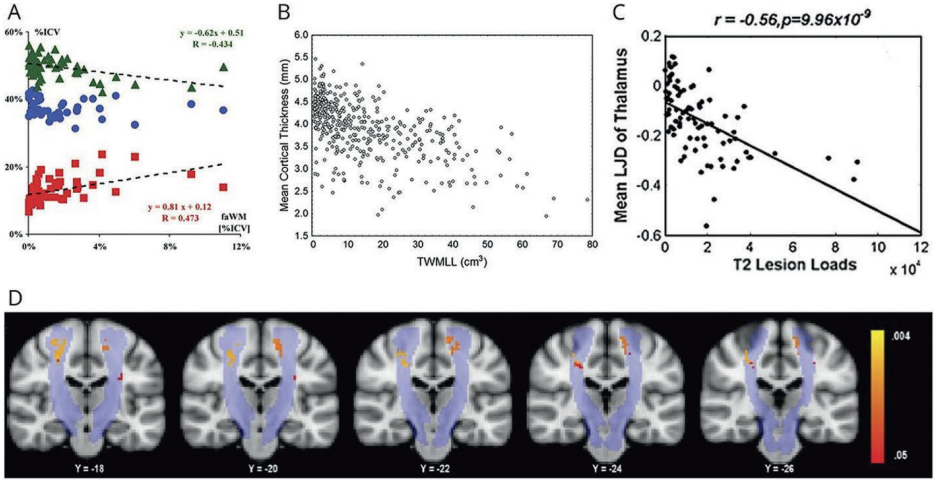

Figure 2 – RRMS shows consistent associations between WM lesions and GM volume.

(A) Scatterplot of the fractional volumes of GM (fGM, green triangles), WM (fWM, blue circles) and CSF (fCSF, red boxes), versus the fractional volume of abnormal WM (faWM), all expressed as percentages of intracranial volume. fGM and fCSF values are adjusted to patients mean age (35.6 years). Where significant, regression lines are shown along with the corresponding equations and R values. Increasing loss of GM volume, with a corresponding increase in CSF volume, is apparent with increasing faWM. fWM (which includes also the WM lesion volume) is not significantly changed with increasing faWM. Reproduced with permission from Quarantelli et al., 2003[e20]. (B) Scatterplot showing the relationship between mean cortical thickness in millimeters and total white matter lesion load (TWMLL) in cubic centimeters in 425 RRMS patients. Reproduced with permission from Charil et al., 2007[e35]. (C) Correlation of the mean logarithm of the Jacobian determinant (a measure of atrophy) with T2 lesion load in 88 RRMS patients for the thalamus. Reproduced with permission from Tao et al., 2009[e37]. (D) Lesional voxels that significantly correlate with primary motor cortex thickness are shown in red-yellow. The probabilistic CST atlas is shown in light blue. Reproduced with permission from Bergsland et al., 2015[e47].

Abbreviations: RRMS, relapsing-remitting multiple sclerosis; GM, gray matter; WM, white matter; fGM, fractional gray matter volume; fWM, fractional white matter volume; fCSF, fractional cerebrospinal fluid volume; faWM, fractional abnormal white matter volume; CST, corticospinal tract.

Global GM in RRMS

The majority of available cross-sectional RRMS studies, i.e., 8 of 10 studies, observed significant associations between global GM volume and global WM lesion load. Eight studies observed a significant association between global GM volumes and T1[e13,e14] and T2 LV[e13-e19] and abnormal WM[e20] (*p* values ranging from <0.001–0.047). In contrast, 2 studies considering T2 LV[e21,e22] and another 2 studies considering gadolinium-enhancing LV[e13,e14] did not.

One cross-sectional study investigated the impact of regional LV on total GM volume and reported a significant correlation with regional T1 and T2 LV in 3 and 4 of 26 WM regions, respectively (*r* values ranging from −0.20 to −0.50, $p < 0.001$).[e23]

Of the 7 longitudinal studies available, 4 did not find an association between global GM atrophy progression and global WM lesion measures. When considering gadolinium-enhancing lesion measures obtained at baseline, 1 study found a significant association ($p = 0.04$),[e24] while 3 others did not find that global GM atrophy progression related to the presence,[e25,e26] number,[e26] or volume[e14] of gadolinium-enhancing lesions (follow-up time ranging from 1–4 years).

Three of 5 studies considering longitudinal WM lesion changes[e14,e24,e27-e29] with a follow-up period between 1 and 4 years observed significant associations between longitudinal changes in T1[e27] and T2[e24,e27,e28] LV and GM atrophy progression (*p* values ranging from 0.0004–0.03).

CGM in RRMS

A majority of cross-sectional studies (14 of 19) considering global WM LV found significant associations. WM LV was found to relate negatively to both total cortical volume (*p* values ranging from <0.0001–0.05)[e2,e15,e30-e32] and global cortical thickness (*p* values ranging from <0.001–<0.05).[e17,e30,e33-e35]

A total of 6 studies explored global T1[e36,e37] and T2[e16,e19,e30,e36-e38] lesions and their relationship with regional cortical volume, with the most consistent and strongest associations in areas in the frontal, temporal, cingulate, and insular cortex. A similar pattern of associations was seen for cortical thickness measures.[e30,e35,e39]

Five studies did not find significant relations for either cortical volume[e21,e40-e42] or cortical thickness.[e39]

In 6 of the 8 cross-sectional studies[e9,e23,e35,e43-e47] considering regional distribution of WM lesions, the results suggested an anatomic or structural relationship between lesion location and regional cortical volume[e44,e45] and thickness.[e9,e35,e46,e47]

Considering global WM lesion measures, 3 of five[e10,e27,e32,e48,e49] longitudinal studies found significant relationships between both baseline WM lesion measures[e49] and on-study changes in WM LV[e27,e48] or numbers[e48] and cortical thinning ($p = 0.040$),[e48] as well as regional ($p < 0.01$) [e27] and total cortical volume loss (*p* values ranging from <0.0001–0.010)[e49] (follow-up time ranging from 1–2 years).

Of the 2 studies assessing regional WM LV, 1 study observed visually that the increase in T2 LV spatially coincided with areas of cortical decrease,[e50] while the other study did not.[e51]

DGM in RRMS

With the exception of 1 study,[e52] all 17 cross-sectional publications reporting global WM LV found significant associations with DGM volume measures. Three studies evaluated DGM volume as a whole (*p* values ranging from <0.0001–0.04),[e17,e49,e53] while the remaining assessed the various structures separately. Thalamic volume and surface displacement[e54] were associated negatively with T1[e36,e37] and T2[e2,e16,e30,e36-e38,e40,e41,e49,e54,e55] LV in 11 studies (*p* values ranging from <0.00001–<0.05). Other DGM structures repeatedly showing significant associations with WM LV were the caudate nucleus (*p* values ranging from <0.0001–<0.05),[e2,e19,e36-e38,e41,e42,e55,e56] putamen (*p* values ranging from <0.00001–<0.05),[e2,e30,e36-e38,e53,e55] and globus pallidus (*p* values ranging from <0.0001–<0.05).[e2,e30,e38,e54,e55]

While 2 cross-sectional studies did not find any associations between regional WM lesion and DGM measures,[e23,e45] the majority of studies did.[e43,e44,e46,e54]

All 4 publications that assessed longitudinal relations between total and regional DGM atrophy and global WM lesion measures observed significant associations.[e10,e48,e49,e57] The associations were found for both baseline WM lesion measures (*p*<0.0001 and 0.037)[e49,e57] and on-study new/enlarging T2 lesions or new gadolinium-enhancing lesions (*p*= .024).[e48]

Secondary Progressive MS

Eleven cross-sectional and 2 longitudinal studies reported patients with secondary progressive MS (SPMS), 5 of which explored associations between WM lesions and global GM volume, while 8 and 10 studies focused on CGM and DGM measures, respectively. Four studies considered regional WM lesion measures.

Included studies are described in **eTables 1** and **2**, and a more detailed discussion of results of each section is given in **eAppendix 2**.

Figure 3 illustrates the main results from this section.

Global GM in SPMS

Three of 4 cross-sectional studies reported negative associations between WM LV and global GM volume (*r* values ranging from −0.36 to −0.72, *p* values ranging from <0.001–<0.01).[e16,e17,e58] The 1 study considering regional T1 and T2 LV and global GM volume found no significant associations.[e23]

Longitudinally, neither baseline nor on-study changes in WM lesion measures predicted changes in GM volume over the 4-year follow-up.[e24]

CGM in SPMS

The observed relationship between cortical volume or thickness and global WM lesion load in patients with SPMS was not consistent in the 4 available studies. Two studies found significant associations with lower cortical volume (*p* < 0.001 and <0.05).[e16,e40] Furthermore, cortical thickness was evaluated in another 2 studies based on 1 and the same study population; neither study found any significant association between T2 LV and global mean cortical thickness.[e17,e34]

One of 4 cross-sectional studies assessing regional WM lesions observed relatively strong correlations between lower cortical volume and T2 LV in the same or adjacent lobes (*r* values

ranging from −0.67 to −0.79, $p < 0.001$).[e40] In the 3 remaining studies, the associations with lower cortical volume or thickness were weak[e45,e46] or nonsignificant.[e23]

The only longitudinal study available investigated atrophied T2 LV (T2-weighted lesional tissue subsequently substituted by CSF) in patients with SPMS and primary progressive MS (PPMS) in a combined progressive MS group, finding no associations with baseline cortical volume or volume change (follow-up time 5.5 years).[e10]

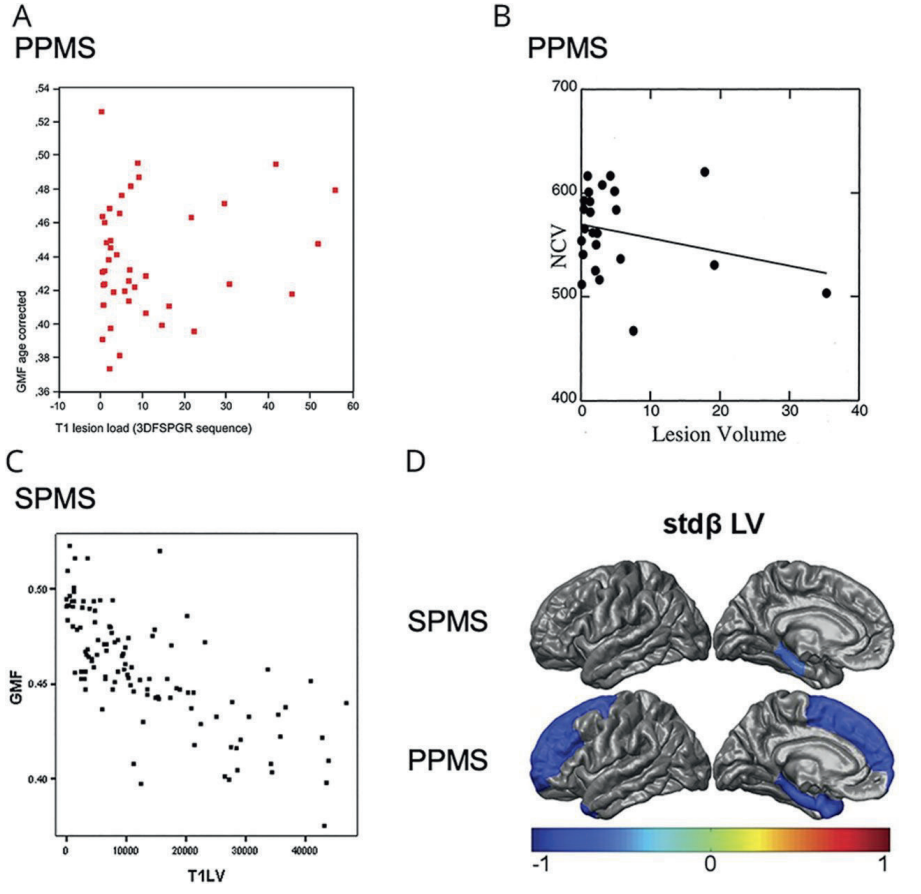

Figure 3 – Progressive MS shows varying associations between WM lesions and GM volume.

(A) In 43 patients with PPMS, GMF corrected for age is plotted against T1 lesion load (in ml; derived from 3D FSPGR scans), illustrating an absence of correlation. Reproduced with permission from Sastre-Garriga et al., 2004[e59]. (B) Data illustrating the absence of correlation between NCV and T2 lesion volume in 25 patients with PPMS (Spearman rank coefficient r=-0.1, p=0.6). Reproduced with permission from De Stefano et al., 2003[e31]. (C) Scatter plot of T1LV against GMF in 117 patients with SPMS, illustrating a significant correlation (r=-0.72, p<0.001). Reproduced with permission from Furby et al., 2009[e58]. (D) Graphical visualization of the cross-sectional relation between regional cortical thickness and WM measures in 53 patients with longstanding SPMS (top row) and 25 patients with longstanding PPMS (bottom row), as assessed through linear regression. In gray areas, lesion volume in the connected WM tracts did not contribute significantly to the model explaining regional cortical thickness, whereas in colored areas, the colors correspond to the standardized beta values of lesion volume in the connected tracts for the respective regional model. Reproduced with permission from Steenwijk et al., 2015[e46].

Abbreviations: MS, multiple sclerosis; LV, lesion volume; GMF, gray matter fraction; SPMS, secondary progressive multiple sclerosis; PPMS, primary progressive MS; FSPGR, fast spoiled gradient recalled echo; NCV, normalized cortical volume; WM, white matter.

DGM in SPMS

In the 6 cross-sectional publications that considered global WM lesions, results were somewhat conflicting. Two studies found no associations with lower DGM volume[e16,e55]; in the other 4 studies, however, T1[e53] or T2[e17] LV was associated significantly with both total DGM volume (*p* values ranging from <0.01–0.04) and separate DGM structures such as the hippocampus,[e52] thalamus, and caudate nucleus[e40] (*r* values ranging from −0.69 to −0.88, *p* < 0.001–0.018). Two of 4 studies[e23,e40,e45,e46] that included results for regional WM LV or distribution showed significant associations[e40,e46] (*p* values ranging from <0.001–<0.05).

The sole longitudinal study found no association between atrophied T2 LV (described in the previous section) and baseline thalamic volume or volume change.[e10]

Primary Progressive MS

The relationship between WM lesions and GM measures in patients with PPMS was assessed in 11 cross-sectional studies and longitudinally in 2 studies.

Three studies performed analyses involving global GM volume, while CGM and DGM measures were each considered in 9 studies. Three studies considered regional WM LV or distribution.

Included studies are described in **eTables 1** and **2**, and a more detailed version of the respective sections is given in **eAppendix 2**.

Figure 3 illustrates the main results from this section.

Global GM in PPMS

The cross-sectional associations between global WM lesion measures and global GM volume in patients with PPMS were variable. One study reported a significant correlation with T2 LV ($r = -0.68$, $p < 0.001$),[e17] while the other found no significant associations with either T2, T1, or gadolinium-enhancing LV or lesion numbers.[e59]

In the available longitudinal study, baseline WM lesion measures were not related to GM volume change over 12 months.[e60]

CGM in PPMS

Cross-sectional results for global WM lesion measures and CGM volume or thickness in patients with PPMS were divided. Three studies found associations between T1[e61] and T2[e34,e40] LV and total ($r = -0.508$, $p < 0.05$)[e61] and regional (*r* values ranging from −0.605 to −0.85, *p* values ranging from <0.001–<0.01)[e40] cortical volume and total cortical thickness ($p < 0.05$).[e34] In the other 3 studies, no significant associations were found for either cortical volume[e31,e62] or thickness.[e17]

Of 3 publications assessing regional WM lesion measures, 1 study found a relationship with cortical volume in anatomically connected areas ($p < 0.001$),[e40] while in the other 2, the associations with cortical thickness or volume were weak[e46] or absent.[e45]

Only 1 longitudinal study was identified, finding no associations between atrophied T2 LV (described in previous section) and baseline cortical volume or volume change.[e10]

DGM in PPMS

All but 1[e55] of the 6 cross-sectional studies reporting the relationship between global WM lesions and DGM volume observed significant associations. In patients with PPMS, correlations were significant for both total DGM volume (*r* values ranging from −0.651 to −0.71, *p* values ranging from <0.001–<0.01)[e17,e61] and the separate structures. The most consistent association with global WM LV was seen for the thalamus[e62] for both T2 (*r* values ranging from −0.48 to −0.94, *p* values ranging from <0.001–<0.05)[e40,e61,e63] and T1 (*r* values ranging from −0.44 to −0.554, *p* values ranging from 0.002–<0.05)[e61,e63] LV.

Of 3 cross-sectional publications assessing regional WM lesions,[e40,e45,e46] lower regional DGM volume was related to regional T2 LV in 2 studies (*p* values ranging from <0.001–<0.05).[e40,e46] Again, only 1 longitudinal study was available, and for both baseline thalamic volume and volume change, no relationship to atrophied T2 LV (described in the previous section) was found.[e10]

Results for Mixed MS Groups

A number of studies[e1,e9,e10,e16,e17,e23,e34,e46,e53,e55,e64-e90] reported analyses relating GM atrophy measures to WM lesion measures in heterogeneous groups of patients with MS encompassing different disease phenotypes. Full results of these studies are reported in **eAppendix 2**. Briefly, in most cross-sectional studies, GM atrophy and WM lesions were significantly associated; in longitudinal studies, results were more variable.

Comparisons Between Disease Phenotypes

Some of the studies discussed in the previous sections included multiple disease phenotypes in a single study. Such a design eliminates differences between image acquisition and image analysis approaches that may otherwise account for differences between disease phenotypes observed from separate studies and therefore can shed the most direct light on whether the relationship between GM atrophy and WM lesions might differ between disease types. **Table 2** summarizes the observed associations in articles including multiple phenotypes, and full reports of the studies are given in **eAppendix 2**. We focus on whether the observations differed between disease types and, when available, on the direct statistical comparisons between disease types. In summary, for global GM, CGM, and DGM, both cross-sectional and longitudinal studies found the most consistent associations with WM lesions in RRMS, while the associations for CIS, SPMS, and PPMS were more variable (**Table 2**). In 11 of 15 studies, the largest patient group consisted of patients with RRMS, often in a great majority. Such imbalance may cause the studies to detect significant associations only in the larger patient group, merely because of power and not due to a lack of true association in the smaller (progressive) patient group.

The included studies are described in **eTables 1** and **2**.

Table 2 – Studies including multiple disease phenotypes.

Study	CIS	RRMS	SPMS	PPMS
Association between investigated measures	Present	Present	Present	Present
Global GM – cross-sectional studies				
Global lesion volume				
Reference e16	—	✓	✓	—
Reference e17	—	✓	✓	✓/✗
Regional lesion volume				
Reference e23	—	✓	✗	—
Global GM – longitudinal studies				
Global lesion volume				
Reference e24	—	✓	✗	—
Cortical GM – cross-sectional studies				
Global lesion volume				
Reference e2	✗	✓	—	—
Reference e16	—	✓	✓	—
Reference e17	—	✓	✗	✓/✗
Reference e31	—	✓	—	✗
Reference e34	—	✓	✗	✓
Reference e40	—	✗	✓	✓
Regional lesion volume				
Reference e9	✓	✓	—	—
Reference e46	—	✓	✓	✓
Reference e23	—	✗	✗	—
Reference e45	—	✓	✓	—
Reference e40	—	—	✓	✓
Cortical GM – longitudinal studies				
Global lesion volume				
Reference e10	✗	✗	✗	✗
Deep GM – cross-sectional studies				
Global lesion volume				
Reference e2	✓	✓	—	—
Reference e17	—	✓	✓	✓/✗
Reference e53	—	✓	✓/✗	—
Reference e52	—	✗	✓	—
Reference e16	—	✓	✓	—
Reference e40	—	✓	✓	✓
Reference e55	—	✓	✗	✗

[continued on next page]

Table 2 – *[continued]*

Study	CIS	RRMS	SPMS	PPMS
Association between investigated measures	Present	Present	Present	Present
Regional lesion volume				
Reference e23	—	✗	✗	—
Reference e45	—	✓	✓	✗
Reference e46	—	✓	✓	✓
Reference e40	—	—	✓	✓
Deep GM – longitudinal studies				
Global lesion volume				
Reference e10	✗	✓	✗	✗

Abbreviations: CIS = clinically isolated syndrome; GM = gray matter; PPMS = primary progressive multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis. Presence or absence of the association between white matter lesion measures and GM atrophy measures is indicated as ✓ (association present), ✗ (association not present), or ✓/✗ (association present for some analyses, not present for other analyses). If the disease phenotype was not investigated in the study, this is indicated as —.

Discussion

This systematic review assessed the existing evidence regarding an association between brain WM lesions and GM atrophy in MS. Surveying results from cross-sectional and longitudinal studies of different phenotypes and with varying anatomic regions of interest has resulted in a comprehensive picture. More WM lesions were associated with more GM atrophy (**Table 3**), especially in RRMS and less consistently so in progressive MS.

Table 3 – Main findings.

Main findings
No 1: More WM lesions, more GM atrophy
In cross-sectional studies in particular, WM lesion volumes were mostly related to global, cortical and deep GM volumes, and those significant associations were almost without exception negative, indicating that higher WM lesion volumes were associated with lower GM volumes or lower cortical thicknesses.
No 2: WM lesions are most clearly linked to GM atrophy in RRMS
The most consistent relation between WM lesions and GM atrophy was seen in patients with RRMS. In this relapsing phenotype, significant associations were found in the majority of studies considering global, cortical and deep GM. A relation with deep GM and especially thalamus volumes was particularly consistent in RRMS, and also in CIS.
No 3: In progressive disease WM lesions are mostly, but less consistently, linked to GM atrophy
Studies of the progressive disease types showed more variable associations: for both SPMS and PPMS, WM lesion measures were related to global GM volume in the majority of studies, but for cortical and deep GM, associations were less consistent.

Abbreviations: CIS = clinically isolated syndrome; GM = gray matter; PPMS = primary progressive multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis; WM = white matter.

The quality of evidence was mostly rated as fair, with no correction for potential confounders (e.g., therapeutic and physiologic factors), short follow-up time, and small or unbalanced disease groups (as highlighted in the previous section) as the main risks of bias.

The clear trend emerging from cross-sectional and longitudinal studies for both global and regional associations was that more WM lesions were related to more or faster GM atrophy. Patients with high WM lesion burden may be expected to also have extensive damage to other brain structures, not necessarily because one causes the other but possibly also because an advanced disease stage acts as a common denominator. To further investigate, each disease type was evaluated and compared; the association was observed frequently in all disease types, most consistently in RRMS. However, the relationship was more variable for longitudinal than for cross-sectional outcomes.

In mixed MS groups, the lack of significant associations in longitudinal studies could be related to group heterogeneity. Furthermore, variable treatment regimens across patients affect the interpretation of all studies, especially more recent longitudinal studies. Here, the time that patients have spent under potent treatment is often considerable and may modulate not only the observed association between WM lesions and GM atrophy but also the main pathologic substrate of the neurodegenerative process.

Current knowledge from neuroimaging and histopathology implies that GM neurodegeneration is driven both by events secondary to WM inflammation and by primary disease mechanisms within the GM. Adding to the complexity, these mechanisms seem to act simultaneously, with additive effects.[21] Strong and consistent associations with WM lesions were found in all GM regions in RRMS and in DGM in CIS; this suggests that early GM neurodegeneration is mainly secondary to damage in the WM: after chronic inflammation in WM, neuronal injury and damage to mitochondria with resulting energy deficiency initiate several neurodegenerative cascades. The degenerative process can move forward toward the axonal terminal (anterograde or wallerian degeneration) or backward toward the cell soma (retrograde degeneration), leading to neuronal loss and atrophy in connected GM regions.[22] In both CIS and RRMS, the most consistent relations were seen in DGM and the thalamus. Connecting and relaying information between subcortical areas and the neocortex through different WM tracts,[23] it seems plausible that thalamic GM components are vulnerable to damage through retrograde degeneration.[24]

In progressive MS phenotypes, while GM atrophy was more widespread, affecting most DGM structures[e40,e46,e55] and cortical areas,[e40,e46] the relationship with focal WM lesions was more varied although still present in the majority of studies.[e17,e34,e46,e55] These results, interpreted together with neuropathologic studies showing continued, widespread GM atrophy development, at least partly independently of focal inflammatory WM lesions,[25,-,27] suggest that in progressive MS the neurodegenerative disease mechanism may be a mainly primary process. Alternatively, in long-standing MS with many tracts affected, the relationships between primary lesional damage and downstream GM atrophy may become too complex and too variable across individuals to disentangle. Furthermore, GM lesions, often found more prominently in progressive MS, may also propagate GM atrophy and contribute to its less consistent association with WM lesions.[e53] In addition, consistent GM atrophy patterns found in patients with CIS[e70] suggest that some primary degenerative processes may be present throughout the disease.

The reviewed literature suggests that the mechanisms of neurodegeneration in MS are not static through the disease course, so the therapeutic targets, interventions, and subsequent monitoring will most likely differ for the various patient groups. To obtain fully individualized and optimized patient treatment, we have summarized important research aims and suggestions for future research in **Table 4**.

Table 4 – Research questions and suggestions for future research.

Research questions	Suggestions
Understand the details of the spatio-temporal relationship between white matter lesions and gray matter atrophy in multiple sclerosis.	- Imaging studies preferably of longitudinal design, focusing on pathology in defined structurally or functionally connected regions. - Minimize technical inter- and intra-study variability in imaging studies, and include acquisitions needed to detect and study relevant pathology (e.g., GM lesions).
Untangle the neurodegenerative processes secondary to focal inflammatory damage, from those primarily arising in the gray matter.	- Combining and/or interpreting results of imaging studies in context with knowledge obtained through histopathological and/or molecular studies.
Determine which neurodegenerative process is the dominant driver of gray matter atrophy in different stages of the disease.	- Investigate each neurodegenerative process separately for each disease phenotype, defined not only by clinical characteristics, but also by biological and imaging markers to better capture the dominant pathological substrate(29). - Consider the type and duration of therapeutic interventions of which included subjects have received.
Develop therapeutic interventions targeting specific neurodegenerative processes.	- Specify the neurodegenerative pathway targeted in clinical trials, interpret results separately for each disease phenotype and compare them directly.

Abbreviations: GM, gray matter.

Our study has several limitations. Diagnostic criteria and hence the separation between CIS and MS varied over time. The effect of physiologic variability and therapeutic interventions was not consistently accounted for in the reviewed articles. Whether treatment was used and what type were mostly stated but rarely adjusted for in the analyses. Therefore, effects of individual treatments on WM lesions or (primary or secondary) GM atrophy, which potentially change the observed relationship between the 2 processes for each patient, could cloud our interpretation of the disease mechanisms.

In longitudinal studies, the group sizes were often smaller, and the majority followed up the patients for ≤2 years. Such short follow-up durations most likely affected the ability to detect temporal associations, considering that neurodegeneration is a slowly progressive process. [e24]

Moreover, brain atrophy is cumulative and may exhibit a ceiling effect and delayed effects from previous exposures or previous pathologic damage.[28]

Technical factors are well known to affect brain measurements[28]: intrastudy and interstudy variability in MRI scanners and acquisitions (e.g., field strength, slice thickness, 2-/3-dimensional acquisitions, pulse sequence type and parameters), image (pre)processing tools, and analysis software. This makes the interpretation and comparison of results challenging. The 20-year time frame of included articles, during which MRI technology has achieved major leaps of improvement, means that earlier studies have to be evaluated in the light of the concurrently available technology and knowledge. Furthermore, the large variability in image acquisition, analysis methods, and outcome measures, combined with uncertainties about potential confounders such as treatment, made it impossible to conduct a meaningful and interpretable meta-analysis of the results reported in the reviewed articles.

Statistical issues may also have influenced results. Sample sizes were often unbalanced between disease types, especially with small progressive groups being compared to larger relapsing-remitting groups. Furthermore, the majority of studies focused on patients with RRMS, which limits our ability to draw conclusions for progressive disease types.

To elucidate the pathophysiologic relationship between inflammatory WM lesions and neurodegenerative changes in GM, this review has an obvious limitation in that statistical associations do not prove causation. However, the many imaging studies included provide the possibility to investigate these relations in vivo in a large number of patients in different disease stages. Although in this study a spatiotemporal relationship between changes in GM structures and WM lesions was found, we cannot draw any conclusions about whether this process starts with demyelination in WM or whether the primary defect is in the axon or neuron itself, with demyelination as a secondary effect.[29] To widen this question of causality, some researchers suggest that the association seen between WM lesions and lower GM volume in certain regions is not causally linked through axonal degradation but is mainly due to a common close proximity to inflammatory soluble factors in the CSF.[e9]

Due to capacity and limiting the scope of this systematic review, we included only MRI measures obtained by conventional MRI sequences. Advanced imaging methods would be interesting to review, which by necessity would require more attention to the myriad technical differences between such studies.

We found that the majority of the literature overwhelmingly reported an association between WM lesions and global or regional GM atrophy. The association was most consistent in RRMS but more variable in progressive phenotypes and CIS. This suggests that GM neurodegeneration is mostly secondary to damage in the WM during early disease stages, while more detached and dominated by other, possibly primary neurodegenerative disease mechanisms in progressive MS. These findings are of great importance for patient treatment and research, indicating that the most effective targets for neuroprotective treatment change throughout the disease course.

To further disentangle the secondary GM atrophy caused by WM damage from primary neurodegenerative disease mechanisms, more studies investigating the spatiotemporal relationship between the 2 pathologic phenomena are needed, preferably with extensive follow-up time and a direct comparison with the different disease phenotypes.

Funding

This work was funded by the Dutch MS Research Foundation, Grant No. 14-876. FB is supported by the NIHR biomedical research centre at UCLH.

Declaration of interest

The authors declare that there is no conflict of interest.

Supplements

https://cdn-links.lww.com/permalink/wnl/b/wnl_2022_02_07_lie_1_sdc1.pdf
eAppendix 1 – Search strategy.
eAppendix 2 – Expanded results section.
eTable 1 – Characteristics of cross-sectional studies.
eTable 2 –Characteristics of longitudinal studies.
eTable 3 – Characteristics of cross-sectional and longitudinal studies.

References

1. Audoin B, Zaaraoui W, Reuter F, et al. Atrophy mainly affects the limbic system and the deep grey matter at the first stage of multiple sclerosis. J Neurol Neurosurg Psychiatry. 2010;81(6):690-695.
2. Popescu V, Klaver R, Voorn P, et al. What drives MRI-measured cortical atrophy in multiple sclerosis? Mult Scler. 2015;21(10):1280-1290.
3. Chard D, Trip SA. Resolving the clinico-radiological paradox in multiple sclerosis. F1000Res. 2017;6:1828.
4. DeLuca GC, Yates RL, Beale H, Morrow SA. Cognitive impairment in multiple sclerosis: clinical, radiologic and pathologic insights. Brain Pathol. 2015;25(1):79-98.
5. Chard DT, Griffin CM, Rashid W, et al. Progressive grey matter atrophy in clinically early relapsing-remitting multiple sclerosis. Mult Scler. 2004;10(4):387-391.
6. Calabrese M, Rinaldi F, Mattisi I, et al. The predictive value of gray matter atrophy in clinically isolated syndromes. Neurology. 2011;77(3):257-263.
7. Di Filippo M, Anderson VM, Altmann DR, et al. Brain atrophy and lesion load measures over 1 year relate to clinical status after 6 years in patients with clinically isolated syndromes. J Neurol Neurosurg Psychiatry. 2010;81(2):204-208.
8. Eshaghi A, Prados F, Brownlee WJ, et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. Ann Neurol. 2018;83(2):210-222.
9. Pellicano C, Kane RL, Gallo A, et al. Cognitive impairment and its relation to imaging measures in multiple sclerosis: a study using a computerized battery. J Neuroimaging. 2013;23(3):445-452.
10. Tekok-Kilic A, Benedict RH, Weinstock-Guttman B, et al. Independent contributions of cortical gray matter atrophy and ventricle enlargement for predicting neuropsychological impairment in multiple sclerosis. NeuroImage. 2007;36(4):1294-1300.
11. Rudick RA, Lee JC, Nakamura K, Fisher E. Gray matter atrophy correlates with MS disability progression measured with MSFC but not EDSS. J Neurol Sci. 2009;282(1-2):106-111.
12. Bergsland N, Horakova D, Dwyer MG, et al. Gray matter atrophy patterns in multiple sclerosis: a 10-year source-based morphometry study. Neuroimage Clin. 2018;17:444-451.Google Scholar
13. Eijlers AJC, van Geest Q, Dekker I, et al. Predicting cognitive decline in multiple sclerosis: a 5-year follow-up study. Brain. 2018;141(9):2605-2618.
14. Amato MP, Bartolozzi ML, Zipoli V, et al. Neocortical volume decrease in relapsing-remitting MS patients with mild cognitive impairment. Neurology. 2004;63(1):89-93.
15. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292-302.
16. Stangel M, Penner IK, Kallmann BA, Lukas C, Kieseier BC. Towards the implementation of "no evidence of disease activity" in multiple sclerosis treatment: the multiple sclerosis decision model. Ther Adv Neurol Disord. 2015;8(1):3-13.
17. Popescu BF, Lucchinetti CF. Pathology of demyelinating diseases. Annu Rev Pathol. 2012;7:185-217.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
19. Rayyan. Accessed September 10, 2019. rayyan.qcri.org
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan: a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210.
21. Geurts JJ, Barkhof F. Grey matter pathology in multiple sclerosis. Lancet Neurol. 2008;7(9):841-851.
22. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. Nat Rev Immunol. 2015;15(9):545-558.
23. Sherman SM. Functioning of circuits connecting thalamus and cortex. Compr Physiol. 2017;7(2):713-739.
24. Klaver R, De Vries HE, Schenk GJ, Geurts JJ. Grey matter damage in multiple sclerosis: a pathology perspective. Prion. 2013;7(1):66-75.
25. Kutzelnigg A, Lucchinetti CF, Stadelmann C, et al. Cortical demyelination and diffuse white matter injury in multiple sclerosis. Brain. 2005;128(pt 11):2705-2712.
26. Vercellino M, Plano F, Votta B, Mutani R, Giordana MT, Cavalla P. Grey matter pathology in multiple sclerosis. J Neuropathol Exp Neurol. 2005;64(12):1101-1107.

27. Wegner C, Esiri MM, Chance SA, Palace J, Matthews PM. Neocortical neuronal, synaptic, and glial loss in multiple sclerosis. Neurology. 2006;67(6):960-967.

28. Amiri H, de Sitter A, Bendfeldt K, et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin. 2018;19:466-475.

29. Stys PK, Zamponi GW, van Minnen J, Geurts JJ. Will the real multiple sclerosis please stand up?. Nat Rev Neurosci. 2012;13(7):507-514.

30. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology. 2014;83(3):278-286.

eReferences

e-1. Roosendaal S.D, Bendfeldt K, Vrenken H et al., Grey matter volume in a large cohort of MS patients: relation to MRI parameters and disability. Mult Scler, 2011. 17(9): p. 1098-106.

e-2. Bergsland N, Horakova D, Dwyer M.G et al., Subcortical and Cortical Gray Matter Atrophy in a Large Sample of Patients with Clinically Isolated Syndrome and Early Relapsing-Remitting Multiple Sclerosis. 2012. 33(8): p. 1573-1578.

e-3. Durhan G, Diker S, Has A.C et al., Assessment of the effect of cigarette smoking on regional brain volumes and lesion load in patients with clinically isolated syndrome. Int J Neurosci, 2016. 126(9): p. 805-811.

e-4. Henry R.G, Shieh M, Okuda D.T, Evangelista A, Gorno-Tempini M.L, Pelletier D, Regional grey matter atrophy in clinically isolated syndromes at presentation. J Neurol Neurosurg Psychiatry, 2008. 79(11): p. 1236-44.

e-5. Labiano-Fontcuberta A, Mato-Abad V, Alvarez-Linera J et al., Gray Matter Involvement in Radiologically Isolated Syndrome. Medicine, 2016. 95(13): p. e3208.

e-6. Dalton C.M, Chard D.T, Davies G.R et al., Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. Brain, 2004. 127(Pt 5): p. 1101-7.

e-7. Varosanec M, Uher T, Horakova D et al., Longitudinal Mixed-Effect Model Analysis of the Association between Global and Tissue-Specific Brain Atrophy and Lesion Accumulation in Patients with Clinically Isolated Syndrome. AJNR Am J Neuroradiol, 2015. 36(8): p. 1457-64.

e-8. Jenkins T.M, Ciccarelli O, Atzori M et al., Early pericalcarine atrophy in acute optic neuritis is associated with conversion to multiple sclerosis. J Neurol Neurosurg Psychiatry, 2011. 82(9): p. 1017-21.

e-9. Jehna M, Pirpamer L, Khalil M et al., Periventricular lesions correlate with cortical thinning in multiple sclerosis. Ann Neurol, 2015. 78(4): p. 530-9.

e-10. Tavazzi E, Bergsland N, Kuhle J et al., A multimodal approach to assess the validity of atrophied T2-lesion volume as an MRI marker of disease progression in multiple sclerosis. J Neurol, 2020. 267(3): p. 802-811.

e-11. Cacciaguerra L, Pagani E, Mesaros S et al., Dynamic volumetric changes of hippocampal subfields in clinically isolated syndrome patients: A 2-year MRI study. Mult Scler, 2019. 25(9): p. 1232-1242.

e-12. Henry R.G, Shieh M, Amirbekian B, Chung S, Okuda D.T, Pelletier D, Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. J Neurol Sci, 2009. 282(1-2): p. 61-6.

e-13. Chard D.T, Griffin C.M, Parker G.J, Kapoor R, Thompson A.J, Miller D.H, Brain atrophy in clinically early relapsing-remitting multiple sclerosis. Brain, 2002. 125(Pt 2): p. 327-37.

e-14. Tiberio M, Chard D.T, Altman D.R et al., Gray and white matter volume changes in early RRMS: a 2-year longitudinal study. Neurology, 2005. 64(6): p. 1001-7.

e-15. Dolezal O, Dwyer M.G, Horakova D et al., Detection of cortical lesions is dependent on choice of slice thickness in patients with multiple sclerosis. Int Rev Neurobiol, 2007. 79: p. 475-89.

e-16. Grothe M, Lotze M, Langner S, Dressel A, The role of global and regional gray matter volume decrease in multiple sclerosis. J Neurol, 2016. 263(6): p. 1137-45.

e-17. Steenwijk M.D, Daams M, Pouwels P.J et al., What explains gray matter atrophy in long-standing multiple sclerosis? Radiology, 2014. 272(3): p. 832-42.

e-18. Toth E, Szabo N, Csete G et al., Gray Matter Atrophy Is Primarily Related to Demyelination of Lesions in Multiple Sclerosis: A Diffusion Tensor Imaging MRI Study. Front Neuroanat, 2017. 11: p. 23.

e-19. Prinster A, Quarantelli M, Lanzillo R et al., A voxel-based morphometry study of disease severity correlates in relapsing-- remitting multiple sclerosis. Mult Scler, 2010. 16(1): p. 45-54.

e-20. Quarantelli M, Ciarmiello A, Morra V.B et al., Brain tissue volume changes in relapsing-remitting multiple sclerosis: correlation with lesion load. Neuroimage, 2003. 18(2): p. 360-6.

e-21. Ceccarelli A, Rocca M.A, Falini A et al., Normal-appearing white and grey matter damage in MS. A volumetric and diffusion tensor MRI study at 3.0 Tesla. J Neurol, 2007. 254(4): p. 513-8.

e-22. Sbardella E, Petsas N, Tona F et al., Assessing the correlation between grey and white matter damage with motor and cognitive impairment in multiple sclerosis patients. PLoS One, 2013. 8(5): p. e63250.

e-23. Antulov R, Carone D.A, Bruce J et al., Regionally distinct white matter lesions do not contribute to regional gray matter atrophy in patients with multiple sclerosis. J Neuroimaging, 2011. 21(3): p. 210-8.

e-24. Fisher E, Lee J.C, Nakamura K, Rudick R.A, Gray matter atrophy in multiple sclerosis: A longitudinal study. 2008. 64(3): p. 255-265.

e-25. Vidal-Jordana A, Sastre-Garriga J, Perez-Miralles F et al., Early brain pseudoatrophy while on natalizumab therapy is due to white matter volume changes. Mult Scler, 2013. 19(9): p. 1175-81.

e-26. Vidal-Jordana A, Sastre-Garriga J, Perez-Miralles F et al., Brain Volume Loss During the First Year of Interferon-Beta Treatment in Multiple Sclerosis: Baseline Inflammation and Regional Brain Volume Dynamics. J Neuroimaging, 2016. 26(5): p. 532-8.

e-27. Bendfeldt K, Kuster P, Traud S et al., Association of regional gray matter volume loss and progression of white matter lesions in multiple sclerosis - A longitudinal voxel-based morphometry study. Neuroimage, 2009. 45(1): p. 60-7.

e-28. Preziosa P, Rocca M.A, Riccitelli G.C et al., Effects of Natalizumab and Fingolimod on Clinical, Cognitive, and Magnetic Resonance Imaging Measures in Multiple Sclerosis. Neurotherapeutics, 2019.

e-29. Masuda H, Mori M, Hirano S et al., Relapse numbers and earlier intervention by disease modifying drugs are related with progression of less brain atrophy in patients with multiple sclerosis. J Neurol Sci, 2019. 403: p. 78-84.

e-30. Al-Radaideh A, Athamneh I, Alabadi H, Hbahbih M, Cortical and Subcortical Morphometric and Iron Changes in Relapsing-Remitting Multiple Sclerosis and Their Association with White Matter T2 Lesion Load : A 3-Tesla Magnetic Resonance Imaging Study. Clin Neuroradiol, 2019. 29(1): p. 51-64.

e-31. De Stefano N, Matthews P.M, Filippi M et al., Evidence of early cortical atrophy in MS: relevance to white matter changes and disability. Neurology, 2003. 60(7): p. 1157-62.

e-32. Zivadinov R, Tekwe C, Bergsland N et al., Bimonthly Evolution of Cortical Atrophy in Early Relapsing-Remitting Multiple Sclerosis over 2 Years: A Longitudinal Study. Mult Scler Int, 2013. 2013: p. 231345.

e-33. Calabrese M, Rinaldi F, Mattisi I et al., Widespread cortical thinning characterizes patients with MS with mild cognitive impairment. Neurology, 2010. 74(4): p. 321-8.

e-34. Steenwijk M.D, Geurts J.J, Daams M et al., Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain, 2016. 139(Pt 1): p. 115-26.

e-35. Charil A, Dagher A, Lerch J.P, Zijdenbos A.P, Worsley K.J, Evans A.C, Focal cortical atrophy in multiple sclerosis: relation to lesion load and disability. Neuroimage, 2007. 34(2): p. 509-17.

e-36. Datta S, Staewen T.D, Cofield S.S et al., Regional gray matter atrophy in relapsing remitting multiple sclerosis: baseline analysis of multi-center data. Mult Scler Relat Disord, 2015. 4(2): p. 124-36.

e-37. Tao G, Datta S, He R, Nelson F, Wolinsky J.S, Narayana P.A, Deep gray matter atrophy in multiple sclerosis: a tensor based morphometry. J Neurol Sci, 2009. 282(1-2): p. 39-46.

e-38. Hasan K.M, Walimuni I.S, Abid H, Datta S, Wolinsky J.S, Narayana P.A, Human brain atlas-based multimodal MRI analysis of volumetry, diffusimetry, relaxometry and lesion distribution in multiple sclerosis patients and healthy adult controls: implications for understanding the pathogenesis of multiple sclerosis and consolidation of quantitative MRI results in MS. J Neurol Sci, 2012. 313(1-2): p. 99-109.

e-39. Narayana P.A, Govindarajan K.A, Goel P et al., Regional cortical thickness in relapsing remitting multiple sclerosis: A multi-center study. Neuroimage Clin, 2012. 2: p. 120-31.

e-40. Ceccarelli A, Rocca M.A, Pagani E et al., A voxel-based morphometry study of grey matter loss in MS patients with different clinical phenotypes. Neuroimage, 2008. 42(1): p. 315-22.

e-41. Duan Y, Liu Y, Liang P et al., Comparison of grey matter atrophy between patients with neuromyelitis optica and multiple sclerosis: a voxel-based morphometry study. Eur J Radiol, 2012. 81(2): p. e110-4.

e-42. Prinster A, Quarantelli M, Orefice G et al., Grey matter loss in relapsing-remitting multiple sclerosis: a voxel-based morphometry study. Neuroimage, 2006. 29(3): p. 859-67.

e-43. Kuceyeski A.F, Vargas W, Dayan M et al., Modeling the relationship among gray matter atrophy, abnormalities in connecting white matter, and cognitive performance in early multiple sclerosis. AJNR Am J Neuroradiol, 2015. 36(4): p. 702-9.

e-44. Riccitelli G, Rocca M.A, Pagani E et al., Mapping regional grey and white matter atrophy in relapsing-remitting multiple sclerosis. Mult Scler, 2012. 18(7): p. 1027-37.

e-45. Riccitelli G, Rocca M.A, Pagani E et al., Cognitive impairment in multiple sclerosis is associated to different patterns of gray matter atrophy according to clinical phenotype. Hum Brain Mapp, 2011. 32(10): p. 1535-43.

e-46. Steenwijk M.D, Daams M, Pouwels P.J et al., Unraveling the relationship between regional gray matter atrophy and pathology in connected white matter tracts in long-standing multiple sclerosis. Hum Brain Mapp, 2015. 36(5): p. 1796-807.

e-47. Bergsland N, Lagana M.M, Tavazzi E et al., Corticospinal tract integrity is related to primary motor cortex thinning in relapsing-remitting multiple sclerosis. Mult Scler, 2015. 21(14): p. 1771-80.

e-48. Damasceno A, Damasceno B.P, Cendes F, No evidence of disease activity in multiple sclerosis: Implications on cognition and brain atrophy. Mult Scler, 2016. 22(1): p. 64-72.

e-49. Gaetano L, Haring D.A, Radue E.W et al., Fingolimod effect on gray matter, thalamus, and white matter in patients with multiple sclerosis. Neurology, 2018. 90(15): p. e1324-e1332.

e-50. Battaglini M, Giorgio A, Stromillo M.L et al., Voxel-wise assessment of progression of regional brain atrophy in relapsing-remitting multiple sclerosis. J Neurol Sci, 2009. 282(1-2): p. 55-60.

e-51. Bendfeldt K, Blumhagen J.O, Egger H et al., Spatiotemporal distribution pattern of white matter lesion volumes and their association with regional grey matter volume reductions in relapsing-remitting multiple sclerosis. Hum Brain Mapp, 2010. 31(10): p. 1542-55.

e-52. Sicotte N.L, Kern K.C, Giesser B.S et al., Regional hippocampal atrophy in multiple sclerosis. Brain, 2008. 131(Pt 4): p. 1134-41.

e-53. Kalinin I, Makshakov G, Evdoshenko E, The Impact of Intracortical Lesions on Volumes of Subcortical Structures in Multiple Sclerosis. American Journal of Neuroradiology, 2020.

e-54. Magon S, Chakravarty M.M, Amann M et al., Label-fusion-segmentation and deformation-based shape analysis of deep gray matter in multiple sclerosis: the impact of thalamic subnuclei on disability. Hum Brain Mapp, 2014. 35(8): p. 4193-203.

e-55. Pontillo G, Cocozza S, Lanzillo R et al., Determinants of Deep Gray Matter Atrophy in Multiple Sclerosis: A Multimodal MRI Study. AJNR Am J Neuroradiol, 2019. 40(1): p. 99-106.

e-56. Hasan K.M, Halphen C, Kamali A, Nelson F.M, Wolinsky J.S, Narayana P.A, Caudate nuclei volume, diffusion tensor metrics, and T(2) relaxation in healthy adults and relapsing-remitting multiple sclerosis patients: implications for understanding gray matter degeneration. J Magn Reson Imaging, 2009. 29(1): p. 70-7.

e-57. Talmage G.D, Coppes O.J.M, Javed A, Bernard J, Natalizumab stabilizes physical, cognitive, MRI, and OCT markers of disease activity: A prospective, non-randomized pilot study. PLoS One, 2017. 12(4): p. e0173299.

e-58. Furby J, Hayton T, Altmann D et al., Different white matter lesion characteristics correlate with distinct grey matter abnormalities on magnetic resonance imaging in secondary progressive multiple sclerosis. Mult Scler, 2009. 15(6): p. 687-94.

e-59. Sastre-Garriga J, Ingle G.T, Chard D.T, Ramio-Torrenta L, Miller D.H, Thompson A.J, Grey and white matter atrophy in early clinical stages of primary progressive multiple sclerosis. Neuroimage, 2004. 22(1): p. 353-9.

e-60. Sastre-Garriga J, Ingle G.T, Chard D.T et al., Grey and white matter volume changes in early primary progressive multiple sclerosis: a longitudinal study. Brain, 2005. 128(Pt 6): p. 1454-60.

e-61. Galego O, Gouveia A, Batista S, Moura C, Machado E, Brain atrophy and physical disability in primary progressive multiple sclerosis: A volumetric study. Neuroradiol J, 2015. 28(3): p. 354-8.

e-62. Sepulcre J, Sastre-Garriga J, Cercignani M, Ingle G.T, Miller D.H, Thompson A.J, Regional gray matter atrophy in early primary progressive multiple sclerosis: a voxel-based morphometry study. Arch Neurol, 2006. 63(8): p. 1175-80.

e-63. Mesaros S, Rocca M.A, Pagani E et al., Thalamic damage predicts the evolution of primary-progressive multiple sclerosis at 5 years. AJNR Am J Neuroradiol, 2011. 32(6): p. 1016-20.

e-64. Ciampi E, Pareto D, Sastre-Garriga J et al., Grey matter atrophy is associated with disability increase in natalizumab-treated patients. Mult Scler, 2017. 23(4): p. 556-566.

e-65. Fisniku L.K, Chard D.T, Jackson J.S et al., Gray matter atrophy is related to long-term disability in multiple sclerosis. Ann Neurol, 2008. 64(3): p. 247-54.

e-66. Fragoso Y.D, Wille P.R, Abreu M et al., Correlation of clinical findings and brain volume data in multiple sclerosis. J Clin Neurosci, 2017. 44: p. 155-157.

e-67. Sanfilipo M.P, Benedict R.H, Sharma J, Weinstock-Guttman B, Bakshi R, The relationship between whole brain volume and disability in multiple sclerosis: a comparison of normalized gray vs. white matter with misclassification correction. Neuroimage, 2005. 26(4): p. 1068-77.

e-68. Tedeschi G, Lavorgna L, Russo P et al., Brain atrophy and lesion load in a large population of patients with multiple sclerosis. Neurology, 2005. 65(2): p. 280-5.

e-69. Zimmermann H, Rolfsnes H.O, Montag S et al., Putaminal alteration in multiple sclerosis patients with spinal cord lesions. J Neural Transm, 2015. 122(10): p. 1465-73.

e-70. Calabrese M, Atzori M, Bernardi V et al., Cortical atrophy is relevant in multiple sclerosis at clinical onset. J Neurol, 2007. 254(9): p. 1212-20.

e-71. Eshaghi A, Marinescu R.V, Young A.L et al., Progression of regional grey matter atrophy in multiple sclerosis. Brain, 2018. 141(6): p. 1665-1677.

e-72. Hier D.B, Wang J, Reduced cortical surface area in multiple sclerosis. Neurol Res, 2007. 29(3): p. 231-2.

e-73. Muhlau M, Buck D, Forschler A et al., White-matter lesions drive deep gray-matter atrophy in early multiple sclerosis: support from structural MRI. Mult Scler, 2013. 19(11): p. 1485-92.

e-74. Pareto D, Sastre-Garriga J, Auger C et al., Juxtacortical Lesions and Cortical Thinning in Multiple Sclerosis. AJNR Am J Neuroradiol, 2015. 36(12): p. 2270-6.

e-75. Sailer M, Fischl B, Salat D et al., Focal thinning of the cerebral cortex in multiple sclerosis. Brain, 2003. 126(Pt 8): p. 1734-44.

e-76. Shiee N, Bazin P.L, Zackowski K.M et al., Revisiting brain atrophy and its relationship to disability in multiple sclerosis. PLoS One, 2012. 7(5): p. e37049.

e-77. Tsagkas C, Chakravarty M.M, Gaetano L et al., Longitudinal patterns of cortical thinning in multiple sclerosis. Human Brain Mapping, 2020. 41(8): p. 2198-2215.

e-78. Sepulcre J, Goni J, Masdeu J.C et al., Contribution of white matter lesions to gray matter atrophy in multiple sclerosis: evidence from voxel-based analysis of T1 lesions in the visual pathway. Arch Neurol, 2009. 66(2): p. 173-9.

e-79. Bermel R.A, Innus M.D, Tjoa C.W, Bakshi R, Selective caudate atrophy in multiple sclerosis: a 3D MRI parcellation study. Neuroreport, 2003. 14(3): p. 335-9.

e-80. Deppe M, Kramer J, Tenberge J.G et al., Early silent microstructural degeneration and atrophy of the thalamocortical network in multiple sclerosis. Hum Brain Mapp, 2016. 37(5): p. 1866-79.

e-81. Hasan K.M, Walimuni I.S, Abid H et al., Multimodal quantitative magnetic resonance imaging of thalamic development and aging across the human lifespan: implications to neurodegeneration in multiple sclerosis. J Neurosci, 2011. 31(46): p. 16826-32.

e-82. Longoni G, Rocca M.A, Pagani E et al., Deficits in memory and visuospatial learning correlate with regional hippocampal atrophy in MS. Brain Struct Funct, 2015. 220(1): p. 435-44.

e-83. Louapre C, Govindarajan S.T, Gianni C et al., Heterogeneous pathological processes account for thalamic degeneration in multiple sclerosis: Insights from 7 T imaging. Mult Scler, 2018. 24(11): p. 1433-1444.

e-84. Mehndiratta A, Treaba C.A, Barletta V et al., Characterization of thalamic lesions and their correlates in multiple sclerosis by ultra-high-field MRI. Mult Scler, 2020: p. 1352458520932804.

e-85. Rocca M.A, Mesaros S, Pagani E, Sormani M.P, Comi G, Filippi M, Thalamic damage and long-term progression of disability in multiple sclerosis. Radiology, 2010. 257(2): p. 463-9.

e-86. Fuchs T.A, Carolus K, Benedict R.H.B et al., Impact of Focal White Matter Damage on Localized Subcortical Gray Matter Atrophy in Multiple Sclerosis: A 5-Year Study. AJNR Am J Neuroradiol, 2018. 39(8): p. 1480-1486.

e-87. Pongratz V, Schmidt P, Bussas M et al., Prognostic value of white matter lesion shrinking in early multiple sclerosis: An intuitive or naive notion? Brain Behav, 2019. 9(12): p. e01417.

e-88. Lee H, Nakamura K, Narayanan S et al., Impact of immunoablation and autologous hematopoietic stem cell transplantation on gray and white matter atrophy in multiple sclerosis. Mult Scler, 2018. 24(8): p. 1055-1066.

e-89. Tedeschi G, Dinacci D, Comerci M et al., Brain atrophy evolution and lesion load accrual in multiple sclerosis: a 2-year follow-up study. Mult Scler, 2009. 15(2): p. 204-11.

e-90. Fox J, Kraemer M, Schormann T et al., Individual Assessment of Brain Tissue Changes in MS and the Effect of Focal Lesions on Short-Term Focal Atrophy Development in MS: A Voxel-Guided Morphometry Study. Int J Mol Sci, 2016. 17(4): p. 489.

CHAPTER 2

Challenges related to lesions in MS

CHAPTER 2.1

Comparing lesion segmentation methods in multiple sclerosis: input from one manually delineated subject is sufficient for accurate lesion segmentation

Merlin M. Weeda
Iman Brouwer
Marlieke L. de Vos
Myrte S. de Vries
Frederik Barkhof
Petra J.W. Pouwels
Hugo Vrenken

NeuroImage: Clinical 24 (2019)

Abstract

Purpose

Accurate lesion segmentation is important for measurements of lesion load and atrophy in subjects with multiple sclerosis (MS). International MS lesion challenges show a preference of convolutional neural networks (CNN) strategies, such as nicMSlesions. However, since the software is trained on fairly homogenous training data, we aimed to test the performance of nicMSlesions in an independent dataset with manual and other automatic lesion segmentations to determine whether this method is suitable for larger, multi-center studies.

Methods

Manual lesion segmentation was performed in fourteen subjects with MS on sagittal 3D FLAIR images from a 3T GE whole-body scanner with 8-channel head coil. We compared five different categories of automated lesion segmentation methods for their volumetric and spatial agreement with manual segmentation: (i) unsupervised, untrained (LesionTOADS); (ii) supervised, untrained (LST-LPA and nicMSlesions with default settings); (iii) supervised, untrained with threshold adjustment (LST-LPA optimized for current data); (iv) supervised, trained with leave-one-out cross-validation on fourteen subjects with MS (nicMSlesions and BIANCA); and (v) supervised, trained on a single subject with MS (nicMSlesions). Volumetric accuracy was determined by the intra-class correlation coefficient (ICC) and spatial accuracy by Dice's similarity index (SI). Volumes and SI were compared between methods using repeated measures ANOVA or Friedman tests with post-hoc pairwise comparison.

Results

The best volumetric and spatial agreement with manual was obtained with the supervised and trained methods nicMSlesions and BIANCA (ICC absolute agreement >0.968 and median SI>0.643) and the worst with the unsupervised, untrained method LesionTOADS (ICC absolute agreement=0.140 and median SI=0.444). Agreement with manual in the single-subject network training of nicMSlesions was poor for input with low lesion volumes (i.e. two subjects with lesion volumes ≤3.0 ml). For the other twelve subjects, ICC varied from 0.593 to 0.973 and median SI varied from 0.535 to 0.606. In all cases, the single-subject trained nicMSlesions segmentations outperformed LesionTOADS, and in almost all cases it also outperformed LST-LPA.

Conclusion

Input from only one subject to re-train the deep learning CNN nicMSlesions is sufficient for adequate lesion segmentation, with on average higher volumetric and spatial agreement with manual than obtained with the untrained methods LesionTOADS and LST-LPA.

Introduction

Multiple sclerosis (MS) is an autoimmune disorder of the central nervous system, characterized by neurodegeneration and demyelination. To enable both atrophy and lesion load measurements in subjects with MS, accurate lesion segmentation is necessary. However, manual lesion segmentation is labor intensive and highly time consuming. Therefore, several automatic lesion segmentation methods have been developed[1,2] with varying amounts of manual input and output optimization possibilities. In general, methods can be supervised and/or trained, i.e. based on a previous set of training images that comes with the algorithm (supervised) or based on a set of training images specific to the dataset in which the method is to be applied (trained).

International MS lesion challenges have shown that especially convolutional neural networks (CNN) deep learning strategies perform well for MS lesion segmentation[3]. However, CNN methods have the disadvantage that they still require a lot of manual reference input data in order to construct the network in the MR domain of choice. Recently, a method was published showing that input from only one single subject could be sufficient for the cascaded CNN method nicMSlesions to outperform manual delineation[4]. The nicMSlesions software consists of an 11 layer CNN source model trained using the public MS databases of the Medical Image Computing and Computer Assisted Intervention (MICCAI) society[5]. Because that original training data was fairly homogeneous, with most data acquired in 3T Siemens systems, we aimed to test the performance of nicMSlesions in an independent dataset with manual lesion segmentations and to compare it with other segmentation methods. In this way, we will be able to determine whether this deep learning method is a reliable lesion segmentation method when compared to other existing lesion segmentation methods, with special attention for the use of minimal input data (i.e. few manual delineations available) for its applicability in real-world data.

In this study, we will focus on four different segmentation methods, with different configurations splitting them in (i) unsupervised, untrained methods (Lesion-Topology preserving Anatomical Segmentation (LesionTOADS)[6]); (ii) supervised, untrained methods (Lesion Segmentation Toolbox with Lesion Probability Algorithm (LST-LPA)[7] and the default network of nicMSlesions[4]); (iii) supervised, untrained methods with threshold adjustment (LST-LPA); (iv) supervised, trained methods (FMRIB Software Library (FSL) Brain Intensity AbNormality Classification Algorithm (BIANCA)[8] and the trained network of nicMSlesions); and (v) supervised, trained method with minimal input (single-subject trained network of nicMSlesions). We quantified volumetric and spatial agreement with manual for all methods and also tested their performance in lesion-negative, healthy control images, in order to determine which type of method is this most suitable (i.e. best performance with least manual labour) for larger, multi-center studies.

Methods

Subjects

From a larger cohort of subjects with RRMS, a total of fourteen subjects scanned between December 2016 and June 2017 were included in this study. Subjects included were over 18

years of age, diagnosed with RRMS for a maximum of five years with a maximum expanded disease disability status scale (EDSS) score of 5.0, and received either first line disease modifying therapy or no therapy at all. Patients were not eligible for participation if they had switched medication in the 6 months prior to their visit, if they had received second-line treatment in the past, or if they had received steroid treatment in the 3 months prior to the MRI examination. Further exclusion criteria were past or current neurological or immunological syndromes other than MS, and inability to undergo MRI examination. From the same cohort, data of five healthy controls was also included in the present study.
This study was approved by the local institutional medical ethics committee and written informed consent was obtained from all individuals, according to the Declaration of Helsinki.

MRI examination

Subjects underwent extensive MRI examination on a 3T whole-body MR scanner (GE Discovery MR750) with an 8-channel phased-array head coil. The protocol included a sagittal 3D T1-weighted fast spoiled gradient echo sequence (FSPGR with TR/TE/TI = 8.2/3.2/450 ms and voxel size 1.0x1.0x1.0 mm) and a sagittal 3D T2-weighted fluid attenuated inversion recovery sequence (FLAIR with TR/TE/TI = 8000/130/2338 ms with voxel size 1.0x1.0x1.2 mm).

MR imaging data analysis

For this research, we investigated the performance of four different lesion segmentation methods in comparison to manual segmentation, all with different levels of optimization and training possibilities, allowing us to study the following five categories of lesion segmentation methods: (i) unsupervised, untrained segmentation, without optimization; (ii) supervised, untrained segmentation, without optimization; (iii) supervised, untrained segmentation, with threshold adjustment; (iv) supervised, fully trained segmentation, with optimization; and (v) supervised, trained segmentation with minimal input. Details on the different lesion segmentation methods are described below and an overview of these methods is shown in **Table 1** and **Figure 1**.

Table 1 – Overview of the different lesion segmentation methods investigated in this study for their performance to manual segmentation.

Supervision	Training	Optimization	Method
no	no	no	(1) LesionTOADS
yes	no	no	(2) LST-LPA default (3) nicMSlesions default
yes	no	yes	(4) LST-LPA adjusted-threshold
yes	yes (*n*=14)	yes	(5) nicMSlesions optimized (6) BIANCA
yes	yes (*n*=1)	no	(7) nicMSlesions single-subject

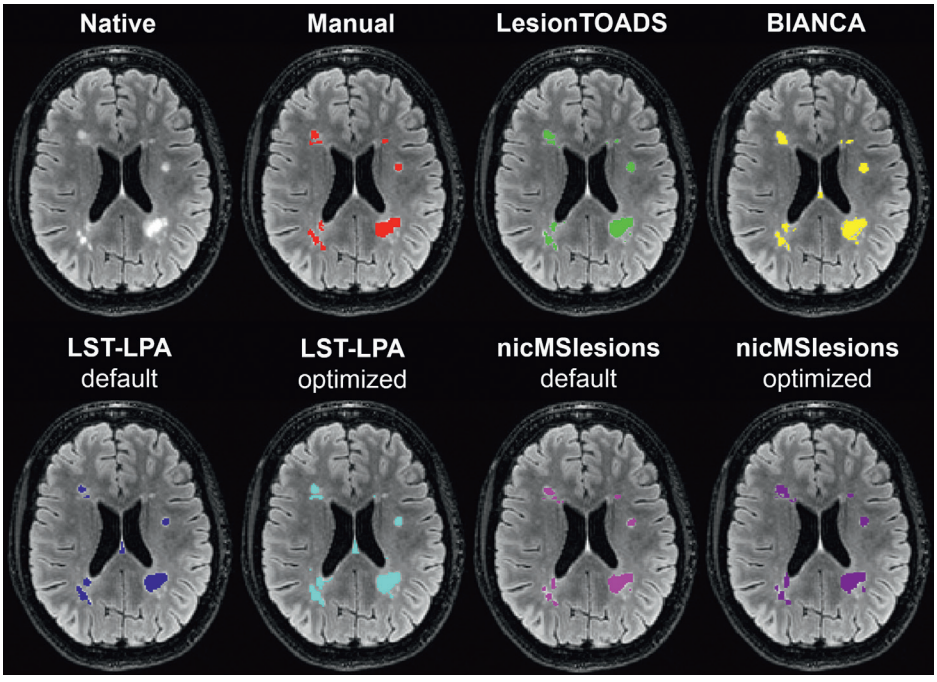

Figure 1 – Example lesion segmentation shown for the different lesion segmentation methods on the original 3D FLAIR image of a 36 year old female with RRMS: manual (red); LesionTOADS (green), BIANCA (yellow), LST-LPA default (blue), LST-LPA adjusted-threshold (turquoise), nicMSlesions default (pink), and nicMSlesions optimized (purple).

Manual segmentation

An expert rater (MLV; experience >10 years) manually delineated the lesions on the 3D FLAIR images. For this, lesions were defined as hyper-intense regions compared to the surrounding tissue with a size of at least three voxels. The rater had access to the 3D T1 images for reference. Three subjects were rated twice to calculate intra-rater agreement.

Unsupervised, untrained segmentation without optimization: LesionTOADS

Lesion-Topology preserving Anatomical Segmentation (LesionTOADS)[6] uses a statistical lesion atlas based on a topology preserving anatomical atlas. Preprocessing consisted of bias field correction and brain extraction from both T1 and FLAIR images, as well as affine linear registration of the brain extracted images from T1 to FLAIR. The method results in a binary lesion segmentation map.

Supervised, untrained segmentation without optimization: LST-LPA and nicMSlesions default

LST-LPA default

Lesion Segmentation Toolbox with Lesion Prediction Algorithm (LST-LPA)[7] uses voxel-wise binary regression with spatially varying intercepts. No preprocessing is applied, because

the algorithm performs the necessary bias field correction and affine registration of T1 to FLAIR images as part of the pipeline. Because the output is a probabilistic map, a threshold has to be defined to obtain binary segmentation files, with the default probability threshold set to 0.5.

nicMSlesions default

nicMSlesions is a deep learning method based on cascaded convolutional neural networks that, in contrast to most trained or deep learning methods, can be used when limited amounts of manual input data are available. For the supervised, untrained segmentation, we segmented our data using the default parameters of the network (called *baseline 2ch*) from nicMSlesions with the outcome probability map thresholded at 0.5. No preprocessing was required.

Supervised, untrained segmentation with threshold adjustment: LST-LPA adjusted-threshold

Threshold adjustment of LST-LPA output was performed by varying the probability threshold from 0.10 to 0.80 with step-size 0.05 and selecting the threshold that resulted in the highest mean similarity index (SI) to manual across all fourteen subjects. Note that this configuration of LST-LPA is dependent on the manual delineations, since they are required for the probability threshold optimization.

Supervised, fully trained segmentation, with optimization: nicMSlesions optimized and BIANCA

nicMSlesions optimized

Next to the default *baseline 2ch* trained segmentation of nicMSlesions, the method allows full re-training of the neural network (all 11 layers) with our own dataset, evaluating performance by using a leave-one-out cross-validation approach (i.e. train on thirteen subjects and apply on the fourteenth subject). No preprocessing was required. From this output, we determined the optimal probability threshold in the same way as for LST-LPA. Although the minimal lesion size can be optimized for nicMSlesions as well, we chose to keep this at the default value of five voxels, since optimization of this parameter only has very limited effect on the performance of the segmentation[5].

BIANCA

BIANCA[8] is a trained segmentation method based on the *k*-nearest neighbor algorithm. Preprocessing of the data consisted of bias field correction and brain extraction from both T1 and FLAIR images, as well as affine registration of the brain extracted images from T1 to FLAIR. Since BIANCA allows for many parameters to be set, we first optimized BIANCA for our full (n=14) dataset with leave-one-out cross-validation. First, we optimized the optimal probability and cluster-size thresholds. Then, we optimized: 1) the number of lesion and non-lesion training points; 2) the location of the non-lesion training points; 3) use of patch and patch size; and 4) spatial weighting.

Supervised, trained segmentation with minimal input: nicMSlesions single-subject

Last, we tested a single-subject configuration of nicMSlesions, in which the last 1 or 2 layers of the 11-layer neural network were re-trained from single-subject input only, again using leave-one-out cross-validation. In our data, nicMSlesions re-trains one layer for input lesion volumes below 5 ml, and two layers for higher input lesion volume. Here, the default probability threshold of 0.5 was used as well.

Statistics

Per segmentation method, true and false positives and negatives were extracted from the native FLAIR images, and corresponding sensitivity and 1-specificity were calculated. Furthermore, two-way random intra-class correlation coefficient (ICC) for absolute agreement and for consistency were calculated to determine volumetric accuracy. Dice's similarity index (SI) compared to manual was calculated to quantify spatial accuracy.

Statistical analysis was performed in IBM SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, N.Y., USA). Repeated measures ANOVA was used for volumetric agreement with manual per segmentation method and Friedman Tests for spatial agreement with manual per segmentation method. For the single-subject analyses of nicMSlesions, Wilcoxon Signed Ranks tests were used.

For the repeated measures ANOVA, Mauchly's test of sphericity was performed to assess equal variances of the differences between all within-subject factors. When the assumption of sphericity was violated, degrees of freedom were corrected using Huyn-Feldt estimates of sphericity. When appropriate, post-hoc analyses were conducted using Mann-Whitney U tests (unpaired) or Wilcoxon Signed Ranks tests (paired).

Inter quartile range was determined by the 25th and 75th percentile. Results were considered statistically significant upon p-value < 0.05.

Results

Demographics

Five of the fourteen subjects with RRMS were male (36%). Mean age was 37.1 ± 5.3 years (range 26.4 to 47.7) and mean disease duration 3.1 ± 1.4 years (range 0.6 to 4.7 years). Disease modifying treatment was used by ten subjects (dimethyl fumarate n=3; glatiramer acetate n=2; interferon-β n=3; teriflunomide n=2) and median EDSS was 3.5 (range 1.0 – 4.0).

From the healthy controls, one of five subjects was male (20%) and mean age was 34.3 ± 8.3 years (range 22.9 to 45.7).

Optimization of lesion segmentation methods

Optimization was performed on LST-LPA, nicMSlesions and BIANCA based on the highest SI to manual segmentation volume. Results are depicted in **Supplemental Table 1** showing an optimal probability threshold of 0.25 for LST-LPA and 0.40 for nicMSlesions. Optimization of BIANCA (**Supplemental Table 2**) in our dataset resulted in the following settings: probability threshold 0.99; cluster size threshold 3; 2000 lesion points and equal number of non-lesion

points in the training set; any location of the non-lesion training points; 3D patch with size 5; and spatial weighting 2.

Performance of lesion segmentation methods: untrained and trained on n=14 subjects with MS

Volumetric and spatial reliability of the manual segmentations (i.e. intra-rater agreement) were good: for the three images that were segmented twice, ICC for absolute agreement was 0.867 and mean SI was 0.76 ± 0.04.

The median lesion volume according to manual delineation was 7.91 ml (interquartile range [IQR]: 4.26 - 10.15 ml). The performance of the different segmentation methods in subjects with RRMS is shown in **Table 2**. Note that the single-subject trained nicMSlesions is added to the table as an average over all fourteen configurations, but results from these variants separately are described in section 3.4. A scatter-plot showing the lesion segmentation volumes compared to manual is depicted in **Figure 2**.

The two supervised and trained methods nicMSlesions optimized and BIANCA showed the highest volumetric agreement (ICC absolute agreement = 0.975 and 0.968, respectively), as well as the highest sensitivity to lesions (0.698 and 0.639, respectively) and the best spatial agreement with manual (SI = 0.660 and 0.643, respectively). From the two supervised, untrained methods (i.e. LST-LPA default and adjusted-threshold, and nicMSlesions default), LST-LPA showed the best volumetric and spatial agreement, as well as best sensitivity to lesions, both for the default and for the probability threshold optimized configuration. The unsupervised, untrained method LesionTOADS performed worst on all volumetric and spatial measures, except for lesion specificity, which was higher than for the other methods.

Statistical analysis showed a significant effect of segmentation method on lesion volumes ($F(6,78)=35.435$, $p < 0.001$), with post-hoc testing showing this difference between manual and the methods LesionTOADS ($p=0.001$), LST-LPA default ($p=0.001$), nicMSlesions default ($p=0.022$), but not for LST-LPA adjusted-threshold ($p = 0.778$), nicMSlesions optimized ($p=0.300$) or BIANCA ($p=0.925$). Friedman test for SI showed a main effect of method as well ($\chi^2(5) = 56.082$, $p<0.001$), with post-hoc testing showing that these differences were significant between all combinations of methods except between: (a) LesionTOADS and nicMSlesions default; and (b) LST-LPA default and nicMSlesions default.

Table 2 – Volumetric and spatial accuracy of the various grouped lesion segmentation methods compared to manual in the fourteen subjects with RRMS.

RRMS *n*=14	**Lesion volume**	**ICC absolute agreement**	**ICC consistency**	**True positive volume**	**True negative volume**	**False positive volume**	**False negative volume**	**Sensitivity**	**1–Specificity (10E-3)**	**SI to manual**
Manual	7.91 (4.26 – 10.15)	NA	NA	NA	NA	NA	NA	NA	NA	
LesionToads	3.67 *** (2.57 – 4.23)	0.140 (-0.119 – 0.503)	0.300 (-0.253 – 0.705)	2.78 (1.18 – 3.28)	11092 (11090 – 11095)	0.72 (0.48 – 1.11)	5.07 (3.20 – 6.20)	0.312 (0.230 – 0.405)	0.065 (0.043 – 0.100)	0.444 (0.336 – 0.542)
LST-LPA default	4.40 *** (1.86 – 7.19)	0.73 (-0.059 – 0.939)	0.947 (0.845 – 0.983)	3.21 (1.48 – 5.81)	11091 (11089 – 11096)	1.06 (0.44 – 2.20)	4.02 (2.94 – 5.80)	0.414 (0.298 – 0.520)	0.096 (0.039 – 0.198)	0.528 (0.425 – 0.581)
LST-LPA adjusted-threshold	7.64 (3.82 – 11.15)	0.917 (0.769 – 0.972)	0.917 (0.764 – 0.973)	4.46 (2.28 – 7.16)	11089 (11086 – 11095)	2.95 (1.62 – 5.48)	2.72 (2.07 – 4.40)	0.592 (0.431 – 0.637)	0.266 (0.145 – 0.494)	0.568 (0.481 – 0.607)
nicMSlesions default	8.91 * (5.03 – 12.51)	0.872 (0.506 – 0.962)	0.911 (0.747 – 0.971)	4.26 (2.23 – 6.18)	11089 (11085 – 11094)	4.75 (2.91 – 6.40)	3.13 (1.81 – 4.85)	0.553 (0.429 – 0.641)	0.428 (0.262 – 0.577)	0.490 (0.424 – 0.586)
nicMSlesions optimized	7.39 (3.25 – 9.93)	0.975 (0.928 – 0.992)	0.975 (0.925 – 0.992)	5.41 (2.22 – 7.11)	11091 (11088 – 11096)	2.00 (1.27 – 3.05)	2.32 (2.02 – 3.61)	0.698 (0.536 – 0.736)	0.180 (0.114 – 0.275)	0.660 (0.613 – 0.716)
nicMSlesions single-subject #	5.33 *** (2.87 – 7.97)	0.746 0.189 – 0.893)	0.854 (0.809 – 0.889)	3.57 (1.77 – 5.29)	11091 (11089 – 11096)	1.54 (0.95 – 2.60)	3.59 (2.55 – 4.93)	0.501 (0.401 – 0.591)	0.139 (0.06 – 0.235)	0.568 (0.490 – 0.638)
BIANCA	7.54 (4.37 – 10.25)	0.968 (0.905 – 0.990)	0.966 (0.898 – 0.989)	5.11 (2.67 – 6.79)	11090 (11087 – 11095)	2.93 (1.92 – 3.78)	2.67 (2.08 – 3.75)	0.639 (0.521 – 0.686)	0.264 (0.173 – 0.341)	0.643 (0.514 – 0.675)

Volumes are shown as median with interquartile range with the first and last quartile; intraclass correlation coefficients (ICC) are shown with 95% confidence interval. Note that the positive and negative volumes are extracted from the native FLAIR image. # nicMSlesions single- subject output is an average over all fourteen variants (also see section 3.4). Statistics from repeated measures ANOVA with post-hoc pairwise Wilcoxon Signed Ranks test; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

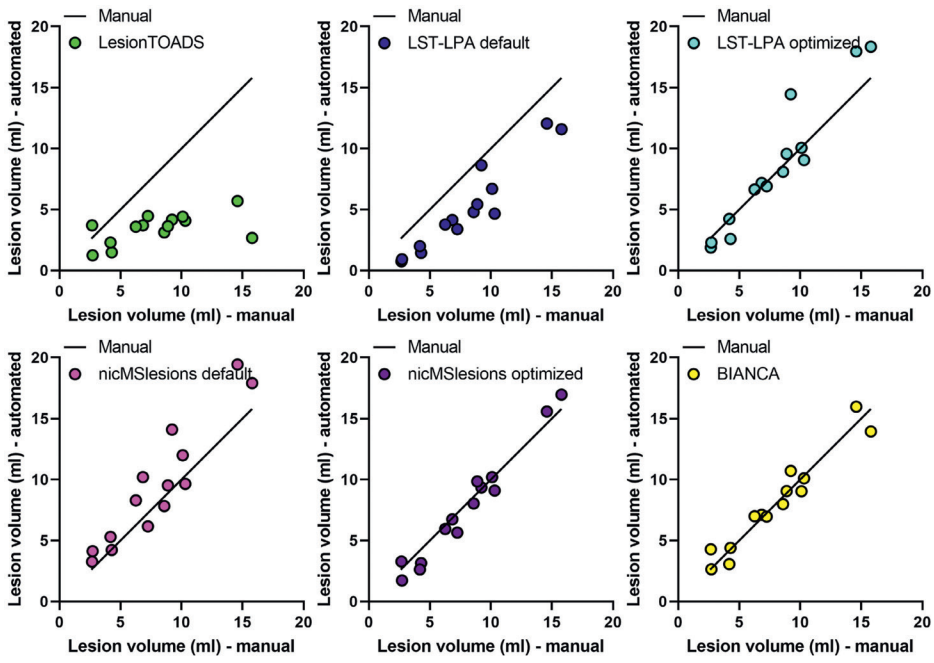

Figure 2 – Scatter plots of the automated segmentation lesion volumes versus manual lesion volumes obtained from LesionTOADS (green), LST-LPA default (blue), LST-LPA adjusted-threshold (turquoise), nicMSlesions default (pink), nicMSlesions optimized (purple), and BIANCA (yellow).

Performance of lesion segmentation methods: trained on n=1 subject with MS

Next, we looked at the results from nicMSlesions with single-subject input. In this case, we excluded the subject that was used for training from the volumetric outcomes, leading to fourteen different mean lesion volumes for manual and for nicMSlesions single-subject (**Table 3**). Here, upon training with the single-subject with the lowest lesion volume (2.68 ml), nicMSlesions failed in one of the thirteen remaining subjects to segment the lesions; this was the only failure. ICC for absolute agreement was lowest upon training single-subjects with the two lowest and the two highest lesion volumes (ICC absolute agreement ≤ 0.760). ICC consistency was also lower upon training single-subjects with low lesion volume (ICC consistency ≤ 0.758), but not upon input from subjects with higher lesion volumes. The best volumetric agreement was seen upon training using data of the single-subjects with lesion volumes of 4.19 or 4.29 ml (ICC absolute agreement = 0.941 and 0.973, respectively, and ICC consistency = 0.937 and 0.971, respectively).

Spatial agreement for the single-subject configuration of nicMSlesions was good, with median SI varying from 0.363 (input lesion volume: 2.68 ml) to 0.606 (input lesion volume: 10.31 ml) and an average median SI of 0.559 (**Table 3**). Excluding the two subjects with lesion volumes ≤ 3.0 ml (i.e. the minimum lesion volume recommended by nicMSlesions for single-subject training[4]), lowest median SI was 0.535 (input lesion volume: 9.22 ml) and the average median SI over all remaining twelve subjects was 0.578.

Table 3 – Volumetric and spatial accuracy of the different single-subjects configuration of nicMSlesions, each evaluated on the remaining thirteen subjects.

Manual volume (ml) of single-subject used for training *i.e. subject excluded from analysis*	**Median manual lesion volume (ml)** *with IQR*	**Median automated lesion volume (ml)** *with IQR*	**ICC absolute agreement** *with 95% CI*	**ICC consistency** *with 95% CI*	**Median Dice's similarity index to manual labels** *with IQR*
2.68 #	8.73 (6.40 – 10.25)	2.94 ** (2.45 – 4.24)	0.131 (-0.088 – 0.499)	0.396 (-0.201 – 0.779)	0.363 (0.311 – 0.417)
2.72	8.58 (5.27 – 10.20)	8.18 (5.17 – 9.71)	0.760 (0.397 – 0.919)	0.758 (0.379 – 0.919)	0.521 (0.462 – 0.581)
4.19	8.58 (5.27 – 10.20)	7.88 (4.63 – 10.79)	0.941 (0.819 – 0.982)	0.937 (0.806 – 0.980)	0.574 (0.500 – 0.640)
4.29	8.58 (5.22 – 10.20)	8.02 (4.94 – 10.24	0.973 (0.914 – 0.992)	0.971 (0.908 – 0.991)	0.595 (0.564 – 0.663)
6.26	8.58 (4.34 – 10.20)	5.44 ** (2.44 – 7.20)	0.753 (-0.071 – 0.944)	0.927 (0.779 – 0.977)	0.596 (0.517 – 0.620)
6.85	8.58 (4.34 – 10.20)	5.48 ** (2.87 – 7.59)	0.856 (-0.014 – 0.969)	0.955 (0.861 – 0.986)	0.562 (0.520 – 0.635)
7.24	8.58 (4.34 – 10.20)	6.09 ** (2.78 – 7.74)	0.844 (-0.029 – 0.967)	0.953 (0.853 – 0.985)	0.599 (0.557 – 0.668)
8.58	7.24 (4.24 – 10.20)	6.15 ** (3.07 – 8.15)	0.930 (0.284 – 0.985)	0.972 (0.910 – 0.991)	0.595 (0.529 – 0.663)
8.88	7.24 (4.24 – 10.20)	4.68 ** (2.31 – 650)	0.727 (-0.072 – 0.938)	0.930 (0.787 – 0.978)	0.575 (0.439 – 0.630)
9.22	7.24 (4.24 – 10.20)	5.50 ** (2.19 – 7.34)	0.819 (-0.052 – 0.963)	0.963 (0.883 – 0.989)	0.535 (0.479 – 0.653)
10.09	7.24 (4.24 – 9.76)	5.45 ** (2.39 – 7.23)	0.855 (-0.034 – 0.970)	0.962 (0.881 – 0.988)	0.590 (0.513 – 0.648)
10.31	7.24 (4.24 – 9.66)	6.34 ** (2.58 – 8.09)	0.885 (0.073 – 0.975)	0.959 (0.870 – 0.987)	0.606 (0.524 – 0.676)
14.59	7.24 (4.24 – 9.66)	4.81 ** (2.37 – 6.68)	0.696 (-0.073 – 0.929)	0.923 (0.769 – 0.976)	0.571 (0.473 – 0.623)
15.79	7.24 (4.24 – 9.66)	3.86 ** (2.01 – 4.94)	0.593 (-0.066 – 0.896)	0.907 (0.723 – 0.971)	0.540 (0.427 – 0.561)

Abbreviations: IQR gives the interquartile range with the first and last quartile. Statistics from Wilcoxon Signed Ranks test; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. # One subject's images could not be segmented by this single-subject variant of nicMSlesions; therefore volume and Dice's similarity index are determined over twelve instead of thirteen subjects.

Wilcoxon Signed Ranks tests showed that in eleven out of fourteen cases, nicMSlesions single-subject volumetric output differed significantly from manual volumes.

Compared to the other supervised, trained methods BIANCA and nicMSlesions optimized, the single-subject trained nicMSlesions showed worse volumetric and spatial (**Figure 3**) agreement, although this was highly dependent on the subject that was used for the training

of the network. However, the single-subject training of nicMSlesions showed better volumetric and spatial agreement than the unsupervised, untrained method LesionTOADS in all thirteen successfully segmented cases; and also better than the supervised, untrained method LST-LPA default in ten (volumetric) or nine (spatial) of thirteen cases. The other two supervised, untrained methods (LST-LPA adjusted-threshold and nicMSlesions default) generally also performed worse than the single-subject nicMSlesions configuration.

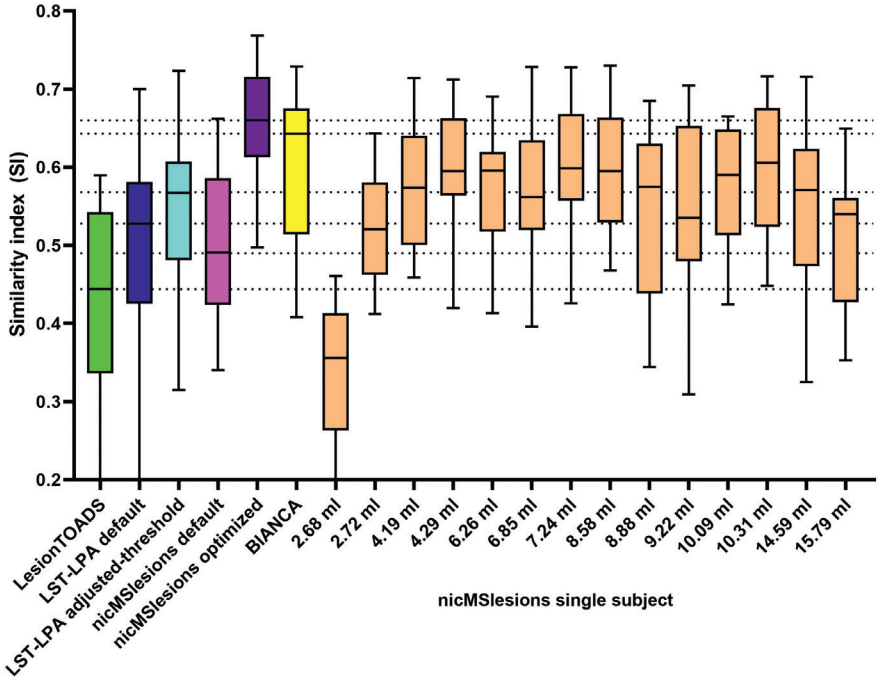

Figure 3 – Box-and-whiskers plot (min-to-max, line at median) showing Dice's similarity index (SI) in comparison to the manual lesion segmentation for LesionTOADS (green), LST-LPA default (blue), LST-LPA adjusted-threshold (turquoise), nicMSlesions default (pink), nicMSlesions optimized (purple), BIANCA (yellow), and the fourteen different nicMSlesions single-subject variants (orange). Horizontal dotted lines indicate the medians of the other automated lesion segmentation methods.

Performance of lesion segmentation methods: healthy controls

Last, we tested the various methods in healthy controls. The mean lesion volume measured in the five subjects, as well as the range from min-to-max is shown in **Table 4**. Since the manual lesion volume was 0 in all subjects, no volumetric or spatial reliability could be calculated.

Results show that the measured lesion volumes in the healthy controls was generally low in all segmentation methods. However, the supervised, untrainedmethod LST-LPA with threshold adjustment, the supervised, untrained version of nicMSlesions and the supervised and trained method BIANCA showed relatively higher false positive lesion volumes in these healthy controls, which was also seen in the subjects with MS (**Table 2**). The lowest false positive rate was seen when the single-subject trained network of nicMSlesions was used to segment the healthy control subjects.

Table 4 – Volumetric accuracy of the various grouped lesion segmentation methods compared to manual in the five healthy control subjects.

HC *n*=5	**Mean lesion volume ± SD** (range min-to-max)
LesionToads	0.81 ± 0.23 (0.45 – 1.08)
LST-LPA default	0.40 ± 0.22 (0.07 – 0.67)
LST-LPA adjusted-threshold	1.36 ± 0.11 (0.40 – 2.49)
nicMSlesions default	2.23 ± 1.77 (0.00 – 5.18)
nicMSlesions optimized	0.32 ± 0.43 (0.00 – 1.53)
nicMSlesions single-subject	0.27 ± 0.34 (0.00 – 0.94)
BIANCA	1.81 ± 0.83 (0.86 – 3.38)

Discussion

In this research, we compared four different lesion segmentation methods in different configurations to investigate the suitability of the deep learning cascaded CNN method nicMSlesions with input from only one manually delineated subject for multi-center trials. Although the supervised and trained methods nicMSlesions (re-train of full 11-layer cascaded CNN with own data) and BIANCA (optimized for own data) show the best volumetric and spatial agreement with manual, the single-subject configuration of nicMSlesions outperformed the unsupervised, untrained method LesionTOADS and the supervised, untrained method LST-LPA for both volumetric and spatial agreement with manual. Furthermore, the single-subject trained network of nicMSlesions showed the least false positives when tested on healthy controls.

There is a need for automatic lesion segmentation in MS, not only to assess lesion accrual itself, but also to assess structural brain changes such as atrophy, because the presence of MS lesions significantly hampers accurate brain segmentation[9]. Several reviews have been published showing an abundance of available MS lesion segmentation methods[2,10] and since the emergence of machine learning even more methods have surfaced, as becomes evident in the various international MS lesion segmentation challenges[3,11,12]. An important feature of automated and accurate lesion segmentation, is robustness to new data from “unseen” centers, with potentially different MR vendors and acquisition protocols than those used during development[13]. In this study, we show that re-training of the nicMSlesions network with data from an unseen center is possible with as little as one manual delineated subject as input data, showing the potential of this method for multi-center studies in MS. Important to note is that a cluster size threshold can be customized for this single-subject trained network of nicMSlesions as well, which may even further decrease the false positive rate of the segmentation.

One limitation of nicMSlesions is that for input lesion volumes below a certain threshold (in our dataset 5.0 ml), nicMSlesions has not enough data points to re-train the last two layers of the network, and therefore it only re-trains the last layer. Segmentation problems for low input lesion volumes are present in other automated lesion segmentation methods as well[14-16], although none of these methods use single-subject input. In our data, we see that training on

subjects with lesion volumes below 3.0 ml yields less accurate segmentations than training on subjects with higher lesion volumes. However, training on subjects with lesion volumes between 3.0 and 5.0 ml shows good performance, even though only the last layer of the network was re-trained in these cases. Valverde *et al.* (2018) propose that this may be due to the importance of higher lesion numbers over higher lesion load, suggesting that input for re-training of the method should be chosen carefully in order to obtain the best results with the cascaded CNN method. The other configurations of nicMSlesions tested here (i.e.the default 11-layer CNN and the 11-layer re-trained CNN with probability threshold optimization) show no problems with the segmentation when four out of the fourteen subjects had lesion volumes below 5.0 ml, but we have no data when the network is re-trained with input from subjects with lesion volumes below 5.0 ml only, which should be tested in future work to assess the suitability of the method for cohorts with low lesion load or few lesions.

Our results as shown in Figure 3 also seem to suggest that choosing training subjects that are most similar to a large part of the study population may be advantageous: while performance is roughly stable across the central part of the lesion volume distribution, partial re-training with subjects with an extreme lesion volume (i.e. very low or very high) appears to lead to reduced SI compared to manual.

We have shown that the default configuration of nicMSlesions can be optimized for protocols from an individual scanner. Even in a harmonized multi-center study with scanners from the same vendor, considerable differences in lesion volume in the same subject have been described[17]. Hardware differences such as coil configuration, gradient and radio frequency amplifiers, and other differences such as acquisition parameters, spatial resolution, and filters used in image reconstruction, can have substantial effects on the appearance of lesion in FLAIR images and on their quantification using automated software. Therefore it is not unreasonable to suspect that training on one's own specific type of images could improve performance of the nicMSlesions segmentation, and our results suggest that this is indeed the case.

As in comparable papers on image analysis methods, while the main message can be appreciated visually from inspecting the graphs and tables reflecting measured data, we did include some statistical analyses. It should be noted that such statistical testing results are reported in this paper as auxiliary information. Given the relatively small number of subjects and the fact that quite a number of comparisons are of interest, as well as the more general debate on null hypothesis testing and the legitimacy of p-values[18], the resulting p-values should be interpreted with caution and are by no means intended to suggest any definitive answers to the questions posed.

This study has some limitations. First, manual delineation was performed by one single rater, thereby the gold standard presented here may be biased towards the individual rater. Furthermore, the data presented here is from one MR vendor, thereby no information on multi-center data could be included. However, the main objective of this study was to investigate the performance of the single-subject configuration of the deep learning cascaded CNN of nicMSlesions in comparison to other methods, while varying the amount of manual input and optimization possibilities, since this has not yet been reported. Of course, although all methods have already shown their multi-vendor validity in their separate proof-of-concept papers, these data should be replicated in other datasets to further prove that single-subject nicMSlesions

outperforms the untrained methods LesionTOADS and LST-LPA. Because this paper focused on the nicMSlesions software, and because retraining is not as readily available for LST-LPA as for nicMSlesions, we did not also retrain LST-LPA using our own data. Such retraining could improve performance of LST-LPA which may affect the reported comparisons. However, it should be noted that the retraining of LST-LPA would require a group of patients rather than a single subject. Last, longitudinal performance of the single-subject configuration of nicMSlesions has not yet been shown, which should be further investigated.

In conclusion, the cascaded CNN lesion segmentation method nicMSlesions can be successfully re-trained with a limited amount of manual input, e.g. with only one manual delineation from new data, showing higher volumetric and spatial agreement with manual lesion segmentation than obtained with the commonly used untrained lesion segmentation methods LST-LPA and LesionTOADS. The method should be further optimized for cohorts with low lesion loads and in longitudinal, multi-center studies.

Funding

This work was supported by the Dutch Multiple Sclerosis Research Foundation (grant number 14-876 MS).

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplements

Supplemental Table 1 – Optimization of the probability threshold of LST-LPA and nicMSlesions in our dataset.

Probability threshold	Mean SI compared to manual (range min-max)	
	LST-LPA	nicMSlesions
0.20	0.540 (0.331 – 0.710)	0.648 (0.513 – 0.743)
0.25	**0.546 (0.315 - 0.723)**	0.653 (0.511 – 0.752)
0.30	0.545 (0.288 – 0.729)	0.656 (0.508 – 0.760)
0.35	0.539 (0.259 – 0.729)	0.658 (0.502 – 0.766)
0.40	0.530 (0.231 – 0.723)	**0.658 (0.497 – 0.769)**
0.45	0.516 (0.208 – 0.714)	0.656 (0.484 – 0.769)
0.50 (default)	0.499 (0.181 – 0.700)	0.652 (0.468 – 0.769)
0.55	0.478 (0.151 – 0.683)	0.648 (0.468 – 0.768)
0.60	0.455 (0.127 – 0.661)	0.642 (0.446 – 0.768)
0.65	0.427 (0.106 – 0.637)	0.632 (0.425 – 0.766)
0.70	0.394 (0.085 – 0.609)	0.620 (0.416 – 0.761)

Supplemental Table 2 – Optimization of the various parameters for BIANCA; results from threshold 0.99 and cluster-size threshold 3.

Configuration BIANCA	Number of lesion training points	Number of non-lesion training points	Location of non-lesion points	Patch	Patch size	Spatial weighting	Mean SI to manual
1	2000	equal	any	nopatch	0	1	0.524
2	2000	5000	any	nopatch	0	1	0.542
3	2000	100000	any	nopatch	0	1	0.499
4	all	equal	any	nopatch	0	1	0.541
5	2000	equal	noborder	nopatch	0	1	0.520
6	2000	equal	surround	nopatch	0	1	0.262
7	2000	equal	any	3D	1	1	0.524
8	2000	equal	any	3D	3	1	0.586
9	**2000**	**equal**	**any**	**3D**	**5**	**1**	**0.602**
10	2000	equal	any	3D	6	1	0.592
11	2000	equal	any	3D	7	1	0.597
12	2000	equal	any	3D	9	1	0.594
13	2000	equal	any	3D	11	1	0.585
14	2001	equal	any	nopatch	0	0	0.428
15	2000	equal	any	nopatch	0	0.5	0.489
16	2000	equal	any	nopatch	0	2	0.547
17	2000	equal	any	nopatch	0	5	0.556
18	2000	equal	any	nopatch	0	10	0.538

References

1 Danelakis, A., Theoharis, T. & Verganelakis, D. A. Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. Comput Med Imaging Graph 70, 83-100, doi:10.1016/j.compmedimag.2018.10.002 (2018).

2 Garcia-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L. & Collins, D. L. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med Image Anal 17, 1-18, doi:10.1016/j.media.2012.09.004 (2013).

3 Commowick, O. et al. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. Sci Rep-Uk 8, doi:ARTN 13650 10.1038/s41598-018-31911-7 (2018).

4 Valverde, S. et al. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. Neuroimage Clin 21, 101638, doi:10.1016/j.nicl.2018.101638 (2019).

5 Valverde, S. et al. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage 155, 159-168, doi:10.1016/j.neuroimage.2017.04.034 (2017).

6 Shiee, N. et al. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. Neuroimage 49, 1524-1535, doi:10.1016/j.neuroimage.2009.09.005 (2010).

7 Schmidt, P. Bayesian Inference for Structured Additive Regression Models for Large-scale Problems with Applications to Medical Imaging, LMU München, (2017).

8 Griffanti, L. et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. Neuroimage 141, 191-205, doi:10.1016/j.neuroimage.2016.07.018 (2016).

9 Gonzalez-Villa, S. et al. Evaluating the effect of multiple sclerosis lesions on automatic brain structure segmentation. Neuroimage Clin 15, 228-238, doi:10.1016/j.nicl.2017.05.003 (2017).

10 Mortazavi, D., Kouzani, A. Z. & Soltanian-Zadeh, H. Segmentation of multiple sclerosis lesions in MR images: a review. Neuroradiology 54, 299-320, doi:10.1007/s00234-011-0886-7 (2012).

11 Styner M., L. J., Chin B., Chin M.S., Commowick O., Tran H., Markovic-Plese S., Jewells V., Warfield S. 3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation. The MIDAS Journal - MS Lesion Segmentation (MICCAI 2008 Workshop) (Nov 2008).

12 Carass, A. et al. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. Neuroimage 148, 77-102, doi:10.1016/j.neuroimage.2016.12.064 (2017).

13 de Sitter, A. et al. Performance of five research-domain automated WM lesion segmentation methods in a multi-center MS study. Neuroimage 163, 106-114, doi:10.1016/j.neuroimage.2017.09.011 (2017).

14 Steenwijk, M. D. et al. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). Neuroimage Clin 3, 462-469, doi:10.1016/j.nicl.2013.10.003 (2013).

15 Khayati, R., Vafadust, M., Towhidkhah, F. & Nabavi, M. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model. Comput Biol Med 38, 379-390, doi:10.1016/j.compbiomed.2007.12.005 (2008).

16 Schmidt, P. et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. Neuroimage 59, 3774-3783, doi:10.1016/j.neuroimage.2011.11.032 (2012).

17 Shinohara, R. T. et al. Volumetric Analysis from a Harmonized Multisite Brain MRI Study of a Single Subject with Multiple Sclerosis. AJNR Am J Neuroradiol 38, 1501-1509, doi:10.3174/ajnr.A5254 (2017).

18 McShane, B. B., Gal, D., Gelman, A., Robert, C. & Tackett, J. L. Abandon Statistical Significance. The American Statistician 73, 235-245, doi:10.1080/00031305.2018.1527253 (2019).

CHAPTER 2.2

Lesion simulation software LESIM: A robust and flexible tool for realistic simulation of white matter lesions

Alexandra de Sitter
Merlin M. Weeda
Iman Brouwer
Menno M. de Boer
Rick J. van Tuijl
Petra J.W. Pouwels
Ferderik Barkhof
Hugo Vrenken

Under revision

Abstract

Background

In MS, the most obvious pathological brain changes are WM-lesions and GM atrophy. any image analysis methods that quantify GM atrophy exhibit impaired performance in the presence of WM-lesions. To allow objective investigations of the effects of WM-lesions and facilitate the development of segmentation methods that are robust to the presence of WM-lesions, we here present a novel lesion simulation method (LESIM).

Methods

The LESIM simulates lesions from a MS-patient into a 3D-T1weighted images of a healthy control (HC). The pipeline performs pre-processing, non-linear registration of patient image to HC, normalization of intensity levels, followed by imputation of lesion voxels with a smooth transition border to the surrounding tissue and adjustment of noise levels to match the target image. This creates a modified 3D-T1weighted image with realistic WM-lesions. To evaluate the software, data from five early relapsing-remitting MS patients and five HC were combined pairwise, to create 25 images with simulated lesions. The simulated images were evaluated by visual inspection as well as quantitative analysis of the effect on FSL-SIENAX GM segmentations.

Results

Visual inspection showed that LESIM simulates natural-looking lesions in the correct locations, and with the correct signal to noise ratio and intensity. Total GM volumes obtained from FSL-SIENAX were increased after lesion segmentation. The influence of lesions was clearly visible in the GM partial volume estimate (PVE) maps, which showed a dramatic increase of GM in regions with simulated lesions compared to GM-PVE maps of HC without simulated lesions.

Discussion

LESIM is a new, robust and flexible tool for reliable WM-lesion simulation. The registration steps, the correction for noise and intensity differences, and the smooth transition from the edges of the lesions to the surrounding tissue results in realistic lesions in HC images. Moreover, the simulated WM-lesions have the expected effect on GM segmentation using FSL-SIENAX.

Introduction

Multiple sclerosis (MS) is a disease of the central nervous system, mainly characterized by inflammatory demyelinating lesions in the white matter (WM)[1] and by neurodegeneration and atrophy of the grey matter (GM)[2]. Magnetic resonance imaging (MRI) enables visualization and quantification of both WM and GM pathology, thereby playing a crucial role in diagnosis and research in MS.

Brain MR image analyses such as image registration, segmentation and volumetric measurement are known to be affected by the presence of MS lesions[3-9]. For example, Popescu et al. (2014) showed that WM and (deep) GM volumes were significantly different in images with or without lesion filling applied[5]. Although those studies investigated the effect of filling lesions, they were restricted to images of patients in which lesions were already present and therefore lacked an independent ground truth without lesions. Instead, Chard et al. (2010) introduced synthetic lesions in MR images of healthy subjects and showed that filling these synthetic lesions had an effect on automated segmentation of WM and GM[7]. While important, those results were still limited because they involved spherical, homogeneous lesions that differ from true lesions in MS and other disease with WM lesions. Therefore, a correct simulation of WM lesions in HCs could help to better understand the effects of real lesions on image analysis methods. Simulation of realistic lesions in healthy control images enables investigation of the isolated effect of real lesions (by reproducing their topology, shapes and intensities), without other possible influencing factors such as brain atrophy (which are often present in patient images).

Currently, two lesion simulation tools are available: MS-MIST[10] is a tool to simulate predefined lesion volumes in multiple sequences (e.g. FLAIR and diffusion tensor imaging), and BrainWeb (http://www.bic.mni.mcgill.ca/brainweb) is a tool to simulate a predefined lesion mask into a 2-D T1-, T2- or proton density-weighted image. However, these tools cannot simulate a range of different lesion volumes, nor do they enforce a correct signal intensity and signal-to-noise ratio to the simulated lesions.

Therefore we here introduce the LESIM software. LESIM is a new lesion simulation method that enables the user to introduce realistic WM lesions from MR images of MS patients in healthy control (HC) images, where lesions are transferred with correct shapes, locations, intensities and signal-to-noise ratios. The manipulated HC images can be used to assess exclusively the effect of WM lesions on image analysis methods (e.g. brain segmentation) by comparing results obtained from these lesion-simulated images to those obtained from the unmodified HC images. We describe the software and test its performance visually, as well as quantitatively by evaluating the effect of introducing simulated lesions on brain volumes obtained by FSL-SIENAX.

Methods

Subjects

From a larger internal Amsterdam UMC cohort study of subjects with relapsing remitting MS (RRMS) and HCs, a total of ten subjects were included in this study, of which five HCs and five RRMS patients who were matched for sex and age.[11] The study was approved by the local institutional medical ethics committee and written informed consent was obtained from all individuals, according to the Declaration of Helsinki.

MRI examination

Subjects underwent MRI examination on a 3T whole-body MRI scanner (GE Discovery MR750) with an 8-channel phased-array head coil. The protocol included a sagittal 3D T1-weighted fast spoiled gradient echo sequence (with TR/TE/TI = 8.2/3.2/450 ms and voxel size 1.0x1.0x1.0 mm) and a sagittal 3D fluid attenuated inversion recovery sequence (FLAIR with TR/TE/TI = 8000/130/2338 ms with voxel size 1.0x1.0x1.2 mm). The WM lesions were manually outlined on FLAIR images of MS patients by an expert rater[11].

Lesion simulation software: LESIM

The LESIM software enables the user to simulate lesions, sampled in this study from MS patients, in HC data. As input for the pipeline, LESIM requires two 3D T1-weighted images: one of a patient and one of a HC. Furthermore, the patient's lesion mask is required, which should be provided in the same space as the patient's T1 image. Optionally, a 3D FLAIR image can additionally be provided for the patient and the patient's lesion mask should then be in the 3D FLAIR space; this option was added to improve ease of use, by not requiring lesion masks in T1 space to be prepared by the user.

LESIM makes use of several different software packages. First, FMRIB Software Library (FSL) version 5.0.10 is used for, amongst others, brain extraction[12,13], segmentation of the tissue[14], (non)-linear registration[15,16], registration to standard space (MNI152, http://www.bic.mni.mcgill.ca/), and several calculations using the *fslmaths* and *fslstats* functions. Second, FreeSurfer[17,18] version 6.0.0 is used for N3 correction[19], and can also be used for tissue segmentation. Furthermore, LEsion Automated Preprocessing (LEAP)[7] is used for lesion filling and Elastix[20] version 4.7 is used for the non-linear registration of patient data to the HC images.

The LESIM pipeline consists of five separate steps, which are discussed below; an overview of the pipeline is shown in **Figure 1** and code is available.

Step 1: Processing of the patient image

The processing of the patient's data is depicted in **Supplementary Figure 1**. In this step, the patient's data is processed in such a way that the lesion mask can later be applied to the HC image. The T1 image of the patient is needed as input, as well as the lesion mask and if needed the FLAIR. The separate steps for the processing of the patient images are:

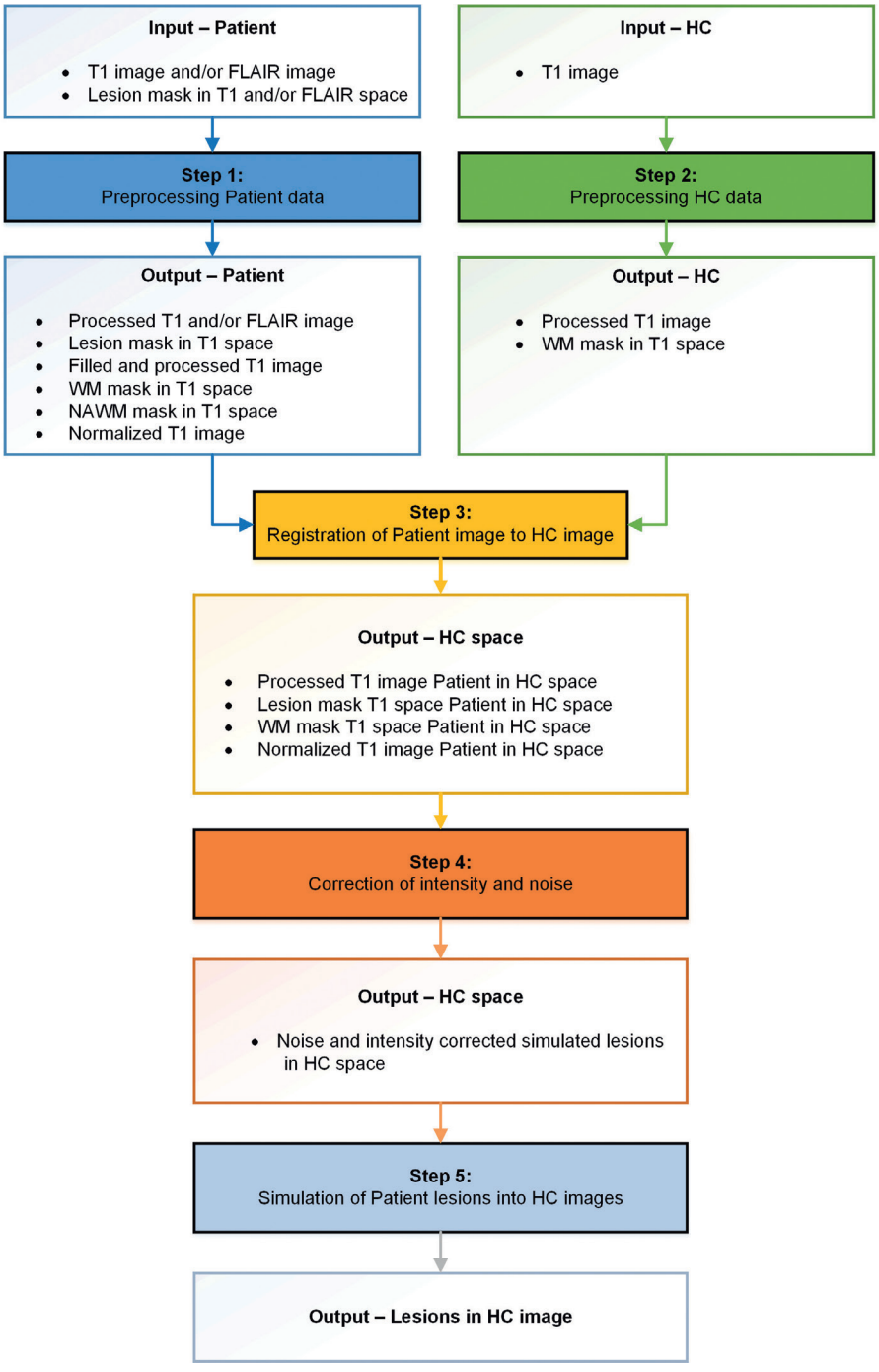

Figure 1 – Overview of pipeline LESIM with 5 main steps and all input and output data.

Abbreviations: HC = healthy control, FLAIR = fluid attenuated inversion recovery, WM = white matter, NAWM = normal appearing white matter.

A) Preprocessing of T1 patient image

This step entails neck removal and brain extraction with FSL, as well as bias field (N3) correction with FreeSurfer, resulting in the preprocessed T1 patient image.

B) Preprocessing of lesion mask

If the lesion mask is in the space of a FLAIR image, continue with step B1. If the lesion mask is the space of a T1 image, continue with step B2.

B1) When the lesion mask is in the space of a FLAIR image:

i. Lesions smaller than 5 voxels are removed from the FLAIR lesion mask to prevent the loss of lesions due to interpolation in later steps.
ii. The FLAIR image is registered to the preprocessed T1 patient image using linear registration with six degrees of freedom with mutual information as cost function (FSL-FLIRT). The inverse transformation is used to bring the brain mask to FLAIR space for N3 bias correction An improved transformation from FLAIR to T1 space is calculated based on bias corrected versions of the images.
iii. The FLAIR lesion mask is transformed to the preprocessed T1 patient image using the transformation matrix obtained in B1-ii with nearest neighbor interpolation (FSL-FLIRT). This results in the lesion mask in the space of the preprocessed T1 patient image.
iv. Again, lesions smaller than 5 voxels are removed from the lesion mask in the space of the preprocessed T1 patient image.

B2) When the lesion mask is in the space of a T1 image:

i. Lesions smaller than 5 voxels are removed from the T1 lesion mask with FSL. This results in the binary lesion mask in the space of the preprocessed T1 patient image.

C) Lesion filling

For correct segmentation of the WM, first the lesion are filled. Lesions in the space of the preprocessed T1 patient image are filled with normal appearing white matter (NAWM) intensities using LEAP. This results in a filled and preprocessed T1 patient image.

D) Normalizing of the T1 patient image

The last step of the patient processing pipeline normalizes the WM intensities in the preprocessed T1 patient image. For this, a WM mask is created from the filled and preprocessed T1 patient image, which can be done in the current implementation of the software with either FSL-FAST (default) or FreeSurfer. A dilated lesion mask is subtracted from this WM mask, to ensure that the new mask will only contain NAWM. From this NAWM mask the mean signal intensity is calculated. Next, the preprocessed T1 patient image is divided by this mean signal intensity of the NAWM, which results in a normalized T1 patient image. These steps are performed with FSL.

Step 2: Processing of the HC image

In this step (**Supplementary Figure 2**), neck removal and brain extraction are performed with FSL, and bias field (N3) correction is performed with FreeSurfer, which results in a processed T1 HC image. Furthermore, the WM from this image is segmented with either FSL-FAST (default) or FreeSurfer, resulting in the WM mask in the space of the preprocessed T1 HC image.

Step 3: Registration of the patient image to the HC image

At this point, the data from the above two pipelines will be combined, leading to the patient's data in the space of the preprocessed T1 HC image. The pipeline (**Supplementary Figure 3**) consists of the following steps:

A) Whole brain nonlinear registration

Registration of the filled and preprocessed T1 patient image to the processed T1 HC image is performed with two-step nonlinear registration using Elastix. The first registration will align the skulls of the patient's and HC's T1 image using a B-spline cost interpolation. The second registration will align the brains in the same manner. Then, a combination of both alignments is computed and written into a parameter file, which contains the matrix needed to achieve the registration.

B) Transformation of patient masks

The parameter file from step 3A is used for transformation of the lesion mask and WM mask, both in the space of the preprocessed T1 patient image (step 1), to the preprocessed T1 HC image (step 2), using nearest neighbor interpolation (Elastix). Both masks are now in the same space as the preprocessed T1 HC image.

C) Transformation of preprocessed T1 patient image

The parameter file is also used for the registration of the processed T1 patient image and normalized T1 patient image (step 1) to the processed T1 HC image (step 2) using a B-spline interpolation (Elastix). Now, the processed and normalized T1 patient images are in the same space as the T1 HC image.

D) Creating the lesion mask and simulate (uncorrected) lesions

In the processed T1 HC image space, the binary lesion mask is overlaid with the HC WM mask (step 2) to ensure that lesions can only be simulated in the WM regions of the HC image. This results in the lesion mask in the space of processed T1 HC image. Applying this mask with the processed T1 HC image results in simulated (uncorrected) lesions (FSL).

Step 4: Correction of intensity and noise

Since the T1 images from the patient and HC can, in general, have different signal intensities and different signal-to-noise ratios (SNRs), these need to be corrected when simulating the patient lesions into the processed T1 HC image. This is done with the use of FSL with the following steps (**Supplementary Figure 4**):

A) Intensity correction

First, the mean signal intensity is calculated in the WM mask in the space of the processed T1 HC image (step 2). The processed T1 HC image is divided by this mean signal intensity, producing a normalized T1 HC image. This has already been done for the patient's image in step 1E. Then, the scaling between the normalized T1 HC and patient image is calculated, and applied to the simulated (uncorrected) lesions in the processed T1 HC image (step 3D), resulting in intensity corrected simulated lesions in the processed T1 HC image.

B) Noise correction

The processed T1 images of both the patient and the HC are linearly registered to standard space (MNI152 , http://www.bic.mni.mcgill.ca/). A ROI defined in MNI-space coordinates (**Supplementary Figure 5**) is then brought to the patient's and HC's processed T1 image spaces using the inverse transforms. This ROI is then dilated and applied to both the HC WM mask and patient WM mask. Within both resulting WM ROIs, the SNR is calculated using $SNR = \frac{\mu^2}{\sigma^2}$, where μ is the mean of signal intensity and σ is the standard deviation. After calculation of the SNR in both the patient and the HC image, there are three options, based on an arbitrarily defined cut-off of an SNR difference of 10%:

i. If the SNR in the patient image is more than ten percent smaller than in the HC, random Gaussian noise is added to intensity corrected simulated lesions in HC space (created in 4A) with the magnitude of the difference between the two SNRs.
ii. If the SNR in the patient image is more than ten percent larger than in the HC, then intensity corrected simulated lesions in HC space (created in 4A) is smoothed using FSL-SUSAN (22), which uses nonlinear filtering to reduce the noise while preserving the different structures.
iii. Else (i.e. less than ten percent difference between SNR of the T1 patient and HC image), there is no need for noise correction.

All three options result in corrected simulated lesions in the space of the processed T1 HC image.

Step 5: Simulation of the patient's lesions into the HC image

The last steps (**Supplementary Figure 6**) of the lesion simulation into the processed T1 HC image are performed with FSL and are as follows:

A) Create border for transition WM and lesion

Since intensities differ between the WM and the corrected simulated lesions, a border of one voxel is created around the lesions. Within this border, the signal intensity is averaged between the WM and lesion intensity, so that a smoother transition from the surrounding WM to the lesion is created.

B) Lesion simulation in HCs

Finally, this border and the corrected simulated lesions replace the corresponding voxels in the processed T1 HC image, which results in the T1 HC image with simulated lesions.

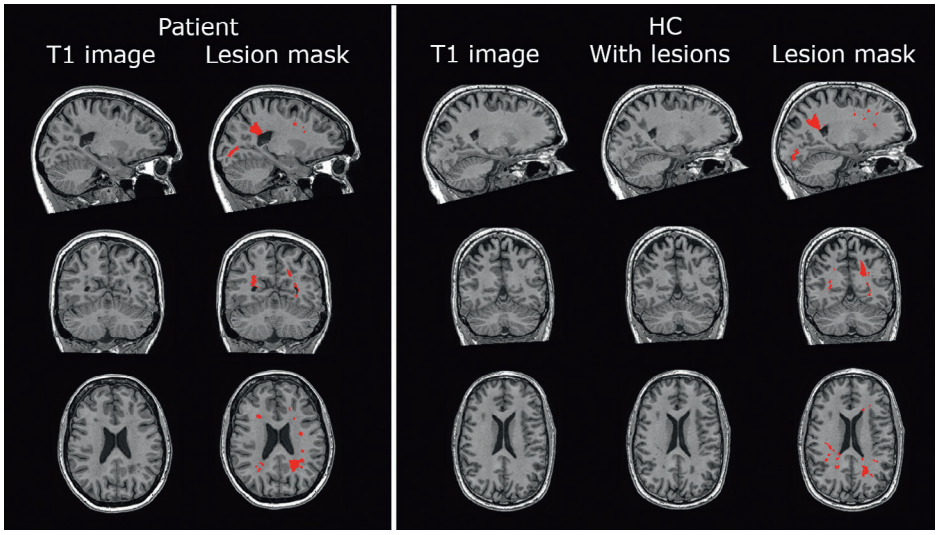

Figure 2 – Image of patient and healthy control with simulated lesions of patient.

From left to right: T1 patient image, lesion mask (red) on T1 patient image, T1 healthy control image, T1 healthy control image with simulated lesions from shown patient and lesion mask (red) on T1 healthy control image. Note :lesion mask of patient was manually outlined on FLAIR image and transformed to T1 image with nearest neighbor interpolation. Abbreviations: HC = healthy control.

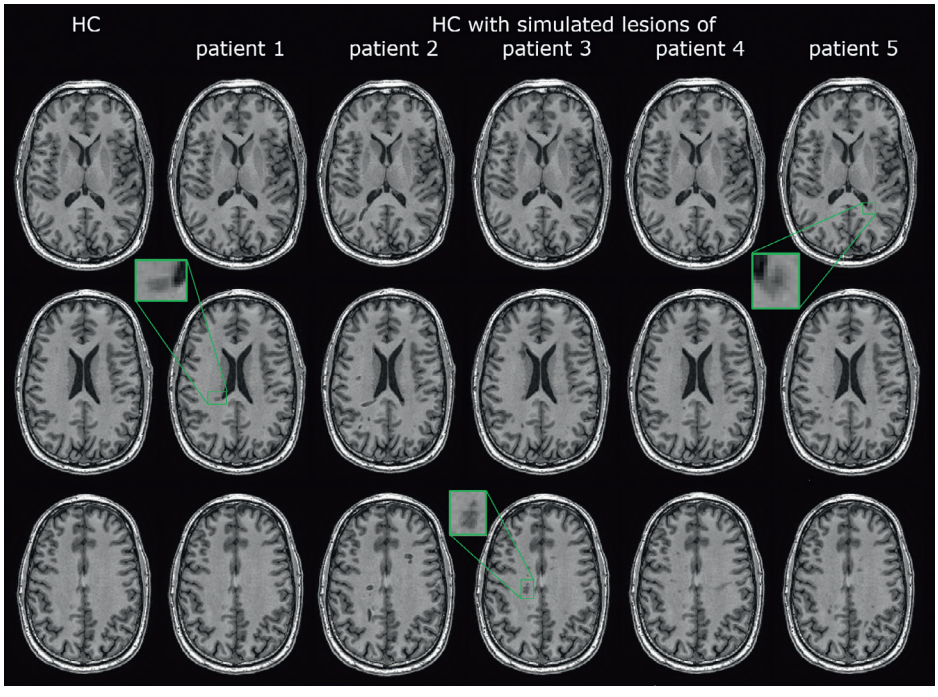

Figure 3 – Axial slices of T1 healthy control image and with simulated lesions of the five patients.

Abbreviations: HC = healthy control.

Evaluation and statistical analysis

For evaluation, we simulated a total of 25 3D T1-weighted images with lesions, by transposing the lesions of all five patients to the images of all five HCs. In order to evaluate whether the simulated lesions have the proposed effect on the segmentation of the HC T1 image, the native and simulate images were both segmented with SIENAX[21,22]. With SIENAX, the brain is segmented in three partial volume estimates: GM, WM, and cerebrospinal fluid (CSF). SIENAX then calculates the volumes of the whole brain, and the GM and WM. For this study, we used the GM volumes of SIENAX.

Results

Demographics

Of the five subjects with RRMS, one was male (20%). The mean age of all patients was 39.4 ± 5.9 years (range: 32.2 – 47.7 years) and the mean disease duration was 3.4 ± 1.5 years (range: 0.9 – 4.7 years) with a mean expanded disability status scale (EDSS) of 3.0 (range: 1.0 – 4.0). Two of the five HCs were male (40%) and their mean age was 32.9 ± 9.1 years (range: 22.9 – 45.7 years). There was no significant difference between the patients and HCs for baseline age ($p = 0.215$) or gender ($p = 0.490$).

Characteristics of simulated lesions

Visual appearance

An example of the lesion simulation from a patient's lesion mask into a HC image is shown in **Figure 2**. For one HC, axial slices of the native HC and HC with simulated lesions from all five patients are shown in **Figure 3**. Different simulated lesions are visible and for three lesions, an enlarged view is provided. The realistic appearance of the simulated lesions is clearly visible.

Signal intensity characteristics

Figure 4 shows an intensity profile taken along a line through a lesion and surrounding tissue, for both the original lesion in the patient and the simulated lesion in the healthy control image. Note that the intensity scales of the two profiles are different because of the different intensity ranges in the patient and control images. The two profiles show great similarity, confirming that LESIM has accurately reproduced the intensity profile of the original lesion in the simulated lesion.

Lesion volumes

Simulated lesion volumes may differ from original lesion volumes. The LESIM software performs nonlinear registration to project lesions as accurately as possible onto the correct anatomical location in the control image, thereby correcting for differences between the patient image and the control image in head size and brain shape (due to, e.g., atrophy). Therefore, the simulated lesion volumes in control T1 space are not expected to be identical to the original lesion volumes in patient T1 space. The results show that they are indeed not the same.

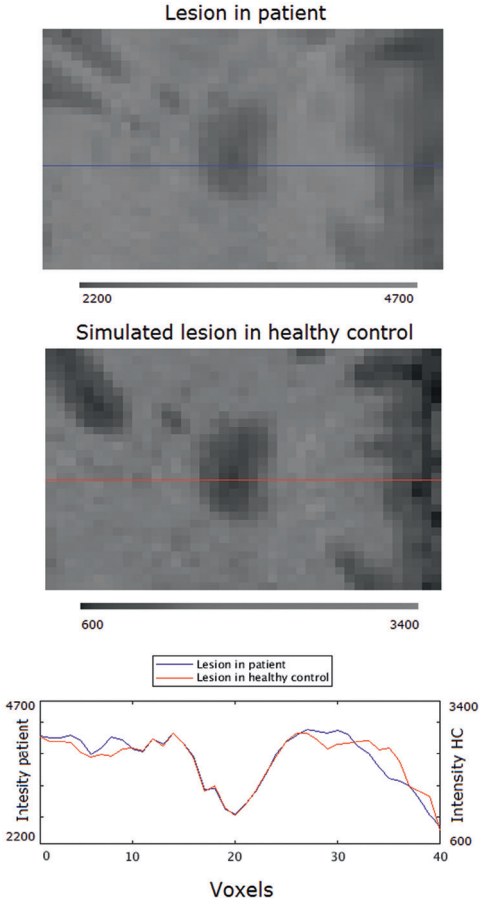

Figure 4 – Intensity plot of the intersection of a lesion (blue) and a simulated lesion (red) with surrounding tissue. Note the intensity axes are different for the lesion in patient and simulated patient.

The total volume of the lesion masks in the patients and of the simulated lesion masks in the HCs are listed in **Table 1**. Mean lesion volume in the five patients was 8.7 ± 4.3 milliliter (ml) (range: 4.3 – 15.8 ml). Upon simulation of the lesions into the five different HCs, mean lesion volume in the five patients was 7.2 ± 4.0 ml (range: 2.3 – 15.4 ml). Besides differences in head sizes and the deformation induced by the non-linear registration, the removal of small lesions and the use of HC WM masks to avoid simulating lesions outside WM regions, can also result in lesion volume differences between simulated and original lesions.

Brain volumes with and without simulated lesions

Mean whole brain volume for the five HCs was 1280 ± 109 ml (range: 1118 – 1451 ml). The mean GM and WM volumes were 579 ± 85 ml (range: 511 – 717 ml) and 701 ± 31 ml (range: 668 – 734 ml), respectively. The segmentation by SIENAX was affected by the simulated lesions. One example of the partial volume estimates (PVE) of the GM in the native HC and HC with simulated lesions is shown in **Figure 5**, as well as the difference between the PVEs of the

GM. In addition, the lesion mask with borders is shown to enable comparison of the simulated lesions with the PVE differences.

Table 1 – Original lesion volumes from the five patients, and the simulated lesion volumes from the five patients in the five healthy controls (HC).

		Lesion volume (ml)				
		Patient 1	**Patient 2**	**Patient 3**	**Patient 4**	**Patient 5**
	Native	6.86	7.14	9.10	15.94	4.16
Simulated in HC	HC 1	7.13	6.22	6.76	15.35	2.76
	HC 2	6.51	5.30	4.83	12.69	2.95
	HC 3	8.74	7.18	7.22	17.37	3.41
	HC 4	6.27	5.61	5.05	11.91	2.73
	HC 5	6.35	5.52	5.39	13.91	2.26

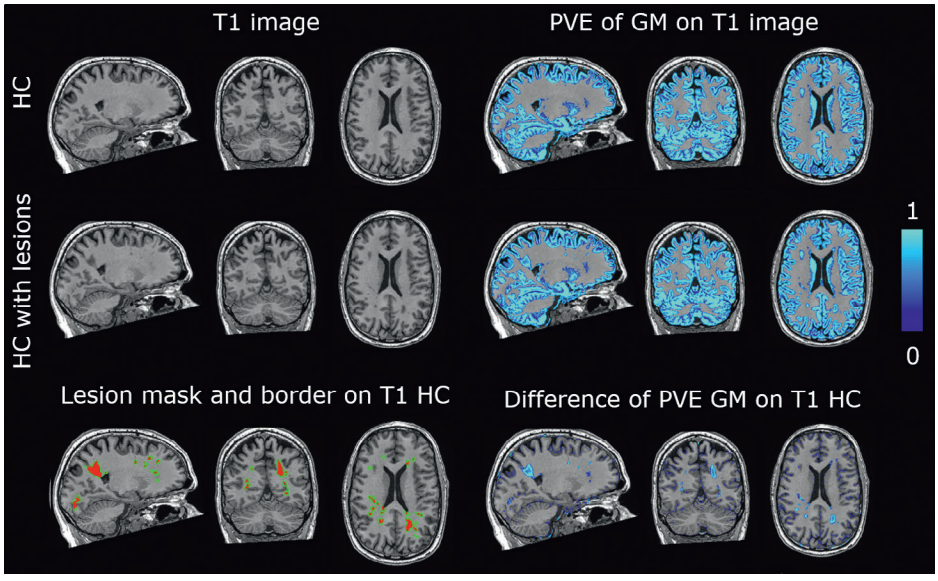

Figure 5 – Effect of simulated lesions of partial volume estimation (PVE) of GM.

With first the T1 image of the healthy control and a healthy control image with simulated lesions and the associated partial volume estimation of GM (scale from 0-1). Secondly, lesion mask (red) and adjusted border (green) on T1 healthy control image and difference of PVE of T1 image of the healthy control and a healthy control image with simulated lesions.

Abbreviations: HC = healthy control, PVE = partial volume estimation, GM = grey matter.

The GM volumes measured in the lesion-simulated HC images, as well as the relative difference with respect to the native HC GM volumes, are listed in **Table 2**. **Figure 6** plots the percentage difference in GM volumes between lesion-simulated and native HC images against the lesion volume in that image, showing a general trend for larger lesion volumes to be associated with a higher percentage of GM volume overestimation.

Table 2 – Grey matter volumes in the native HC images and the lesion simulated HC images, as well as the percentage change between native and lesion simulated images.

		HC 1		HC 2		HC 3		HC 4		HC 5	
		volume (ml)	difference (%)	volume (ml)	difference (%)	volume (ml)	difference (%)	volume (ml)	difference (%)	volume (ml)	difference (%)
Native HC		716.01		667.71		733.78		720.06		668.01	
HC with lesions from	Patient 1	717.93	0.27	669.42	0.26	738.72	0.67	721.03	0.14	669.26	0.19
	Patient 2	715.74	-0.04	667.18	0.02	735.10	0.18	720.25	0.03	668.06	0.01
	Patient 3	717.86	0.26	669.18	0.22	738.49	0.64	721.77	0.24	669.95	0.29
	Patient 4	718.99	0.42	670.89	0.48	739.57	0.79	723.02	0.41	671.39	0.51
	Patient 5	716.03	0.00	668.74	0.15	737.19	0.46	720.96	0.12	668.78	0.12

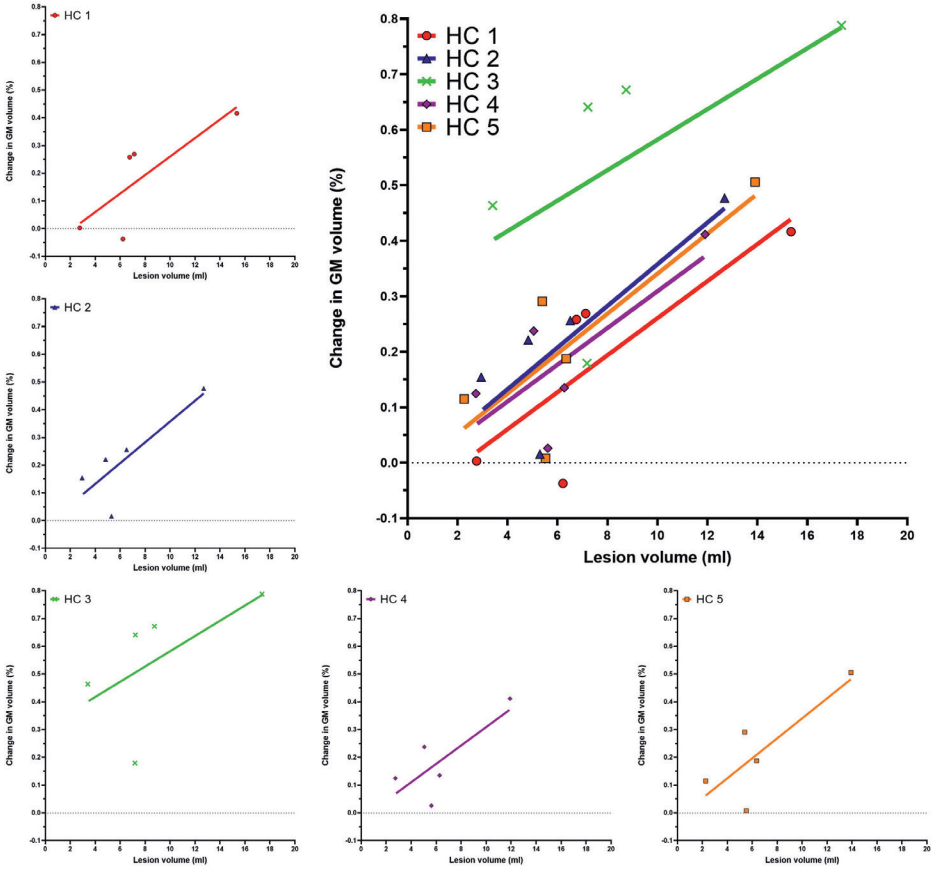

Figure 6 – The percentage change in GM volumes between native and simulated images plotted against the lesion volume in that image.

Discussion

This paper introduced the lesion simulation software LESIM, which enables the implementation of realistic WM lesions from MR images of MS patients into HC images. In addition to their realistic visual appearance, MS lesions simulated using LESIM also affected the GM volumes measured by FSL-SIENAX similarly to real MS lesions. Inspection of the signal intensity profiles confirms that the software reproduces the lesion signal intensity profiles correctly. This confirms that LESIM provides a realistic method to study the effect of lesions on image analysis while eliminating other influences.

To ensure anatomical correspondence between original and simulated lesions, the emphasis in LESIM was placed on the correct registration of the lesions into the HCs WM. Therefore, we filled the lesions in the patient image before registration, since previous studies suggested that this improves registration[7,8]. Furthermore, we optimized the registration from patient to HC data by using a two-step registration procedure in Elastix[20]. Even though the registration is optimized, it cannot be fully excluded that lesions could still be registered to regions of GM or CSF in the HC image. By applying WM masks, we ensure that lesions can only be simulated in WM of HC images.

An important feature of LESIM that has not been applied in other methods, is the correction for differences in signal intensity and SNR between patient and HC images. The difference in signal intensity between patient and HC images could be due to arbitrary MR signal scaling[23] and/or due to physiological variability (e.g. food intake, body fat, age)[24,25]. Furthermore, correction of differences in SNR and signal intensity is especially needed when using different scanners and/or different acquisition sequences. The correction of the intensity and SNR ensures that the lesions are simulated as both visually and quantitatively realistic lesions.

Visual inspection of the 25 images with simulated lesions generated in this study showed that the simulated lesions had the correct signal intensity as well as correct anatomical locations. A more quantitative analysis was performed with use of SIENAX, which is known to be affected by the presence of lesions[5,26]. Our data clearly show the effect of the simulated lesions on the brain segmentation, as can be seen in e.g. the GM partial volume estimate maps. This effect is in line with results in the literature and confirms our hypothesis that the simulated lesions affect image analysis software in a manner that is highly similar to that of real lesions[5].

This study is not without limitations. In order to maintain focus of the present paper, which is description of the software, only a small initial validation was performed, involving five patient images and five healthy control images. Future work should investigate more thoroughly the behavior of the software in larger datasets and its robustness across scanners. Nonetheless the small validation already revealed the possibility that lesions taken from different patients may exert different effects. GM volumes after simulating the lesions of patient 2 were less affected compared to simulated lesions taken from other patients. Visual inspection showed that lesions in the patient 2 T1 image had comparatively low signal intensities, which were transformed into low signal intensities in the HC control images and then mostly segmented as CSF by SIENAX, with little effect on GM segmentation. Lesion intensities cannot be adjusted in LESIM; instead the software reproduces the relative intensities from the patient images.

While this allows a detailed study of how different true lesion intensity patterns affect image analysis methods, a future extension of the method might incorporate a method to control the degree of hypointensity of lesions in order to provide the user with more control of this aspect. Similarly, the anatomical distribution of lesions is defined by the lesion mask, which is currently used without modification; future extensions may incorporate the ability to select or exclude lesions in specific locations. LESIM can be used with lesions from either FLAIR or T1 images from the patient. In this study, lesions were manually segmented in 3D FLAIR images. Therefore, it is possible that not all segmented lesions are also visible in the T1 images. In the future, adjustment of the LESIM pipeline may accommodate lesion simulation in FLAIR images as well which could help to investigate lesion segmentation of lesion filling methods[27]. Finally, LESIM may also be optimized and evaluated for other types of WM lesions and disease, such as vascular lesions and stroke.

In conclusion, LESIM is a new, robust and flexible tool for reliable lesion simulation from MS patients into HCs. Using this tool, the effect of lesions on different image processing tools can be studied in isolation without interference from other brain anomalies. In the future, LESIM may be adjusted so that it is able to simulate lesions in FLAIR images as well, and for other disease types than MS.

Funding

This work was supported by the Dutch MS Research Foundation, grant number 14-876. FB is supported by the NIHR biomedical research centre at UCLH.

Data availability statement

Data is not available for other research groups, because of ethical and privacy issues. The raw code of the tool is available.

Supplements

Supplemental Figure 1 – Pipeline from Step 1: Processing of Patient image
Supplemental Figure 2 – Pipeline from Step 2: Processing of HC image
Supplemental Figure 3 – Pipeline from Step 3: Registration of Patient image to HC image
Supplemental Figure 4 – Pipeline from Step 4: Correction of intensity and noise
Supplemental Figure 5 – Pipeline from Step 5: Simulation of Patient's lesions into HC image

References

1 Lucchinetti, C. et al. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. Ann Neurol 47, 707-717 (2000).

2 Bergsland, N. et al. Subcortical and cortical gray matter atrophy in a large sample of patients with clinically isolated syndrome and early relapsing-remitting multiple sclerosis. AJNR Am J Neuroradiol 33, 1573-1578, doi:10.3174/ajnr.A3086 (2012).

3 Amiri, H. et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin 19, 466-475, doi:10.1016/j.nicl.2018.04.023 (2018).

4 Sdika, M. & Pelletier, D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. Hum Brain Mapp 30, 1060-1067, doi:10.1002/hbm.20566 (2009).

5 Popescu, V. et al. Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. Neuroimage Clin 4, 366-373, doi:10.1016/j.nicl.2014.01.004 (2014).

6 Nakamura, K. & Fisher, E. Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. Neuroimage 44, 769-776, doi:10.1016/j.neuroimage.2008.09.059 (2009).

7 Chard, D. T., Jackson, J. S., Miller, D. H. & Wheeler-Kingshott, C. A. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 32, 223-228, doi:10.1002/jmri.22214 (2010).

8 Battaglini, M., Jenkinson, M. & De Stefano, N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum Brain Mapp 33, 2062-2071, doi:10.1002/hbm.21344 (2012).

9 Gelineau-Morel, R. et al. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. Hum Brain Mapp 33, 2802-2814, doi:10.1002/hbm.21402 (2012).

10 da S Senra Filho, A. C. S., F.H.; dos Santos, A.C.; Murta Junior, L.O. Multiple Sclerosis multimodal lesion simulation tool (MS-MIST). Biomedical Physics & Engineering Express 5 (2019).

11 Weeda, M. M. et al. Comparing lesion segmentation methods in multiple sclerosis: Input from one manually delineated subject is sufficient for accurate lesion segmentation. Neuroimage Clin 24, 102074, doi:10.1016/j.nicl.2019.102074 (2019).

12 Smith, S. M. Fast robust automated brain extraction. Hum Brain Mapp 17, 143-155, doi:10.1002/hbm.10062 (2002).

13 Popescu, V. et al. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. Neuroimage 61, 1484-1494, doi:10.1016/j.neuroimage.2012.03.074 (2012).

14 Zhang, J., Jiang, W., Wang, R. & Wang, L. Brain MR image segmentation with spatial constrained K-mean algorithm and dual-tree complex wavelet transform. J Med Syst 38, 93, doi:10.1007/s10916-014-0093-2 (2014).

15 Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825-841, doi:10.1016/s1053-8119(02)91132-8 (2002).

16 Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. Med Image Anal 5, 143-156, doi:10.1016/s1361-8415(01)00036-6 (2001).

17 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179-194, doi:10.1006/nimg.1998.0395 (1999).

18 Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195-207, doi:10.1006/nimg.1998.0396 (1999).

19 Zheng, W., Chee, M. W. & Zagorodnov, V. Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. Neuroimage 48, 73-83, doi:10.1016/j.neuroimage.2009.06.039 (2009).

20 Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. P. elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 29, 196-205, doi:10.1109/TMI.2009.2035616 (2010).

21 Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1, S208-219, doi:10.1016/j.neuroimage.2004.07.051 (2004).

22 Smith, S. M. et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 17, 479-489, doi:10.1006/nimg.2002.1040 (2002).

23 Amiri, H. et al. Novel imaging phantom for accurate and robust measurement of brain atrophy rates using clinical MRI. Neuroimage Clin 21, 101667, doi:10.1016/j.nicl.2019.101667 (2019).

24 Roberto, C. A. et al. Brain tissue volume changes following weight gain in adults with anorexia nervosa. Int J Eat Disord 44, 406-411, doi:10.1002/eat.20840 (2011).

25 Janowitz, D. et al. Association between waist circumference and gray matter volume in 2344 individuals from two adult community-based samples. Neuroimage 122, 149-157, doi:10.1016/j.neuroimage.2015.07.086 (2015).

26 Guo, C., Ferreira, D., Fink, K., Westman, E. & Granberg, T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. Eur Radiol 29, 1355-1364, doi:10.1007/s00330-018-5710-x (2019).

27 McDonald, W. I. et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 50, 121-127, doi:10.1002/ana.1032 (2001).

CHAPTER 3

Atrophy and lesions in the MS spinal cord

CHAPTER 3.1

Validation of mean upper cervical cord area (MUCCA) measurement techniques in multiple sclerosis (MS): High reproducibility and robustness to lesions, but large software and scanner effects

Merlin M. Weeda
Sander M. Middelkoop
Martijn D. Steenwijk
Marita Daams
Houshang Amiri
Iman Brouwer
Joep Killestein
Bernard M.J. Uitdehaag
Iris Dekker
Carsten Lukas
Barbara Bellenberg
Frederik Barkhof
Petra J.W. Pouwels
Hugo Vrenken

NeuroImage: Clinical 24 (2019)

Abstract

Introduction

Atrophy of the spinal cord is known to occur in multiple sclerosis (MS). The mean upper cervical cord area (MUCCA) can be used to measure this atrophy. Currently, several (semi-)automated methods for MUCCA measurement exist, but validation in clinical magnetic resonance (MR) images is lacking.

Methods

Five methods to measure MUCCA (SCT-PropSeg, SCT-DeepSeg, NeuroQLab, Xinapse JIM and ITK-SNAP) were investigated in a predefined upper cervical cord region. First, within-scanner reproducibility and between-scanner robustness were assessed using intra-class correlation coefficient (ICC) and Dice's similarity index (SI) in scan-rescan 3DT1-weighted images (brain, including cervical spine using a head coil) performed on three 3T MR machines (GE MR750, Philips Ingenuity, Toshiba Vantage Titan) in 21 subjects with MS and 6 healthy controls (dataset A). Second, sensitivity of MUCCA measurement to lesions in the upper cervical cord was assessed with cervical 3D T1-weighted images (3T GE HDxT using a head-neck-spine coil) in 7 subjects with MS without and 14 subjects with MS with cervical lesions (dataset B), using ICC and SI with manual reference segmentations.

Results

In dataset A, MUCCA differed between MR machines ($p < 0.001$) and methods ($p < 0.001$) used, but not between scan sessions. With respect to MUCCA values, Xinapse JIM showed the highest within-scanner reproducibility (ICC absolute agreement = 0.995) while Xinapse JIM and SCT-PropSeg showed the highest between-scanner robustness (ICC consistency = 0.981 and 0.976, respectively). Reproducibility of segmentations between scan sessions was highest in Xinapse JIM and SCT-PropSeg segmentations (median SI ≥ 0.921), with a significant main effect of method ($p < 0.001$), but not of MR machine or subject group. In dataset B, SI with manual outlines did not differ between patients with or without cervical lesions for any of the segmentation methods ($p > 0.176$). However, there was an effect of method for both volumetric and voxel wise agreement of the segmentations (both $p < 0.001$). Highest volumetric and voxel wise agreement was obtained with Xinapse JIM (ICC absolute agreement = 0.940 and median SI = 0.962).

Conclusion

Although MUCCA is highly reproducible within a scanner for each individual measurement method, MUCCA differs between scanners and between methods. Cervical cord lesions do not affect MUCCA measurement performance.

Introduction

Multiple sclerosis (MS) is a demyelinating and neurodegenerative disease of the central nervous system (CNS). Abnormalities in the spinal cord such as lesions and atrophy often manifest early in the disease course and have been shown to be important predictors of disease progression and prognosis.[1,2] Relevance of mean upper cervical cord area (MUCCA) from magnetic resonance imaging (MRI) has been shown in early as well as late stages of MS[3-6].

Since manual MUCCA measurements are labor-intensive and can suffer from large intra- and inter-rater variability[7-10], several (semi-)automated methods have been developed, such as SCT-PropSeg[10], SCT-DeepSeg[11], NeuroQLab[12], Xinapse JIM[13] and ITK-SNAP[14]. However, these methods are generally developed for dedicated cord imaging instead of head imaging, and although previous research has shown that it is possible to obtain accurate MUCCA measures not only from cord imaging, but from brain imaging as well[15,16], research regarding their reproducibility and robustness in whole brain images is not yet available. This knowledge is needed to incorporate MUCCA measurement from head images in standardized clinical care, and to facilitate research on cervical cord atrophy in MS, specifically to analyze MUCCA retrospectively in data in which 3D T1-weighted cervical cord images are not available as well as prospectively without the need for separate spinal cord imaging.

Moreover, subjects with MS often exhibit lesions in the upper cervical cord[17]. Since lesions in the brain are known to severely affect brain atrophy measurements[18-23], it is important to investigate the effect of lesions in spinal cord on MUCCA measurements as well. Although a study using SCT PropSeg did not observe an obvious effect of lesion on the spinal cord segmentation through visual inspection[24], a quantitative assessment of the effect of lesions on the quality of cervical spinal cord segmentation and MUCCA measurement is lacking.

Therefore, the aim of this study was twofold: (a) to assess the reproducibility and robustness of these (semi-)automatic spinal cord segmentation methods in whole-brain 3D T1-weighted images by measuring MUCCA in scan and rescan images (reproducibility) acquired on three different MR machines (robustness); and (b) to quantitatively investigate whether the presence of lesions in the cervical spinal cord affects the performance of these segmentation methods.

Methods

Subjects

The institutional review board approved the study protocols and written informed consent was obtained from all individuals, according to the Declaration of Helsinki.

For this study, two different existing datasets were used, which are hereafter referred to as dataset A for reproducibility and robustness and dataset B for the effect of lesions. For both datasets, subjects with MS according to McDonald 2010 criteria[25] were included and were allowed to use disease modifying treatment. All subjects were enrolled at the same institution. Dataset A consisted of 6 healthy controls and 21 subjects with MS (relapsing remitting MS n=16; secondary progressive MS n=1; and primary progressive MS n=4). All subjects underwent two sessions of MRI examinations (hereafter defined as ‘scan’ and ‘rescan’) on three 3T MR

machines from different vendors (General Electric [GE], Philips and Toshiba) in the same center. The scan and rescan sessions within one MR machine were always performed on the same day and the different MR machine examinations were all performed preferably over the course of a single day, or within a maximum of eight days from each other.

Dataset B consisted of 21 subjects with RRMS selected from a larger cohort of 196 patients with relatively long disease duration[26,27] who were scanned on a 3T GE scanner. Selection of these subjects was based on the amount of cervical cord MS lesions as previously reported[26], and subjects were divided into two groups: with (at least 11 counted) lesions in the cervical cord (*n*=14) and without lesions (*n*=7). The two groups were balanced for gender, mean EDSS as well as for mean MUCCA values as obtained previously in the C1-C2 area using an older version of NeuroQLab[26] (see **Inline Supplementary Table 1**).

Inline Supplementary Table 1 – Subjects from dataset B, selected from a previous study (Daams *et al*., 2014).

	Without lesions (*n*=7)	**With lesions** (*n*=14)
Age (years, mean ± SD)	53.32 ± 11.76	49.74 ± 8.86
Sex (m/f, %f)	2/5 (40%)	4/10 (40%)
EDSS score (median, Q1-Q3)	4.0 (3.5 – 4.5)	4.0 (3.0 – 7.0)
MUCCA[#] (mm^2, mean ± SD)	74.43 ± 10.23	72.64 ± 10.18

[#] MUCCA as measured by NeuroQLab in the study from Daams *et al*., 2014.

MRI examination

The subjects in dataset A underwent MRI examinations on three different 3T whole body MRI scanners with a head coil, all including a sagittal 3D T1-weighted sequence: (1) GE Discovery MR750 (GE Healthcare, USA) with a fast spoiled gradient echo sequence (FSPGR with TR/TE/TI = 8.2/3.2/450 ms and resolution 1.0x1.0x1.0 mm); (2) Philips Ingenuity with a turbo field echo sequence (TFE with TR/TE/TI = 7.9/4.5/900 ms and resolution 1.0x1.0x1.0 mm); and (3) Toshiba Vantage Titan with a fast field echo sequence (FFE with TR/TE/TI = 5.7/2.4/1050 ms and resolution 1.0x1.0x1.2 mm).

Subjects from dataset B underwent MRI examination as described earlier[26]. In summary, subjects were scanned using a 3T HDxt GE scanner (GE Healthcare, USA) with a head-neck-spine coil with a sagittal 3D T1-weighted FSPGR sequence (TR/TE/TI = 7.3/3.0/450 ms with acquired resolution 1.09x1.09x1.0 mm, reconstructed to 0.55x0.55x1.0 mm). The aforementioned lesion count was performed in a previous study[26] based on a separate cervical 2D PD/T2-weighted image (TR/TE = 6,200/21-84 ms, resolution 0.57x0.57x4.0 mm) covering the entire cervical cord.

MR image analysis

Preprocessing

To correct for gradient nonlinearity effects on the MUCCA measurement[28], all images from dataset A were corrected with 3D distortion correction available on each scanner, and images from dataset B were corrected off-line using the grad_unwarp software[29].
All T1-weighted images were bias field corrected using the default options from the segmentation tool FAST from the FSL Toolbox[30].

Region selection

In order to perform an objective, unbiased comparison, all methods were evaluated within the same region, which was pre-defined in each subject by selecting a fixed set of contiguous axial slices starting from the most superior point of C1 and ending at a position 30 mm more inferior, where the 30 mm length was measured perpendicular to the axial plane. For Xinapse JIM, this entailed definition of manually selected input points, as described below under “Xinapse JIM”. For all methods, it entailed creating a subselection of the segmentations produced by accepting only the portion of the segmentations that fell within these pre-defined sets of contiguous slices, as described below under “Post-processing”.

Spinal cord segmentation

For the segmentation of the spinal cord, five (semi-)automated methods were used: SCT-PropSeg, SCT-DeepSeg, NeuroQLab, Xinapse JIM, and ITK-SNAP. In addition, a manual segmentation was created for dataset B to obtain a ground truth segmentation. The segmentation methods are summarized below.

SCT-PropSeg

SCT-PropSeg[10] is a fully automated spinal cord segmentation method incorporated in the Spinal Cord Toolbox (SCT version 3.0.8)[31]. It has a two-step working mechanism. First, spinal cord detection is done by maximizing mutual information of the left and right part of an axial slice and finding the central line. The image is cropped in an area of 5 cm around this medial line, and a Hough transform is applied assuming an approximate guess of the spinal cord radius of 4 mm (default). These steps are then performed on multiple axial slices, after which the results are validated based on the contrast between cerebrospinal fluid and the spinal cord. After spinal cord detection, propagation of the spinal cord segmentation is started and conducted using a mesh deformation model based on maximizing the local contrast gradient[10].

SCT-DeepSeg

SCT-DeepSeg is another fully automated spinal cord segmentation method incorporated in the Spinal Cord Toolbox (SCT version 3.1.1)[11,31]. It is based on a deep learning convolutional neural network (CNN) module trained to effectively segment the spinal cord from MRI images. The version of the method used in this paper operates in a 2D fashion, treating each slice separately.

NeuroQLab

NeuroQLab (MeVisLab, Fraunhofer Mevis, Bremen, Germany) also provides a semi-automated method for cervical cord segmentation[32]. For this, the user selects a cuboid ROI manually, after which interactive watershed transformation (IWT) is applied to the image, removing non-CNS matter. IWT results in over-inclusive spinal cord segmentation, therefore fully automated regional histogram analysis is performed to accurately quantify the spinal cord tissue volume. NeuroQLab only provides MUCCA values but does not give voxel wise labeled segmentation files as output, and therefore no overlap measures could be computed for this method (see section 2.5.3 and 2.6).

Xinapse JIM

Xinapse JIM[33] contains (semi-)automated software for spinal cord segmentation based on the 2D active surface contour method[13]. For this, the Cord Finder tool was used with fixed parameters for all datasets (number of shape coefficients: 24; order of longitudinal variation: 10; nominal cord diameter: 8 mm). In this tool, the centerline of the spinal cord is defined semi-automatically, followed by automatic determination of the cord contour on each axial slice. As indicated above ("Region selection"), the cervical cord was segmented in a section of 30 slices (1 mm slice thickness) starting at the top of the C1 vertebra. The mean cervical spinal cord area (CSA) was obtained by dividing the generated cervical cord volume by the section length (30 mm). For further comparison with the other evaluation methods, binary mask images were generated from the segmented cord sections of each individual. For this purpose the segmented cord outlines of each slice were saved as region-of-interest files and converted into binary masks using the Masker-tool provided in the JIM software package.

ITK-SNAP

ITK-SNAP is a semi-automated method based on active contour models which can be applied to segment any image, and is not specialized for the spinal cord (ITK-SNAP version 3.6.0)[14]. For this study, segmentation was based on region competition, which requires user input to provide the intensity window on which this competition should be based. Then, a 3D snake is propagated with a velocity based on the manually selected intensity window. The snake propagation is both initiated and terminated by the user.

Manual segmentation

For dataset B, manual outlining was needed to establish a ground truth segmentation, since no scan-rescan images were available. Manual segmentation was performed on the 3D FSPGR images in ITK-SNAP (version 3.6.0)[14] by a single rater by in-painting the axial slices on a slice-by-slice basis, resulting in a binary segmentation image. Five scans were segmented twice on two separate occasions to assess intra-rater variability and Dice's similarity index (SI). Manual segmentation was performed from the most superior slice of C1 and continued for 56 slices (30 mm).

Post-processing

All methods resulted in a binary spinal cord segmentation, except for NeuroQLab, which only gives MUCCA measurements as output. The output image of each of the other methods were post-processed as follows: for each segmentation, a mask was manually created to select a subset of the axial slices starting from the most superior point of C1 and ending at a position 30 mm more inferior For dataset A 31 axial (reformatted) slices with a 1.0 mm slice thickness were selected and for dataset B 56 axial (reformatted) slices with a 0.55 mm slice thickness, thereby covering the cervical cord from section C1 up to section C2, depending on the subject's orientation in the coil[34]. This mask was applied to the full binary segmentations in order to obtain a segmentation image of the relevant area in the upper cervical cord and to ensure a direct comparison between the different segmentation methods All subsequent data-analysis was performed on these post-processed segmentation images.

MUCCA

The "SCT Process Segmentation" routine included in the SCT Toolbox[31] was used to obtain actual values for MUCCA for all segmentations. The post-processed segmentation images from all methods (except NeuroQLab) were used as input, from which SCT Process Segmentation calculates the total spinal cord area perpendicular to the centerline for each slice, and then averages this over the length of the section to obtain MUCCA. In this way, the orientation of the cord with respect to the slices is taken into account when calculating MUCCA, thereby diminishing the effects of both slice orientation and cord curvature on MUCCA values.

Dataset A: reproducibility and robustness

Reproducibility: within-scanner intra-class correlation coefficient (ICC)

To evaluate the reproducibility of the various spinal cord segmentation methods, we calculated the within-scanner ICCs for absolute agreement (ICC_{abs}) with their 95% confidence intervals between the scan and rescan images. Furthermore, we calculated the coefficient of variation (COV) as the ratio of the standard deviation of within-scanner differences to the mean MUCCA (scan and rescan) as a measure of dispersion.

Robustness: between-scanner intra-class correlation coefficient (ICC)

To evaluate the robustness of the various spinal cord segmentation methods, we calculated the between-scanner ICCs for absolute agreement and consistency (ICC_{con}) with their 95% confidence interval. For the between-scanner ICCs, only the images of the first scan session were used.

Voxel wise agreement: Dice's similarity index (SI)

Rescan images were linearly registered to the scan images using default FSL FLIRT with an affine transformation (12 parameters), correlation ratio cost function, and tri-linear interpolation followed by a threshold of 0.5[35,36], after which fslmaths and fslstats were used to calculate Dice's similarity index (SI) between scan and rescan segmentations.

Dataset B: effect of lesions

Intra-rater variability of the manual segmentations was assessed by calculating SI and COV. Next, MUCCA was measured as described above and ICCs for absolute agreement and consistency were calculated between MUCCA values from manual segmentations and each of the (semi-)automated methods. Furthermore, SI was calculated between the manual and automated segmentations using the same pipeline described above.

Statistical analyses

Statistical analysis was performed in IBM SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, N.Y., USA). All parameters were tested for normality with a Shapiro-Wilk test. In dataset A, repeated measures ANOVA (parametric data) and Friedman Test (non-parametric data) were used for MUCCA per scan session, segmentation method, MR machine and subject group (HC vs MS); and for SI per segmentation method, MR machine and subject group.
In dataset B, repeated measures ANOVA (parametric data) and Friedman Test (non-parametric data) were used for MUCCA per segmentation method and subject group (with or without lesions); and for SI per segmentation method, MR machine and subject group.
For the repeated measures ANOVA, Mauchly's test of sphericity was performed to assess equal variances of the differences between all within-subject factors. When the assumption of sphericity was violated, degrees of freedom were corrected using Huyn-Feldt estimates of sphericity.
When appropriate, post-hoc analyses were conducted using Mann-Whitney U tests (unpaired) or Wilcoxon Signed Ranks tests (paired) for single effects, and Bonferroni correction for interaction effects. Inter quartile range was determined by the 25th and 75th percentile. Results were considered statistically significant upon p-value < 0.05.

Results

Dataset A: reproducibility and robustness

MUCCA across scan sessions, MR machines, and methods

An example of the various spinal cord segmentation methods is shown in **Figure 1**. The segmented MUCCA is depicted in mm^2, showing great differences between the methods in this particular example subject (NB. images from a patient at the first scan session on the GE MR machine). **Figure 2** and **Supplementary Table 1** show the mean MUCCA values obtained in all subjects, separated by scan session, MR machine and upper cervical cord segmentation method. Here, mean MUCCA also clearly varied between methods and between MR machine (e.g. mean MUCCA in GE varied depending on the method from 49.50 to 73.31 mm^2; and mean MUCCA in SCT-PropSeg varied depending on the MR machine from 66.49 to 70.55 mm^2), as shown by the repeated measures ANOVA for MUCCA that found a significant interaction between MR machine and method (F(8,200) = 3.804, p = 0.025). As expected, no effect of scan session was found (F(1,25) = 0.972, p = 0.334), indicating that scan and rescan MUCCA values did not differ systematically. There was no significant effect of subject group (F(1,25) = 2.756, p = 0.109).

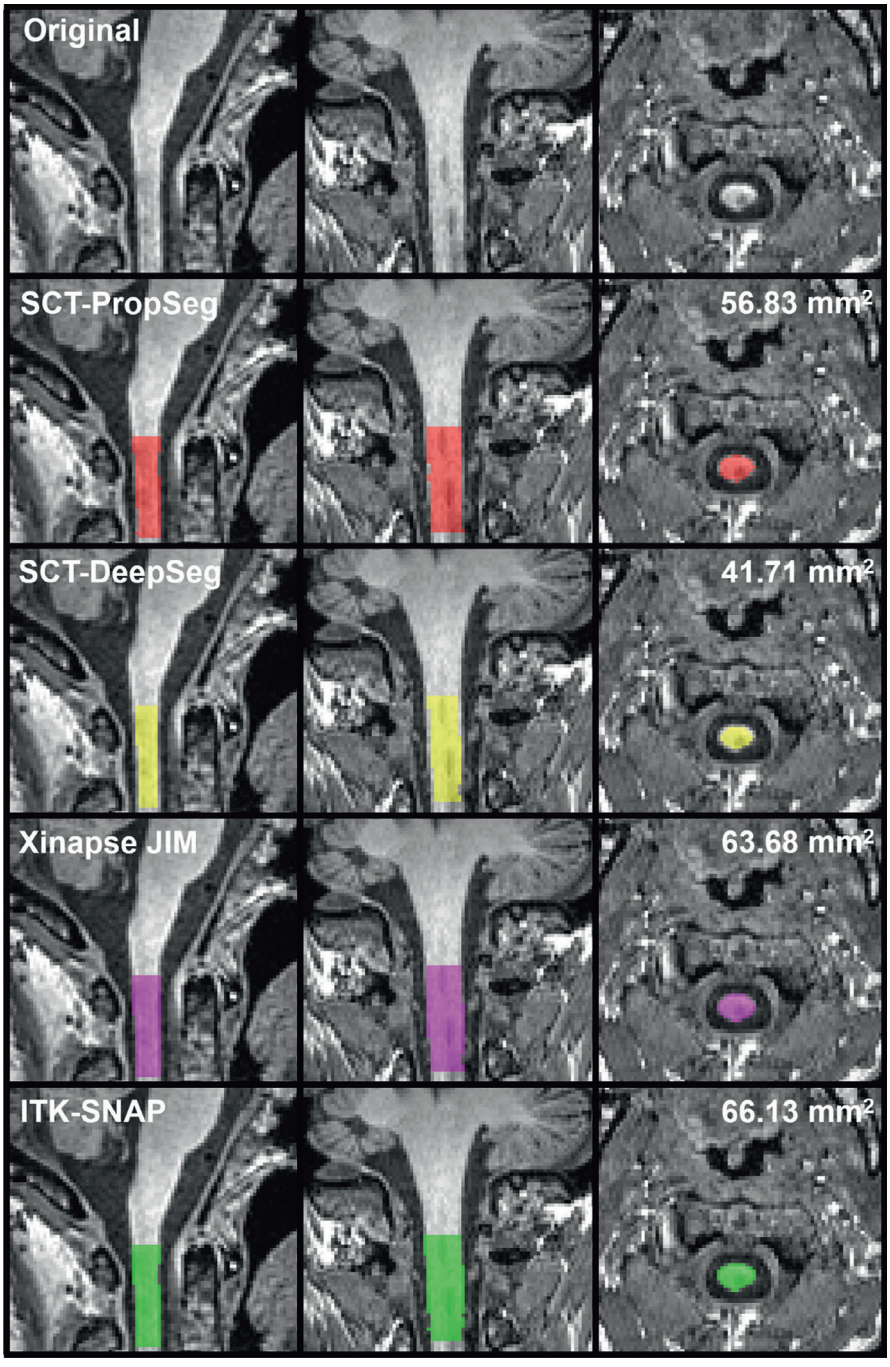

Figure 1 – Example of the spinal cord segmentation methods; scans obtained from a 52 year old female with RRMS in the first scan session on the GE machine.

In this particular case, MUCCA differs between the methods from 41.71 mm^2 obtained from SCT-DeepSeg (yellow) to 66.13 mm^2 obtained from ITK-SNAP (green).

Because of the interaction effect of MR machine and method, a post-hoc analysis was performed separately for the effect of MR machine per method and the effect of method per MR machine. For all methods, MUCCA differed between GE and Philips (all $p \leq 0.001$) and between GE and Toshiba (all $p \leq 0.001$). However, between Philips and Toshiba images, only SCT-DeepSeg MUCCA were different ($p < 0.001$), but not MUCCA obtained from the other methods.

All pairwise MUCCA differences between methods were significant for all MR machines (all $p \leq 0.014$), except for the comparison between NeuroQLab and Xinapse JIM MUCCA for Philips and for Toshiba.

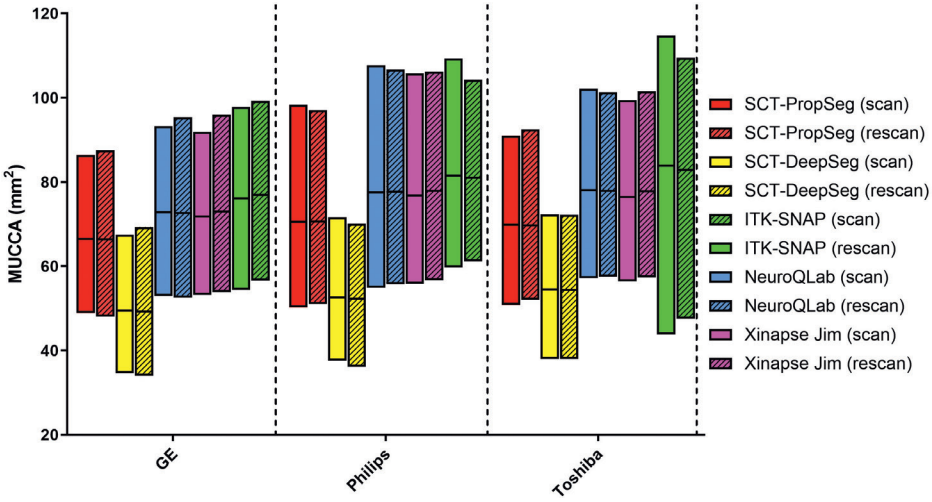

Figure 2 – Interleaved low-high floating bar plot (line at mean) of MUCCA (mm^2) from all subjects (i.e. HC and MS grouped) per MR machine (GE [left], Philips [middle], Toshiba [right]), per segmentation method (SCT-PropSeg [red], SCT-DeepSeg [yellow], NeuroQLab [blue], Xinapse JIM [pink] and ITK-SNAP [green]) and per scan session (scan [clear], rescan [striped]).

Pairwise differences can be seen between segmentation methods and between MR machines, but not between scan sessions.

Reproducibility and robustness: ICC and COV

The within-scanner agreement (i.e. reproducibility) assessed by means of the ICC_{abs} and COV, and the between-scanner agreement (i.e. robustness) assessed by means of the ICC_{con}, are shown per segmentation method and per MR machine in **Table 1**. In general, within-scanner agreement was high ($ICC_{abs} \geq 0.904$ and COV ≤ 5.03%) with the most reproducible results obtained in Toshiba images with NeuroQLab segmentations ($ICC_{abs} = 0.996$ and COV = 0.97%) and Xinapse JIM segmentations ($ICC_{abs} = 0.996$ and COV = 0.88%).

Between-scanner agreement was highest between Philips and Toshiba images (ICC_{con} range from 0.827 and 0.984). The most robust segmentation method was Xinapse JIM (ICC_{con} range from 0.978 to 0.982). However, the most robust results were obtained in SCT-PropSeg segmentations between GE and Philips images ($ICC_{con} = 0.985$).

Table 1 – Reproducibility (i.e. within-scanner agreement) and robustness (i.e. between-scanner agreement) of the different MR machines and different methods.

		SCT-PropSeg	SCT-DeepSeg	NeuroQLab	Xinapse JIM	ITK-SNAP
Within-scanner agreement						
GE	ICC_{abs} [95% CI]	0.994 [0.986, 0.997]	0.994 [0.987, 0.997]	0.995 [0.989, 0.998]	0.995 [0.989, 0.998]	0.954 [0.903, 0.979]
	COV	1.20	1.38	1.06	1.00	3.28
Philips	ICC_{abs} [95% CI]	0.995 [0.989, 0.998]	0.988 [0.973, 0.994]	0.983 [0.963, 0.992]	0.995 [0.988, 0.998]	0.919 [0.831, 0.962]
	COV	1.10	1.86	2.03	1.11	4.35
Toshiba	ICC_{abs} [95% CI]	0.994 [0.988, 0.997]	0.990 [0.978, 0.995]	0.996 [0.991, 0.998]	0.996 [0.992, 0.998]	0.904 [0.803, 0.955]
	COV	1.15	1.64	0.97	0.88	5.03
Between-scanner agreement						
GE vs Philips	ICC_{con} [95% CI]	0.970 [0.934, 0.986]	0.985 [0.967, 0.993]	0.971 [0.938, 0.987]	0.978 [0.951, 0.990]	0.905 [0.804, 0.956]
GE vs Toshiba	ICC_{con} [95% CI]	0.983 [0.964, 0.992]	0.977 [0.950, 0.989]	0.980 [0.956, 0.991]	0.982 [0.962, 0.992]	0.882 [0.758, 0.944]
Philips vs Toshiba	ICC_{con} [95% CI]	0.982 [0.961, 0.992]	0.976 [0.948, 0.989]	0.984 [0.966, 0.993]	0.982 [0.961, 0.992]	0.827 [0.657, 0.917]

Abbreviations: ICC_{abs} = intraclass correlation coefficient, within-scanner absolute agreement; COV = coefficient of variance; ICC_{con} = intraclass correlation coefficient, between-scanner consistency; CI = confidence interval.

Anatomical reproducibility: Dice's similarity index

The anatomical reproducibility of the segmentation methods, assessed as Dice's similarity index (SI) between segmentations from scan and rescan images, is shown in **Table 2**. Since NeuroQLab does not provide segmentation images but solely gives MUCCA values as output, no overlap between scan and rescan could be calculated for this method. SI was generally high, with median SI in all cases above 0.910, but varied significantly between methods (Friedman Test: $\chi^2(3) = 87.504$, $p < 0.001$). SI did not vary systematically with MR machine ($\chi^2(2) = 1.352$, $p = 0.509$) or subject group (Mann Whitney U test $p \geq 0.405$). Post-hoc analysis showed differences between all combinations of methods (all $p \leq 0.001$), except between SCT-PropSeg and Xinapse JIM.

Table 2 – Dice's similarity index between scan and rescan images for the three MR machines and four segmentation methods.

Dice's similarity index	SCT-PropSeg	SCT-DeepSeg	Xinapse JIM	ITK-SNAP
GE	0.927 (0.909 – 0.946)	0.910 (0.894 – 0.928)	0.928 (0.906 – 0.941)	0.925 (0.898 – 0.937)
Philips	0.923 (0.900 – 0.944)	0.911 (0.891 – 0.939)	0.921 (0.898 – 0.943)	0.916 (0.891 – 0.939)
Toshiba	0.922 (0.901 – 0.939)	0.922 (0.894 – 0.923)	0.929 (0.905 – 0.939)	0.920 (0.897 – 0.931)

Legend: SI listed as median with interquartile range (Q1-Q3); because NeuroQLab does not provide segmentation images, no SI could be calculated for NeuroQLab.

Dataset B: effect of lesions

Lesions found in C1-C2 region

In dataset B, we compared subjects with and without lesions in the spinal cord, based on the study from Daams *et al.,* (2014). Since we measure MUCCA only at C1-C2 level, we checked the subjects that were selected in the lesions group for the actual presence of lesions at the C1-C2 level, i.e. in our spinal cord mask. Per subject, at least 3 lesions were found in our region of interest with a median of 5 lesions per subject (Q1-Q3: 3.75 – 5.50 lesions).

Manual segmentations

Quality of the manual segmentations was assessed through intra-rater reproducibility in five cases that were performed twice in dataset B, which showed high intra-rater agreement with a mean SI of 0.966 ± 0.005 (range 0.957-0.971) and an overall COV of 3.65%.

MUCCA across lesion groups and methods

Mean MUCCA values by lesion group and by method are provided in **Figure 3** and **Supplementary Table 2**. No effect of lesion group on MUCCA was found ($F(1,19) = 0.925$, $p = 0.348$), which is consistent with the matching of groups on previously determined MUCCA values as per the study design.

Mean MUCCA varied between methods in both lesion groups, e.g. ranging from 53.36 ± 4.51 mm² in SCT-PropSeg to 80.23 ± 9.39 mm² in manual segmentations in the no lesions group. The effect of method on MUCCA was significant (repeated measures ANOVA $F(5,95) = 260.036$, $p < 0.001$). Post-hoc analysis showed that these differences between methods were present for almost all pairwise comparisons ($p \leq 0.022$). Methods that did not show significant different MUCCA values from one another were: (a) SCT-PropSeg versus SCT-DeepSeg; (b) NeuroQLab versus manual or Xinapse JIM; (c) ITK-SNAP versus manual or NeuroQLab or Xinapse JIM.

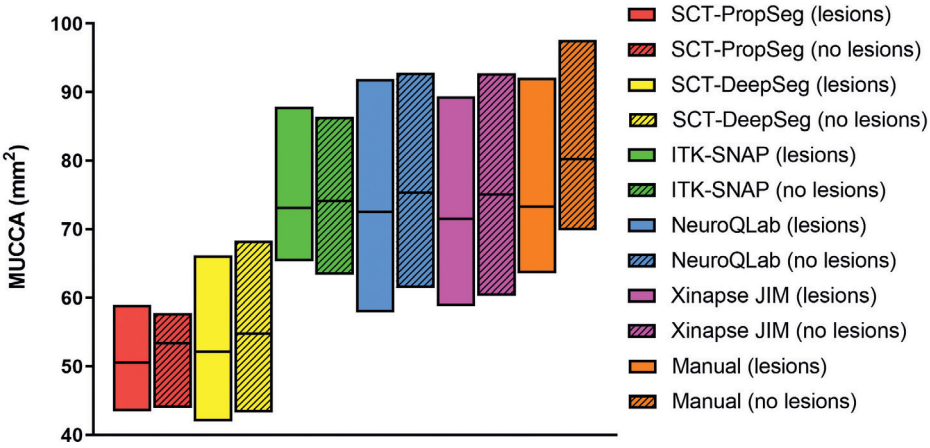

Figure 3 – Interleaved low-high floating bar plot (line at mean) showing MUCCA (mm²) in subjects with with lesions (clear) and without lesions (striped) per segmentation method (SCT-PropSeg [red], SCT-DeepSeg [yellow], NeuroQLab [blue], Xinapse JIM [pink], ITK-SNAP [green] and Manual [orange]).
Differences can be seen between segmentation methods.

Volumetric agreement: ICC

Table 3 shows the volumetric agreement assessed by ICC with manual MUCCA per segmentation method and per lesion group. The differences between manual and some automated MUCCA measurements reported above (**Figure 3**) are reflected by low values for ICC_{abs}, e.g. 0.075 (SCT-PropSeg, group without lesions) or 0.184 (SCT-DeepSeg, group with lesions). Compared to manual, the best volumetric agreement was obtained by NeuroQLab in the group with lesions ($ICC_{abs} = 0.846$), and by Xinapse JIM in the group without lesions ($ICC_{abs} = 0.940$). ICC_{con} values for the automated methods versus manual MUCCA were generally higher and all > 0.57.

Voxel wise agreement: SI

The voxel wise agreement compared to manual in the two lesion groups is shown for the different methods in **Table 4**, showing that mean SI varies between methods and between lesion groups (range: 0.791 to 0.962). SI was generally high, with all SI above 0.79, but varied significantly between methods (Friedman Test: $\chi^2(3) = 51.229$, $p < 0.001$). Post-hoc analysis showed differences between all combinations of methods (all $p \leq 0.033$), except between

ITK-SNAP and Xinapse JIM. Wilcoxon Signed Ranks test found no effect of lesion group in any of the methods ($p \geq 0.176$).

Table 3 – Volumetric agreement of the different methods with manual MUCCA per lesion group.

Within-scanner agreement		SCT-PropSeg	SCT-DeepSeg	Neuro-QLab	Xinapse JIM	ITK-SNAP
Without lesions	ICC_{abs} [95% CI]	0.106 [-0.023, 0.427]	0.181 [-0.013, 0.582]	0.931 [0.804, 0.977]	0.940 [0.774, 0.982]	0.883 [0.674, 961]
	ICC_{con} [95% CI]	0.707 [0.304, 0.896]	0.894 [0.703, 0.965]	0.930 [0.797, 0.977]	0.954 [0.863, 0.985]	0.876 [0.658, 0.958]
With lesions	ICC_{abs} [95% CI]	0.075 [-0.028, 0.438]	0.184 [-0.006, 0.665]	0.846 [-0.032, 0.976]	0.837 [0.031, 0.973]	0.577 [-0.089, 0.907]
	ICC_{con} [95% CI]	0.572 [-0.226, 0.911]	0.946 [0.720, 0.990]	0.947 [0.725, 0.991]	0.924 [0.628, 0.987]	0.685 [-0.042, 0.938]

Abbreviations: ICC_{abs} = intraclass correlation coefficient, within-scanner absolute agreement; ICC_{con} = intraclass correlation coefficient, between-scanner consistency; CI = confidence interval.

Table 4 – Dice's similarity index between manual and automated segmentation methods in the lesion groups.

Dice's similarity index	SCT-PropSeg	SCT-DeepSeg	Xinapse JIM	ITK-SNAP
Without lesions	0.782 (0.756 – 0.823)	0.824 (0.789 – 0.829)	0.956 (0.941 – 0.968)	0.955 (0.904 – 0.971)
With lesions	0.818 (0.793 – 0.833)	0.835 (0.811 – 0.849)	0.962 (0.959 – 0.968)	0.959 (0.953 – 0.964)

Legend: SI listed as median with interquartile range (Q1-Q3).

Discussion

In this study, we investigated the performance of five (semi-)automated spinal cord segmentation methods to measure MUCCA in brain images by quantifying within-scanner reproducibility, between-scanner robustness, and performance in the presence of cervical cord lesions. Within-scanner volumetric and anatomical reproducibility of MUCCA were high, but MUCCA varied between scanners and segmentation methods. Interestingly, the volumetric and voxel wise agreement of the MUCCA measurements did not differ between subjects with or without upper cervical cord lesions.

Despite our small sample size in dataset A, we found significant differences between MUCCA from the acquisition protocols of GE, Philips and Toshiba. In general, MUCCA measured in GE images was lower than images obtained in Philips and Toshiba images. It is unlikely that spatial resolution had a major influence, since this was slightly different only on Toshiba. We can only speculate that small variability in contrast between spinal cord and surrounding CSF, arising from acquisition differences, especially in timing parameters, could lead to differences in partial volume effects at the border of the spinal cord. Previous research in which acquisition was

homogenized between centers still showed low between-scanner agreement[28,37], supporting our finding that cross-sectional MUCCA cannot be easily compared in multi-center, multi-vendor research. This emphasizes the importance of relative instead of absolute MUCCA comparisons in multi-vendor studies, i.e. with normalized UCCA percent change over time as a measure for upper cervical cord atrophy[38,39].

Next to differences in MUCCA between MR machines, we also found significant differences in MUCCA between methods. In both dataset A and B, SCT-PropSeg and SCT-DeepSeg segmentations resulted in lower MUCCA than for other methods, which was also shown in an earlier study[24]. This under estimation may be due to the use of the head and head-neck coil instead of a spine coil, for which the Spinal Cord Toolbox is optimized[10,31]. Differences in spinal cord to CSF contrast-to-noise ratio may affect these methods. Previous research did show that MUCCA values obtained with either head or head-neck images are comparable for both NeuroQLab and XinapseJIM[15,16], but the effect of the coil (head, head-neck, spine) used should be further investigated for the other spinal cord segmentation methods. In addition to these volumetric differences, the amount of manual labor needed differed between methods as well. Dependent on the study type and the amount of data, it is important to take these differences in robustness into account in order to assure equal comparisons between subjects; for example, ITK-SNAP is less suited for larger studies due to the manual labor required, and NeuroQLab may be less suited when segmentation images are preferred in addition to MUCCA values only. In the present study we did not systematically optimize parameters for the automated methods; it is possible that performance of the automated methods could be improved by adjusting parameters for each dataset, which should be investigated further.

With our unique scan-rescan research design, we were able to show high volumetric and voxel wise agreement between scan sessions for all method-scanner combinations, with the exception of NeuroQLab, since it only provides MUCCA output and no segmentation images. This high anatomical reproducibility shows that all methods are robust to changes in cervical cord orientation that occur in a scan/rescan setting, which has not been reported previously. This suggests that MUCCA measurement from head images is suitable for single-patient monitoring in the clinic, as long as subjects are examined on the same MR scanner and MUCCA measurement is performed with the same method, which may lead to more accurate predictions of disability and possible disease course[3-6].

It would be interesting to study if differences between processing methods project to meaningful sample-size differences in clinical studies. While within-scanner reproducibility was high for each of the cross-sectional methods investigated here, there were some differences in COV (**Table 1**). It remains to be investigated how this translates to sensitivity of longitudinal MUCCA evaluations. A comparable scan-rescan investigation of both inherently longitudinal and repeated cross-sectional methods in a cohort with long enough follow-up to expect sufficient MUCCA decrease, will allow quantification of the reproducibility of the methods, their sensitivity to MUCCA change, and the required sample sizes in a clinical trial setting with MUCCA change as outcome.

Another point of attention for longitudinal MUCCA measurement is the ROI selection. In this study, since we were interested in the differences between segmentation methods,

we pre-selected the upper cervical cord area as a mask, starting from the top of C1 and continuing for 30 mm. Although this fixed length may lead to minor differences between subjects concerning the portion of the cervical cord selected, in the present study this allowed unbiased comparison of segmentation methods. However, to achieve clinical implementation of MUCCA measurement, automated selection of this region is needed in order to ensure MUCCA measurement in equal segments of the cord over time. Future work should therefore further validate methods such as SCT_process_segmentation for the automated selection of the desired cord section, especially in head images, to allow high-throughput (clinical) processing in a more automated fashion.

In subjects with MS, an important question is whether the presence of lesions in the upper cervical cord influences MUCCA measurement performance, since previous studies have found that the brain's grey matter volume may be underestimated in the presence of lesions[18]. Extending on previous research, which found that SCT-PropSeg segmentations were not influenced by lesions upon visual inspection[24], we observed that the presence of lesions in the upper cervical cord did not affect MUCCA measurement performance, neither on a volumetric (ICC) nor on a voxelwise (SI) level for any of the investigated MUCCA measurement methods. This is an important finding, especially in view of the potential clinical application of MUCCA measurement in MS, as it implies that no additional measures are required and existing software can be applied as is if overall MUCCA is the desired outcome, in contrast to brain MRI. Since no specific analyses of lesion volume, lesion location or T1-weighted lesion intensity was performed here, all of which are known to severely influence brain MRI segmentations, future research for their effect on MUCCA in the different software packages should be performed for more anatomical detailed analysis of cervical cord atrophy in MS

An important topic of research is how lesions and atrophy of the spinal cord develop in MS and how they are related to each other. Valsasina and colleagues, using Xinapse JIM, did not find a relation between upper cervical cord lesions and MUCCA[40], but did not investigate whether the presence of cervical cord lesions affected their MUCCA measurement. Our current results show that accurate MUCCA measurement is possible in the presence of cervical cord lesions, for Xinapse JIM as well as for the other methods, thereby reinforcing the finding reported by Valsasina et al. on the independence of the two disease phenomena. In the current study, we could not investigate such disease-related questions directly, because we matched subjects for their previously obtained MUCCA values.[26] By design, we were therefore unable to answer any disease-related questions about MUCCA or its relation to cervical lesions. Nevertheless this is an interesting topic that warrants further investigation in future research, for which the current study lays the foundation by demonstrating that measurement of MUCCA can be performed without an effect of cervical cord lesions for any of the investigated methods.

In conclusion, the choice of automated spinal cord segmentation method has a large effect on MUCCA measurement, as well as the type of MR machine used. However, all methods show high within-scanner agreement between scan and rescan session, both for volumetric and voxel wise MUCCA measures. Most importantly, performance of the MUCCA software tested was not affected by the presence of upper cervical cord lesions.

Funding

This study is supported by the Dutch MS Research Foundation (grant 14-876 MS); Amsterdam Neuroscience (grant PoC-2014-BIT-3); Novartis Pharma (grant SP037.15/432282); and the German Federal Ministry for Education and Research, BMBF, German Competence Network Multiple Sclerosis KKNMS (grant 01GI1601I and 01GI0914). FB is supported by the NIHR Biomedical Research Centre at UCLH.

Declaration of interest

The authors declare that there is no conflict of interest.

Acknowledgements

The authors thank Prof. Horst Hahn and Dr. Florian Weiler of Fraunhofer Mevis for supplying a free license to NeuroQLab for this study.

Supplements

Supplementary Table 1 – MUCCA from dataset A per method and per MR machine, shown for both scan and rescan sessions .

MUCCA	SCT-PropSeg		SCT-DeepSeg		NeuroQLab		Xinapse JIM		ITK-SNAP	
	Scan	Rescan	Scan	Rescan	Scan	Rescan	Scan	Rescan	Scan	Rescan
GE	66.49 ± 9.78	66.33 ± 10.10	49.50 ± 8.57	49.22 ± 8.99	72.81 ± 10.71	72.66 ± 11.04	73.31 ± 10.32	73.04 ± 10.69	72.81 ± 10.71	76.08 ± 11.38
Philips	70.55 ± 11.22	70.59 ± 11.11	52.57 ± 8.73	52.25 ± 8.73	77.59 ± 12.33	77.73 ± 12.01	78.31 ± 11.75	77.87 ± 11.85	77.59 ± 12.33	81.53 ± 12.86
Toshiba	69.87 ± 10.66	69.70 ± 10.51	54.49 ± 9.20	54.38 ± 8.73	78.04 ± 12.04	77.90 ± 11.71	77.89 ± 11.49	77.83 ± 11.33	78.04 ± 12.04	83.92 ± 14.21

MUCCA shown as mean (mm^2) ± standard deviation.

Supplementary Table 2 – MUCCA from dataset B per method and per lesion group.

MUCCA	Manual	SCT-PropSeg	SCT-DeepSeg	Neuro-QLab	Xinapse JIM	ITK-SNAP
Without lesions	80.23 ± 9.39	53.36 ± 4.51	54.79 ± 8.22	75.30 ± 10.20	75.33 ± 10.96	74.13 ± 8.49
With lesions	73.31 ± 8.49	50.52 ± 4.37	52.12 ± 6.43	72.51 ± 8.61	71.69 ± 8.29	73.12 ± 7.72

MUCCA shown as mean (mm^2) ± standard deviation.

References

1 Kearney, H., Miller, D. H. & Ciccarelli, O. Spinal cord MRI in multiple sclerosis--diagnostic, prognostic and clinical value. Nat Rev Neurol 11, 327-338, doi:10.1038/nrneurol.2015.80 (2015).

2 Casserly, C. et al. Spinal Cord Atrophy in Multiple Sclerosis: A Systematic Review and Meta-Analysis. J Neuroimaging 28, 556-586, doi:10.1111/jon.12553 (2018).

3 Rashid, W. et al. Increasing cord atrophy in early relapsing-remitting multiple sclerosis: a 3 year study. J Neurol Neurosurg Psychiatry 77, 51-55, doi:10.1136/jnnp.2005.068338 (2006).

4 Lukas, C. et al. Relevance of spinal cord abnormalities to clinical disability in multiple sclerosis: MR imaging findings in a large cohort of patients. Radiology 269, 542-552, doi:10.1148/radiol.13122566 (2013).

5 Biberacher, V. et al. Atrophy and structural variability of the upper cervical cord in early multiple sclerosis. Mult Scler 21, 875-884, doi:10.1177/1352458514546514 (2015).

6 Hagstrom, I. T. et al. Relevance of early cervical cord volume loss in the disease evolution of clinically isolated syndrome and early multiple sclerosis: a 2-year follow-up study. J Neurol 264, 1402-1412, doi:10.1007/s00415-017-8537-5 (2017).

7 Cadotte, A. et al. Spinal Cord Segmentation by One Dimensional Normalized Template Matching: A Novel, Quantitative Technique to Analyze Advanced Magnetic Resonance Imaging Data. PLoS One 10, e0139323, doi:10.1371/journal.pone.0139323 (2015).

8 El Mendili, M. M. et al. Validation of a semiautomated spinal cord segmentation method. J Magn Reson Imaging 41, 454-459, doi:10.1002/jmri.24571 (2015).

9 Kearney, H. et al. Improved MRI quantification of spinal cord atrophy in multiple sclerosis. J Magn Reson Imaging 39, 617-623, doi:10.1002/jmri.24194 (2014).

10 De Leener, B., Kadoury, S. & Cohen-Adad, J. Robust, accurate and fast automatic segmentation of the spinal cord. Neuroimage 98, 528-536, doi:10.1016/j.neuroimage.2014.04.051 (2014).

11 Gros, C. et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. Neuroimage 184, 901-915, doi:10.1016/j.neuroimage.2018.09.081 (2019).

12 Lukas, C. et al. Spinal cord atrophy in spinocerebellar ataxia type 3 and 6 : impact on clinical disability. J Neurol 255, 1244-1249, doi:10.1007/s00415-008-0907-6 (2008).

13 Horsfield, M. A. et al. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. Neuroimage 50, 446-455, doi:10.1016/j.neuroimage.2009.12.121 (2010).

14 Yushkevich, P. A. et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116-1128, doi:10.1016/j.neuroimage.2006.01.015 (2006).

15 Liu, Z. et al. Cervical cord area measurement using volumetric brain magnetic resonance imaging in multiple sclerosis. Mult Scler Relat Disord 4, 52-57, doi:10.1016/j.msard.2014.11.004 (2015).

16 Liu, Y. et al. Multicenter Validation of Mean Upper Cervical Cord Area Measurements from Head 3D T1-Weighted MR Imaging in Patients with Multiple Sclerosis. AJNR Am J Neuroradiol 37, 749-754, doi:10.3174/ajnr.A4635 (2016).

17 Eden, D. et al. Spatial distribution of multiple sclerosis lesions in the cervical spinal cord. Brain 142, 633-646, doi:10.1093/brain/awy352 (2019).

18 Gonzalez-Villa, S. et al. Evaluating the effect of multiple sclerosis lesions on automatic brain structure segmentation. Neuroimage Clin 15, 228-238, doi:10.1016/j.nicl.2017.05.003 (2017).

19 Sdika, M. & Pelletier, D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. Hum Brain Mapp 30, 1060-1067, doi:10.1002/hbm.20566 (2009).

20 Chard, D. T., Jackson, J. S., Miller, D. H. & Wheeler-Kingshott, C. A. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 32, 223-228, doi:10.1002/jmri.22214 (2010).

21 Battaglini, M., Jenkinson, M. & De Stefano, N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum Brain Mapp 33, 2062-2071, doi:10.1002/hbm.21344 (2012).

22 Popescu, V. et al. Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. Neuroimage Clin 4, 366-373, doi:10.1016/j.nicl.2014.01.004 (2014).

23 Amiri, H. et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin 19, 466-475, doi:10.1016/j.nicl.2018.04.023 (2018).

24 Yiannakas, M. C. et al. Fully automated segmentation of the cervical cord from T1-weighted MRI using PropSeg: Application to multiple sclerosis. Neuroimage Clin 10, 71-77, doi:10.1016/j.nicl.2015.11.001 (2016).

25 Polman, C. H. et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 69, 292-302, doi:10.1002/ana.22366 (2011).

26 Daams, M. et al. Mean upper cervical cord area (MUCCA) measurement in long-standing multiple sclerosis: relation to brain findings and clinical disability. Mult Scler 20, 1860-1865, doi:10.1177/1352458514533399 (2014).

27 Steenwijk, M. D. et al. What explains gray matter atrophy in long-standing multiple sclerosis? Radiology 272, 832-842, doi:10.1148/radiol.14132708 (2014).

28 Papinutto, N. et al. Gradient nonlinearity effects on upper cervical spinal cord area measurement from 3D T1 -weighted brain MRI acquisitions. Magn Reson Med 79, 1595-1601, doi:10.1002/mrm.26776 (2018).

29 Jovicich, J. et al. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436-443, doi:10.1016/j.neuroimage.2005.09.046 (2006).

30 Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20, 45-57, doi:10.1109/42.906424 (2001).

31 De Leener, B. et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage 145, 24-43, doi:10.1016/j.neuroimage.2016.10.009 (2017).

32 Lukas, C. et al. Benefit of repetitive intrathecal triamcinolone acetonide therapy in predominantly spinal multiple sclerosis: prediction by upper spinal cord atrophy. Ther Adv Neurol Disord 2, 42-49, doi:10.1177/1756285609343480 (2009).

33 Xinapse JIM 8.0. http://www.xinapse.com/Manual/index.html, 2018).

34 Panjabi, M. M., Duranceau, J., Goel, V., Oxland, T. & Takata, K. Cervical human vertebrae. Quantitative three-dimensional anatomy of the middle and lower regions. Spine (Phila Pa 1976) 16, 861-869 (1991).

35 Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825-841 (2002).

36 Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. Med Image Anal 5, 143-156 (2001).

37 Lukas, C. et al. Quantification of spinal cord atrophy in MS: which software, which vertebral level, spinal cord or brain MRI? A multi-centric, longitudinal comparison of three different volumetric approaches. Mult Scler J 24, 88-90 (2018).

38 Lukas, C. et al. Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. J Neurol Neurosurg Psychiatry 86, 410-418, doi:10.1136/jnnp-2014-308021 (2015).

39 Valsasina, P., Rocca, M. A., Horsfield, M. A., Copetti, M. & Filippi, M. A longitudinal MRI study of cervical cord atrophy in multiple sclerosis. J Neurol 262, 1622-1628, doi:10.1007/s00415-015-7754-z (2015).

40 Valsasina, P. et al. Cervical Cord T1-weighted Hypointense Lesions at MR Imaging in Multiple Sclerosis: Relationship to Cord Atrophy and Disability. Radiology 288, 234-244, doi:10.1148/radiol.2018172311 (2018).

CHAPTER 3.2

Upper cervical cord atrophy is independent of cervical cord lesion volume in early multiple sclerosis: a two-year longitudinal study

Merlin M. Weeda
Sofia Zywicki
Iman Brouwer
Bastiaan Moraal
Joep Killestein
Paolo Gallo
Frederik Barkhof
Petra J.W. Pouwels
Hugo Vrenken

Multiple Sclerosis and Related Disorders 60 (2022)

Abstract

Background

Upper cervical cord atrophy and lesions have been shown to be associated with disease and disability progression already in early relapsing-remitting multiple sclerosis (RRMS). However, their longitudinal relationship remains unclear.

Objective

To investigate the cross-sectional and longitudinal relation between focal T2 cervical cord lesion volume (CCLV) and regional and global mean upper cervical cord area (UCCA), and their relations with disability.

Methods

Over a two-year interval, subjects with RRMS (n=36) and healthy controls (HC, n=16) underwent annual clinical and MRI examinations. UCCA and CCLV were obtained from C1 through C4 level. Linear mixed model analysis was performed to investigate the relation between UCCA, CCLV, and disability over time.

Results

UCCA at baseline was significantly lower in RRMS subjects compared to HCs (p=0.003), but did not decrease faster over time (p≥0.144). UCCA and CCLV were independent of each other at any of the time points or cervical levels, and over time. Lower baseline UCCA, but not CCLV, was related to worsening of both upper and lower extremities function over time.

Conclusion

UCCA and CCLV are independent from each other, both cross-sectionally and longitudinally, in early MS. Lower UCCA, but not CCLV, was related to increasing disability over time.

Introduction

Spinal cord (SC) pathology is frequently seen in multiple sclerosis (MS) and is a strong contributor to disability and disease progression[1-3]. SC pathology includes lesions, which have a major role in diagnosis and prognosis of MS[4-6], and tissue loss or atrophy[7-10], which is generally assessed from upper cervical cord area (UCCA) measurements[2,11,12] and is already present in early stages of the disease[13,14].

The relation between cervical cord focal lesions and atrophy is not fully understood. The *in vivo* measurement of UCCA has shown to be robust in the presence of lesions[11], focal lesions and diffuse abnormalities in the cervical cord have been found to be associated with cord atrophy[14-16]. However, there is no longitudinal data on the relation between upper cervical cord area (UCCA) and cervical cord lesion volume (CCLV) in MS.

Therefore, we aimed to investigate the longitudinal evolution and interrelations of UCCA and CCLV in early RRMS over a two-year follow-up with annual visits. Secondly, we aimed to investigate the relation of UCCA and CCLV with physical disability.

Methods

Subjects

The institutional review board approved the study protocol and all participants gave written informed consent prior to participation, according to the Declaration of Helsinki.

To enable studying of the early disease course, patients included were diagnosed with clinical definite RRMS according to McDonalds 2010 criteria[17], with a maximal disease duration of 5 years and maximum expanded disability status scale (EDSS) score of 5.0. Subjects were using first-line treatment, or no treatment at all. In case of switching of treatment, MRI examinations were planned with at least 4-6 months delay[18]. When steroids were used, MRI was delayed by 3 months[19]. Patients were excluded (over the course of the study) in case of (switching to) second-line treatment.

To differentiate age-related MRI changes from disease-related MRI changes, a group of age, sex and education matched healthy controls (HCs) was included. HC and MS subjects were not eligible for participation when they could not undergo MRI examination, or when they had past or current clinically relevant neurological, psychiatric or (auto)immune disorders other than MS. Subjects visited a single-centre three times, with one-year intervals, for MRI and clinical and neuropsychological evaluation. Patients' disability was measured with the EDSS questionnaire[20]. The 9 Hole Peg Test (9-HPT, both dominant-hand [DH] and non-dominant hand [NDH]) and the 25 Foot Walk Test (25-FWT) were measured according to the Multiple Sclerosis Functional Composite (MSFC) scoring manual[21]. History taking included occurrence of relapses and change in therapy, and patients completed the Multiple Sclerosis Neuropsychological Screening Questionnaire (MSNQ)[22] and the Multiple Sclerosis Walking Scale (MSWS-12)[23]. All subjects completed the Checklist Individual Strength (CIS-20)[24] and the Athens Insomnia Scale (AIS)[25] questionnaires within four days of the visit.

MR image acquisition

MR imaging was performed on a 3T whole-body MR scanner GE Discovery MR750 (GE Healthcare, Milwaukee, WI., USA) using an an eight-channel phased-array head coil. No MRI hardware or software upgrades occurred during the study. The MR protocol included a 2D T2-weighted sequence planned perpendicular to the cervical cord (with TR/TE = 5035/85 ms, voxel size 0.4x0.4x4.0 mm, 26 consecutive slices), covering 104 mm of upper cervical cord starting at the most inferior surface of the pons.

For brain imaging, a sagittal 3D T1-weighted fast spoiled gradient echo sequence (FSPGR with TR/TE/TI = 8.2/3.2/450 ms, voxel size 1.0x1.0x1.0 mm), and a sagittal 3D T2-weighted fluid attenuated inversion recovery sequence (FLAIR with TR/TE/TI = 8000/130/2338 ms, voxel size 1.0x1.0x1.2 mm) with full coverage of the cerebrum and cerebellum, as well as (part of) the upper cervical cord.

MR image analysis

The upper cervical cord was segmented from the T2-weighted images using using SCT-PropSeg[26,27], a fully automated method incorporated in Spinal Cord Toolbox version 4.2. In a few cases the automatic detection of the cervical cord level failed and the label was selected by the operator using the sct_label_vertebrae option *initz*. A careful visual examination of segmentation output was performed and in case of failure, the segmentations were excluded from further analysis. Due to bending of the cord, C3 and C4 area could not be measured in all subjects at all time points. Therefore, we calculated UCCA for C1 through C4 separately.

One rater (SZ) manually delineated and segmented the cord lesions using ITK- SNAP Toolbox 3.6.0[28] under the supervision of an expert neuroradiologist (BM). Lesions were identified as focal hyperintensities on T2-weighted axial images, well demarcated from the surrounding normal-appearing tissue. To avoid false positives, a conservative approach was followed in which lesions were selected only if they were visible on at least two consecutive axial slices and could not be attributed to partial volume effects. The anatomical image was registered to the SC MRI template PAM50 and the T2 cervical cord lesion volume (CCLV) computed using the command line *sct_analyze_lesion*.

Details on brain imaging analysis can be found in **Supplemental Methods**.

Statistical analysis

Statistical analyses were performed using SPSS26 (IBM SPSS, Chicago, USA). Shapiro-Wilk test was used to assess the normally distribution of the variables. Group comparisons for baseline demographics were performed by independent samples *t*-test, Mann Whitney *U*-test, or Chi Square test, when appropriate.

Since not all subjects had an equal number of measurements (i.e. fewer time points and/or fewer cord levels reliably analyzed), linear mixed model (LMM) analysis was used to investigate: (a) differences in UCCA between HC and RRMS; (b) the relation between UCCA and the presence of lesions over the upper cervical cord; and (c) the relation between UCCA and CCLV at a given cervical level. Details on the LMM analyses are given in the **Supplemental Methods**. In short, the relation between UCCA and CCLV was investigated using two approaches: (1)

over all cervical levels (i.e. RRMS subjects with a lesion at any cervical level, referred to as "MS+CL", versus RRMS subjects without any lesion at any cervical level, "MS-CL"); and (2) for each cervical level separately (e.g., in relation to UCCA at the C1 level, only those patients with lesions at the C1 level were classified as having a lesion, and similarly for C2, C3 and C4). This second approach allowed an anatomical refinement of the LMM in the previous analysis by tying the UCCA to the presence or absence of cord lesions at the same level.
In addition, LMM was also used to investigate the relation between disability and UCCA in all RRMS subjects, and between disability and cervical lesions (presence and/or volume) in the MS+CL subjects (for details, see **Supplemental Methods**).

Results

Demographics

In total, 40 subjects with early RRMS and 15 age-and-sex matched HCs were included in the study. Data from 6 subjects were excluded from the analysis, because they had only a baseline examination (*n*=2 loss to follow-up, *n*=2 switch to second-line therapy), or because SCT-PropSeg failed at baseline (*n*=2). The baseline demographics of the final cohort of 13 HCs and 36 MS subjects are depicted in **Table 1**. Based on the presence of lesions in the upper cervical cord (C1 through C4) at baseline, the patient group was subdivided in MS-CL (i.e. without any lesions, *n*=20 subjects) and MS+CL (i.e. with lesions in at least one cord level, *n*=16 subjects).
At baseline, subjects with MS had significantly higher normalized brain lesion volume (NLV, *p*=0.008), CIS scores (*p*=0.004) and AIS scores (*p*=0.004) than HCs. The MS-CL group contained significantly fewer males than the MS+CL group (*p*=0.030), but no other baseline differences were seen between the two MS groups.
One HC (female) was unable to undergo MRI at year-1 but had a scan at year-2. Five MS subjects (female, all from MS-CL group) only had baseline and year 1, but not year-2, measurements (*n*=3 switch to second-line therapy; *n*=1 unable to undergo MRI; *n*=1 lost to follow-up).

Upper cervical cord area, lesions and lesion volume

Table 2 lists the UCCA in HC and MS subjects at each time point, and the change from baseline to year-2. Values are provided for the entire C1:C4 cord length (for those subjects in which this was available) and for each level separately. Within the MS group, we listed the number of patients without and with lesions at a specific cord level, and the UCCA of each group at that cord level, and CCLV when applicable.

Differences in UCCA between HC and MS subjects

LMM analysis showed that UCCA was lower in MS subjects than in HC (*B*=6.203, SE=1.975, *p*=0.003). In addition, UCCA reduced over time from year-0 to year-2 (*B*=0.986, SE=0.390, *p*=0.012) and from year-1 to year-2 (*B*=0.690, SE=0.348, *p*=0.048) in the entire cohort, but there was no interaction between time and subject type (i.e. HC or MS, *p*≥0.144). Neither sex nor age was a significant variable in the model.

Table 1 – Baseline demographics of HC, MS and MS-CL and MS+CL subjects.

	HC (*n*=13)	**MS** (*n*=36)	**MS-CL** (*n*=20)	**MS+CL** (*n*=16)
Age in years, mean ± SD	37.3 ± 12.2	35.6 ± 7.6	34.4 ± 7.7	37.0 ± 7.4
Sex, m/f (% m)	4/9 (31)	7/29 (19)	1/19 (5) *	6/10 (38)
Disease duration in years, mean ± SD	-	2.5 ± 1.3	2.3 ± 1.4	2.7 ± 1.4
ARR, median (range)	-	0 (0 – 2)	0 (0 – 2)	0 (0 – 2)
Treatment[1], *n* (%)				
none		8 (22)	3 (15)	5 (31)
interferon		4 (11)	2 (10)	2 (13)
glatiramer acetate	-	5 (17)	4 (20)	2 (13)
dimethyl fumarate		13 (36)	7 (35)	6 (38)
teriflunomide		2 (6)	1 (5)	1 (6)
Treatment duration in years, mean ± SD (range)	-	1.7 ± 1.1	1.7 ± 1.1	1.9 ± 1.2
EDSS, median (range)	-	3.0 (2.5 – 3.5)	3.0 (2.5 – 3.5)	3.0 (1.5 – 3.5)
9-HPT in seconds, mean ± SD				
dominant hand	-	19.0 ± 2.1	18.5 ± 2.0	19.6 ± 2.2
non-dominant hand		20.5 ± 2.8	20.1 ± 2.1	21.0 ± 3.4
25-FWT in seconds, mean ± SD	-	4.4 ± 1.2	4.5 ± 1.2	4.3 ± 1.4
CIS, median (IQR)	52 (38 – 64)	82 (51 – 92) **	85 (70 – 92)	59 (37 – 92)
AIS, median (IQR)	2 (2 – 4)	5 (3 – 6) **	5 (3 – 8)	4 (2 – 6)
MSWS, median (IQR)	-	18 (14 – 29)	18 (16 – 30)	17 (12 – 26)
MSNQ, median (IQR)	-	25 (15 – 33)	25 (20 – 31)	20 (8 – 35)
NBV in ml, mean ± SD	1538 ± 31	1530 ± 52	1517 ± 56	1545 ± 45
NLV in ml, mean ± SD	1.39 ± 3.32	4.68 ± 3.98 **	4.09 ± 3.72	5.42 ± 4.28

Legend: abbreviations: SD = standard deviation; ARR = annualized relapse rate; EDSS = expanded disability status scale; 9-HPT = 9-hole peg test; 25-FWT = 25-foot walk test; CIS = checklist individual strength; AIS = Athens insomnia scale; MSWS = multiple sclerosis walking scale; MSNQ = multiple sclerosis neurological screening questionnaire; NBV = normalized brain volume; NLV = normalized brain lesion volume; IQR = interquartile range; FU = follow-up. Statistics: * $p \leq 0.05$; ** $p \leq 0.01$.
[1] Treatment group interferon consists of interferon beta-1a (Avonex®, Rebif®), beta-1b (Betaferon®) and peginterferonbeta-1a (Plegridy®); other treatments are glatiramere acetate (Copaxone®), dimethyl fumarate (Tecfidera®), and teriflunomide (Aubagio®).

Relation between UCCA (change) and the presence of lesions in the upper cervical cord in MS subjects

No difference in baseline UCCA was seen between patients who had at least one lesion in the entire C1-C4 region (MS+CL) and those who did not (MS-CL) (B=-0.009, SE=2.264, p=0.997). Considering this comparison at each cord level separately, similarly, UCCA did not differ between patients with lesions at that level and those without (B=-0.271, SE=0.657, p=0.681). This was true for all time points, and neither sex nor age influenced these results.
Looking at change over time, UCCA change (Δ from year-0 to year-2) was independent of the presence of cord lesions at baseline for all cervical levels (p=0.996 for MS-CL vs MS+CL, and p=0.391 for subjects with and without lesions at a given cervical level, respectively). Neither sex nor age influenced these results.

Table 2 – Upper cervical cord area (UCCA) in the four cervical levels for HC and MS, as well as split for MS subjects without or with lesions in a given cervical level, as well as lesion volume in the latter.

	HC		MS		MS without lesions in given cervical level		MS with lesions in given cervical level		
Cervical level	***n***	**UCCA (mm^2)** mean ± SD	***n***	**UCCA (mm^2)** mean ± SD	***n***	**UCCA (mm^2)** mean ± SD	***n***	**UCCA (mm^2)** mean ± SD	**CCLV (µl)** median (IQR)
C1									
year 0	13	76.76 ± 6.99	36	71.71 ± 6.29	24	71.47 ± 5.29	12	72.20 ± 8.19	35.08 (14.04 – 66.40)
year 1	12	78.17 ± 5.97	35	71.39 ± 6.62	25	71.43 ± 6.28	10	71.31 ± 7.77	33.68 (17.20 – 159.90)
year 2	13	76.11 ± 7.21	31	70.84 ± 6.38	21	70.81 ± 5.68	10	70.90 ± 8.01	40.53 (12.38 – 111.19)
Δ y0-2	13	-0.65 ± 2.26	31	-1.07 ± 1.72	21	-0.70 ± 1.68	10	-1.87 ± 1.61	+5.24 (-49.88 – 82.19)
C2									
year 0	13	76.75 ± 7.30	36	70.87 ± 6.22	24	70.70 ± 5.72	12	71.20 ± 7.39	106.73 (58.45 – 154.41)
year 1	12	77.77 ± 6.86	36	71.00 ± 6.05	25	71.38 ± 5.79	11	70.13 ± 6.82	79.27 (12.62 – 141.67)
year 2	13	76.65 ± 6.82	31	69.97 ± 6.74	15	71.49 ± 6.30	16	68.55 ± 7.02	46.30 (29.97 – 87.02)
Δ y0-2	13	-0.10 ± 2.08	31	-1.02 ± 1.67	15	-0.83 ± 1.70	16	-1.20 ± 1.68	+15.15 (-63.34 – 34.42)
C3									
year 0	12	78.11 ± 7.60	35	72.56 ± 6.25	23	73.88 ± 6.04	12	70.03 ± 6.08	23.99 (14.34 – 213.37)
year 1	10	79.59 ± 8.12	31	72.43 ± 6.59	19	73.22 ± 6.23	12	71.17 ± 7.22	76.88 (32.26 – 155.87)
year 2	11	78.40 ± 7.41	27	71.84 ± 7.05	15	73.02 ± 7.00	12	70.37 ± 7.13	68.25 (31.77 – 228.43)
Δ y0-2	11	+0.35 ± 2.22	26	-0.83 ± 2.10	14	-0.41 ± 2.38	12	-1.32 ± 1.69	+10.29 (-8.75 – 56.95)
C4									
year 0	8	80.15 ± 6.94	26	74.75 ± 5.35	20	74.66 ± 5.80	6	75.03 ± 3.92	47.87 (10.58 – 167.44)
year 1	8	81.47 ± 6.58	29	74.63 ± 5.41	25	74.89 ± 5.79	4	72.99 ± 1.09	37.85 (10.35 – 158.98)
year 2	6	78.91 ± 5.69	20	73.58 ± 7.08	14	73.27 ± 6.81	6	74.28 ± 8.30	37.05 (15.35 – 73.94)
Δ y0-2	6	+0.42 ± 2.01	16	-0.62 ± 3.90	12	-0.99 ± 3.16	4	+0.50 ± 6.09	-11.94 (-57.65 – 25.22)

Legend: abbreviations: SD = standard deviation, IQR = inter-quartile range. Please note that values for Δ y0-2 are calculated from subjects with data at the given cervical level available at both year 0 and year 2, indicated by *n*, which may deviate from the arithmetic difference of the means at year 0 and year 2 for all available subjects.

Table 3 – Disability measures in MS-CL and MS+CL subjects at year 0, year 1, year 2 and over time (Δ y0-2).

		MS-CL		MS+CL
EDSS	***n***	**median (IQR)**	***n***	**median (IQR)**
year 0	20	3.0 (2.5 – 3.5)	16	3.0 (1.5 – 3.5)
year 1	19	3.0 (2.5 – 3.5)	16	3.5 (1.5 – 3.5)
year 2	15	3.5 (2.5 – 3.5)	16	3.0 (1.5 – 3.5)
Δ y0-2	15	+0.5 (-0.5 – 1.0)	16	0.0 (-0.5 – 0.0)
9-HPT, DH	***n***	**mean ± SD**	***n***	**mean ± SD**
year 0	20	18.5 ± 2.0	16	19.6 ± 2.2
year 1	20	18.6 ± 1.5	16	19.6 ± 4.2
year 2	15	18.3 ± 1.8	16	19.4 ± 4.0
Δ y0-2	15	-0.23 ± 1.74	16	-0.14 ± 2.68
9-HPT, NDH	***n***	**mean ± SD**	***n***	**mean ± SD**
year 0	20	20.1 ± 2.1	16	21.0 ± 3.4
year 1	19	20.0 ± 2.0	16	20.4 ± 3.7
year 2	15	19.3 ± 2.5	16	20.6 ± 4.3
Δ y0-2	15	-0.69 ± 1.75	16	-0.44 ± 2.10
25-FWT	***n***	**mean ± SD**	***n***	**mean ± SD**
year 0	20	4.5 ± 1.2	16	4.3 ± 1.4
year 1	19	4.5 ± 1.1	16	4.2 ± 1.3
year 2	15	4.9 ± 2.1	16	4.9 ± 2.4
Δ y0-2	15	+0.30 ± 0.90	16	+0.58 ± 1.23

		MS-CL		MS+CL
CIS	***n***	**median (IQR)**	***n***	**median (IQR)**
year 0	20	85 (70 – 92)	16	59 (37 – 92)
year 1	19	82 (61 – 96)	16	64 (41 – 85)
year 2	15	75 (61 – 92)	16	63 (46 – 86)
Δ y0-2	15	-5 (-18 – 8)	16	-1 (-18 – 24)
AIS	***n***	**median (IQR)**	***n***	**median (IQR)**
year 0	20	5 (3 – 8)	16	4 (2 – 6)
year 1	19	5 (3 – 9)	16	4 (2 – 5)
year 2	15	6 (4 – 9)	16	3 (2 – 8)
Δ y0-2	15	0 (0 – 2)	16	1 (-2 – 2)
MSWS	***n***	**median (IQR)**	***n***	**median (IQR)**
year 0	20	18 (16 – 30)	16	17 (12 – 26)
year 1	19	21 (15 – 27)	16	16 (13 – 26)
year 2	15	21 (14 – 27)	16	14 (12 – 23)
Δ y0-2	15	0 (-7 – 5)	16	0 (-3 – 1)
MSNQ	***n***	**median (IQR)**	***n***	**median (IQR)**
year 0	20	25 (20 – 31))	16	20 (8 – 35)
year 1	19	26 (19 – 32)	16	19 (11 – 35)
year 2	15	23 (16 – 30)	16	19 (9 – 34)
Δ y0-2	15	-2 (-8 – 2)	16	0 (-9 – 6)

Legend: 9-HPT and 25-FWT in seconds; abbreviations SD = standard deviation; IQR = inter-quartile range.

Relation between UCCA (change) and CCLV (change) by cervical level in MS subjects

In MS subjects with cord lesions at the respective cord level, UCCA was independent of CCLV (*B*=-0.008, SE=0.007, *p*=0.278) and CCLV was independent of UCCA (*B*=-0.428, SE=1.990, *p*=0.830) over all cervical levels. No effects of age, sex or time were found for either analysis. In addition, UCCA change over time (Δ year-0 to 2) was independent of baseline CCLV across all levels (**Figure 1a**, *B*=-0.001, SE=0.004, *p*=0.889), or in any of the cervical levels separately (data not shown). Furthermore, UCCA change over the second year (Δ year-1 to 2) was independent of CCLV change over the first year (Δ year-0 to 1) (**Figure 1b**, *B*=-0.001, SE=0.004, *p*=0.872). Sex, age and cervical level did not influence these results.

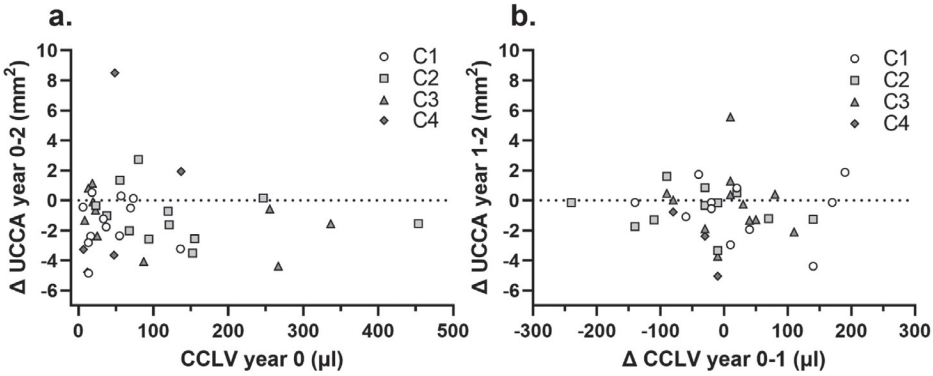

Figure 1 – Relationship between CCLV (µl) and UCCA (mm²) over four cervical levels.
(a) Relation between CCLV year 0 and UCCA from year-0 to year-2; (b) relation between CCLV change over the first year and UCCA change over the second year.

Relations between cervical cord area and cervical cord lesions with disability over time

Table 3 provides the disability measures in the overall MS-CL and MS+CL groups for each time point. One subject (male, MS+CL group) was unable to complete the 9-HPT correctly at multiple time points, and was therefore excluded from the 9-HPT analyses.
LMM analysis over all cervical levels and all time points showed no significant differences between the MS-CL and MS+CL groups for any of the disability measures, except for a trend towards slightly higher EDSS scores in the MS-CL group compared to the MS+CL group (*B*=0.831, SE=0.424, *p*=0.058).

Relation between disability and UCCA in MS subjects

Lower UCCA was related to higher 9-HPT-NDH times (i.e. poorer performance; *B*=-0.072, SE=0.020, *p*<0.001) (**Figure 2a**), with a significant contribution from age (*B*=0.131, SE=0.048, *p*=0.010). UCCA was not associated with any of the other disability measures (EDSS, 9-HPT DH, 25-FWT, CIS, AIS or MSWS scores). However, a trend was seen for higher MSNQ outcome in MS subjects with lower UCCA (*B*=-0.134, SE=0.080, *p*=0.095).

Relation between disability and cervical lesions (presence and/or volume) in MS subjects

Using the cord level-specific classification into patients with or without lesions, EDSS did not differ between MS subjects with lesions at that level and those without (*p*=0.879). A group * time interaction was found, where subjects without lesions at a specific level performed better than subjects with lesions for 9-HPT NDH scores over time (year-1 to year-2) (*B*=0.549, SE=0.253, *p*=0.031) as shown in **Figure 2b**, but age significantly affected this relationship (*B*=0.127, SE=0.048, *p*=0.012).
A similar interaction was found for 25-FWT outcome (**Figure 2c**), where subjects without lesions in a specific cervical level also performed better over time than subjects with lesions (*B*=0.330, SE=0.150, *p*=0.029 for year-0 to year-2 and *B*=0.418, SE=0.153, *p*=0.006 for year-1 to year-2). Across all time points, the groups did not differ significantly from each other regarding the 25-FWT, although a trend was observed (*B*=-0.238, SE=0024, *p*=0.080). Other disability

measures CIS, AIS, MSWS and MSNQ did not differ between patients with or without lesions in a specific cervical level.

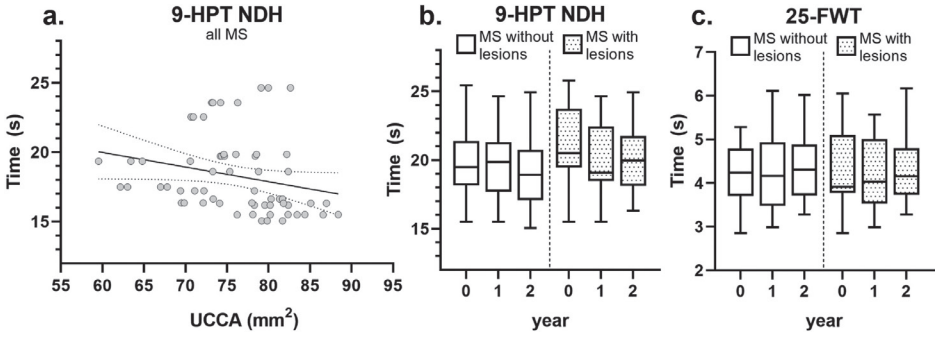

Figure 2 – Disability in MS subjects.

(a) 9-HPT NDH outcome with regard to UCCA shows a negative relation between the two; (b) Subjects without lesions in a given cervical level (white) have lower 9-HPT non-dominant hand times from year-1 to year-2 when compared to subjects with lesions in a given cervical level (dotted); (c) Comparable results are seen for 25-FWT, where subjects with lesions in a given cervical level (dotted) have a greater increase in 25-FWT time from year-0 to year-2 than subjects without lesions in a given cervical level (white).

Looking at disability progression over time in patients with lesions at a specific cervical cord level, we found that 9-HPT progression over time was not influenced by baseline CCLV (**Figure 3a-b**), but it was associated with baseline UCCA (**Figure 3c-d**) (DH: *B*=-0.056, SE=0.028 *p*=0.051; NDH: hand *B*=-0.066, SE=0.027, *p*=0.015), where lower baseline UCCA was predictive of poorer 9-HPT performance. EDSS, CIS, AIS, MSWS and MSNQ progression over time could not be explained by baseline UCCA, and none of the disability outcomes could be explained by baseline CCLV.

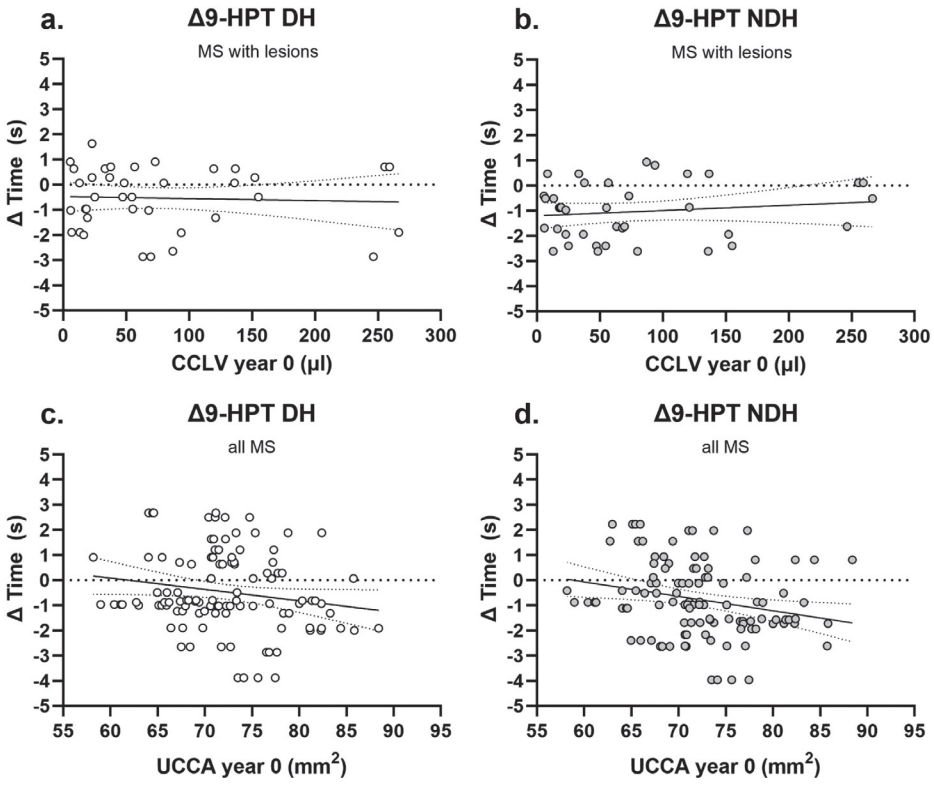

Figure 3 – 9-HPT progression over time from the dominant (white circles) and non-dominant (grey circles) hand shows no relation with baseline CCLV (top), but a negative relation with baseline UCCA (bottom).

Discussion

In this study, we investigated the cross-sectional and longitudinal relationship between upper cervical cord area and lesions in subjects with early RRMS. The study design and statistical analyses presented here allowed us to investigate the longitudinal dynamics and relations between cervical cord lesions (presence or volume) and cervical cord area, both globally (i.e. over the C1:C4 area) and locally (i.e. per cervical level).

We found a lower UCCA in subjects with RRMS compared to HCs but decrease of UCCA over time occurred independent of group. Moreover, UCCA was not related to the presence or volume of cervical cord lesions (CCLV), neither cross-sectionally nor longitudinally. While we found a relation between lower baseline UCCA and faster worsening of 9-HPT-NDH and 25-FWT scores over time, there was no relation between disability change over time and the presence or volume of lesions in the upper cervical cord.

We found no relation between the presence or volume of cervical cord lesions and UCCA at any time point (cross-sectional) or considering their change over time (longitudinal). Although some studies have reported these associations in progressive MS types[16,29,30], in RRMS and CIS this association has not been found[31,32]. These results seem to be confirmed by post-

mortem studies, which observed that neither lesion size nor lesion number in the SC correlated with the degree of local atrophy of the cord when correcting for disease duration[15,33]. Results of several *in vivo* cord studies confirmed this, by showing that local demyelination[34], increased diffusivity[35] and decreased axonal and neurite density[16,36], rather than lesion size or number, were substrates of decreased UCCA in MS.

RRMS subjects with lower baseline UCCA showed subsequent performance worsening on 9-HPT-NDH (upper extremity function) and 25-FWT (lower extremity function) over time in our study, which was also found previously[14,32]. UCCA was not related to EDSS scores in our study, in line with previous studies in non-progressive MS[5,8,16,31,37,38]. These results underline the importance of investigating the different functional systems when looking at the relation between disability and UCCA in early MS. It appears that in early MS, cervical cord atrophy is correlated to pyramidal, and mainly manual, impairment[29,31,32], rather than overall disability (i.e. EDSS). We found no associations between CCLV and disability in our early MS cohort, where this relation was present in more progressive MS phenotypes with moderate to severe disability[5,37,39] or long-standing MS[40]. Interestingly, the relationship of lower UCCA with worsening upper extremity function has also been reported in spinal cord injury due to other causes[41], which seems to support the hypothesis that overall spinal cord damage rather than MS-related inflammation (e.g. CCLV) may be the physiological basis of upper extremity disability.

Our result that subjects with RRMS exhibited lower UCCA compared to HC is in line with most studies[7,15,29,31], although some studies did not find a difference in UCCA between RRMS and HC subjects[42,43]. An important factor in these differences is disease duration[15,31], as well as possible swelling in the cervical cord due to inflammation[31,44], which may increase UCCA, especially in earlier MS stages such as clinically isolated syndrome (CIS)[7]; since we studied subjects early in their disease course, this may have played a role in our results as well. In addition, the cervical cord atrophy rate appears to be similar for HC and RRMS subjects[44] and therefore seems to accelerate mostly in the progressive stages of the disease[42,43,45].

This study has a few limitations. The MS+CL group had significantly fewer males, therefore we included sex in all statistical analyses. Furthermore, bending of the cervical cord makes it difficult to acquire C3 and C4 results in all subjects, therefore LMM analysis was used to overcome the problem of missing data. In addition, did not use the generalized boundary-shift integral (GBSI) for the longitudinal cord atrophy measurements, but instead used LMM analysis to minimize subject-specific measurement variability. Last, it is important to note that our relatively small cohort consisted of subjects with a low degree of disease activity, treated only with first-line immunomodulatory therapies or no treatment at all. These inclusion criteria allow to study *in vivo* biological phenomena with little or no effect of therapy, but potentially exclude those patients in whom the pathological mechanisms occurring in MS (including neurodegeneration and neuroinflammation) are more pronounced.

Conclusion

In conclusion, we found that there is no relation between cervical cord lesions (presence and/or volume) and cervical cord atrophy in subjects with early RRMS, neither cross-sectionally nor longitudinally. In addition, RRMS subjects with lower UCCA values appear to worsen in both upper and lower extremity function over time. Results should be replicated over longer follow-up periods.

Funding

This work was funded by the Dutch MS Research Foundation, Grant No. 14-876. FB is supported by the NIHR biomedical research centre at UCLH.

Declaration of interest

The authors declare that there is no conflict of interest.

Supplements

Supplementary materials associated with this article can be found, in the online version, at https://www.doi.org/10.1016/j.msard.2022.103713.

MR image analysis

MR image analysis was performed with FMRIB software library (FSL) version 5.0.10, FreeSurfer version 6.0.0, nicMSlesions version 0.2, LEsion Automated Preprocessing (LEAP), Spinal Cord Toolbox (SCT) version 4.2.2, and ITK-SNAP version 3.6.0.

Brain imaging

Since the upper cervical cord was included in the T1 and FLAIR field of view (FOV), neck removal was performed using FSL standard_space_roi. Next, brain extraction was performed using FSL BET[46] optimized for the current scan type[47]. Segmentation of the brain into GM, WM and cerebrospinal fluid (CSF) partial volumes was performed with FSL FAST[48]. FreeSurfer was used for N3 bias field correction. Lesion segmentation was performed with the deep-learning algorithm nicMSlesions[49,50] which was optimized for our data in an earlier study[51]. In summary, full re-training of the nicMSlesions neural network (11 layers) was done with the use of manual lesion segmentations available from fourteen subjects with MS. All parameters were set at default, and the optimized probability threshold was set at 0.4.

Lesion filling was performed with LEAP[52] and lesion filled images were processed with the longitudinal pipeline of FreeSurfer[53-55], using a template-driven approach to provide a detailed parcellation and segmentation of the cortex and subcortical structures[56].

All brain volumes were normalized for head size, by multiplying the raw volume with the normalization factor (i.e. standard brain size 1948.106 ml divided by the subjects' estimated total intracranial volume [ETIV]).

Statistical analysis

Differences in UCCA between HC and MS subjects

For each cord level separately as well as for the entire C1:C4 cord length (for those subjects in which this was available), mean UCCA was calculated in HCs and people with MS, at each time point, as well as the change from year-0 to year-2 (year-2 minus year-0 = Δ year 0-2). In addition, within the MS group, numbers of patients with and without lesions, mean UCCA in each of these groups, and mean cervical cord lesion volume (CCLV) in the group with lesions, were calculated for each case. LMMs were used to compare UCCA between MS and HC, using UCCA as dependent variable, cord level as the repeated factor within subjects, group as independent variable and age and sex as confounders. A second LMM with time as an additional repeated factor within subjects, was used to analyze the change over time and compare this between MS and HC.

Relation between UCCA and the presence of lesions in the upper cervical cord in MS subjects

Within the MS group, the relation between the presence or absence of cord lesions and cord UCCA was assessed in two ways. First, to assess the global relation between the presence of any lesions in the cervical cord and UCCA at the different cord levels, we compared UCCA between subjects who had cervical cord lesions at any level of the cord (MS+CL) and those who did not have any cervical cord lesions (MS-CL). Again, UCCA at each cord level (C1 through C4) was the repeated dependent variable in these LMM analyses. Since UCCA can vary with sex and age, both were included in the linear mixed model.

Second, to assess the relation between lesions and UCCA within a certain cervical level, we subdivided the MS patients into those without and with cord lesions depending on the precise cord level being analyzed. E.g., in relation to UCCA at the C1 level, only those patients with lesions at the C1 level were classified as having a lesion, and similarly for C2, C3 and C4. This approach allowed a refinement of the LMM in the previous analysis by tying the UCCA to the presence or absence of cord lesions at the same level as that of the UCCA measurement. Other than that, the same LMM as in the previous analysis was used.

Relation between UCCA and CCLV in a given cervical level in MS subjects

Within the MS group, the relation between the cord level-specific lesion volume and cord-level specific UCCA was studied using LMM analysis, reproducing the model used above but replacing presence or absence of lesions by the lesion volume, and including, at each cord level, lesion volumes only for those subjects who actually had any lesions at that cord level. To complete the cross-sectional assessment of any possible relation between these two variables, we additionally performed similar LMM analyses in which UCCA was the independent variable and CCLV was the dependent variable. Next, a longitudinal model was applied incorporating all time points, including time as an additional repeated factor within subjects. Last, we also calculated UCCA and CCLV change over time (i.e. over year 0-1, year 1-2, year 0-2 by subtracting the former year from the latter) and performed LMM analysis with baseline CCLV as dependent and UCCA change as independent variable, and vice versa.

Relation between disability and UCCA in MS subjects

LMMs were used to compare disability parameters EDSS, 9-HPT, 25-FWT, CIS, AIS, MSWS and MSNQ between MS-CL and MS+CL subjects. In this, the disability parameter was the dependent variable, UCCA was the independent variable, cord level was the repeated factor, and age and sex were possible confounders. In a second LMM, time was added as an additional repeated factor within subjects to analyse the change over time.

Relation between disability and cervical lesions (presence and/or volume) in MS subjects

The presence of cervical lesions and its relation with disability was investigated with an LMM as described in the section above, with group (i.e. MS-CL or MS+CL) as additional independent variable.

The relation between disability and CCLV was investigated with an LMM where disability was the dependent variable, CCLV was the independent variable, cord level and time were the repeated factors, and age and sex were possible confounders.

References

1 Gass, A. et al. MRI monitoring of pathological changes in the spinal cord in patients with multiple sclerosis. Lancet Neurol 14, 443-454, doi:10.1016/S1474-4422(14)70294-7 (2015).

2 Moccia, M. et al. Advances in spinal cord imaging in multiple sclerosis. Ther Adv Neurol Disord 12, 1756286419840593, doi:10.1177/1756286419840593 (2019).

3 Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B. & Ciccarelli, O. Multiple sclerosis. Lancet 391, 1622-1636, doi:10.1016/S0140-6736(18)30481-1 (2018).

4 Rocca, M. A., Preziosa, P. & Filippi, M. What role should spinal cord MRI take in the future of multiple sclerosis surveillance? Expert Rev Neurother 20, 783-797, doi:10.1080/14737175.2020.1739524 (2020).

5 Kearney, H. et al. Cervical cord lesion load is associated with disability independently from atrophy in MS. Neurology 84, 367-373, doi:10.1212/WNL.0000000000001186 (2015).

6 Thompson, A. J. et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 17, 162-173, doi:10.1016/S1474-4422(17)30470-2 (2018).

7 Biberacher, V. et al. Atrophy and structural variability of the upper cervical cord in early multiple sclerosis. Mult Scler 21, 875-884, doi:10.1177/1352458514546514 (2015).

8 Brownlee, W. J. et al. Association of asymptomatic spinal cord lesions and atrophy with disability 5 years after a clinically isolated syndrome. Mult Scler 23, 665-674, doi:10.1177/1352458516663034 (2017).

9 Losseff, N. A. et al. Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. Brain 119 (Pt 3), 701-708, doi:10.1093/brain/119.3.701 (1996).

10 Bot, J. C. et al. Spinal cord abnormalities in recently diagnosed MS patients: added value of spinal MRI examination. Neurology 62, 226-233, doi:10.1212/wnl.62.2.226 (2004).

11 Weeda, M. M. et al. Validation of mean upper cervical cord area (MUCCA) measurement techniques in multiple sclerosis (MS): High reproducibility and robustness to lesions, but large software and scanner effects. Neuroimage Clin 24, 101962, doi:10.1016/j.nicl.2019.101962 (2019).

12 Sastre-Garriga, J. et al. MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. Nature Reviews Neurology 16, 171-182, doi:10.1038/s41582-020-0314-x (2020).

13 Hagstrom, I. T. et al. Relevance of early cervical cord volume loss in the disease evolution of clinically isolated syndrome and early multiple sclerosis: a 2-year follow-up study. J Neurol 264, 1402-1412, doi:10.1007/s00415-017-8537-5 (2017).

14 Lukas, C. et al. Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. J Neurol Neurosurg Psychiatry 86, 410-418, doi:10.1136/jnnp-2014-308021 (2015).

15 Evangelou, N., DeLuca, G. C., Owens, T. & Esiri, M. M. Pathological study of spinal cord atrophy in multiple sclerosis suggests limited role of local lesions. Brain 128, 29-34, doi:10.1093/brain/awh323 (2005).

16 Pravata, E. et al. Influence of CNS T2-focal lesions on cervical cord atrophy and disability in multiple sclerosis. Mult Scler J 26, 1402-1409, doi:Artn 1352485198659891019.1177/1352458519865989 (2020).

17 Polman, C. H. et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 69, 292-302, doi:10.1002/ana.22366 (2011).

18 De Stefano, N. et al. Evidence of early cortical atrophy in MS: relevance to white matter changes and disability. Neurology 60, 1157-1162 (2003).

19 Zivadinov, R. et al. Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years. AJNR Am J Neuroradiol 34, 1931-1939, doi:10.3174/ajnr.A3503 (2013).

20 Lechner-Scott, J. et al. Can the Expanded Disability Status Scale be assessed by telephone? Mult Scler 9, 154-159, doi:10.1191/1352458503ms884oa (2003).

21 Cohen, J. A. et al. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. Arch Neurol 58, 961-967, doi:10.1001/archneur.58.6.961 (2001).

22 Sonder, J. M., Mokkink, L. B., van der Linden, F. A., Polman, C. H. & Uitdehaag, B. M. Validation and interpretation of the Dutch version of the Multiple Sclerosis Neuropsychological Screening Questionnaire. Journal of the neurological sciences 320, 91-96 (2012).

23 Hobart, J., Riazi, A., Lamping, D., Fitzpatrick, R. & Thompson, A. Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). Neurology 60, 31-36 (2003).

24 Vercoulen, J. H. et al. Dimensional assessment of chronic fatigue syndrome. Journal of psychosomatic research 38, 383-392 (1994).

25 Soldatos, C. R., Dikeos, D. G. & Paparrigopoulos, T. J. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. Journal of psychosomatic research 48, 555-560 (2000).

26 De Leener, B., Kadoury, S. & Cohen-Adad, J. Robust, accurate and fast automatic segmentation of the spinal cord. Neuroimage 98, 528-536, doi:10.1016/j.neuroimage.2014.04.051 (2014).

27 Gros, C. et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. Neuroimage 184, 901-915, doi:10.1016/j.neuroimage.2018.09.081 (2019).

28 Yushkevich, P. A. et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116-1128, doi:10.1016/j.neuroimage.2006.01.015 (2006).

29 Valsasina, P. et al. Cervical Cord T1-weighted Hypointense Lesions at MR Imaging in Multiple Sclerosis: Relationship to Cord Atrophy and Disability. Radiology 288, 234-244, doi:10.1148/radiol.2018172311 (2018).

30 Petrova, N., Carassiti, D., Altmann, D. R., Baker, D. & Schmierer, K. Axonal loss in the multiple sclerosis spinal cord revisited. Brain Pathol 28, 334-348, doi:10.1111/bpa.12516 (2018).

31 Valsasina, P. et al. Characterizing 1-year development of cervical cord atrophy across different MS phenotypes: A voxel-wise, multicentre analysis. Mult Scler, 13524585211045545, doi:10.1177/13524585211045545 (2021).

32 Zurawski, J. et al. The impact of cervical spinal cord atrophy on quality of life in multiple sclerosis. J Neurol Sci 403, 38-43, doi:10.1016/j.jns.2019.04.023 (2019).

33 Gilmore, C. P. et al. Spinal Cord Neuronal Pathology in Multiple Sclerosis. Brain Pathol 19, 642-649, doi:10.1111/j.1750-3639.2008.00228.x (2009).

34 Lee, L. E. et al. Cervical cord myelin abnormality is associated with clinical disability in multiple sclerosis. Mult Scler J, doi:Artn 13524585211001780 10.1177/13524585211001780 (2021).

35 Wolanczyk, M. et al. Diffusion tensor imaging of normal-appearing cervical spinal cords in patients with multiple sclerosis: Correlations with clinical evaluation and cerebral diffusion tensor imaging changes. Preliminary experience. Adv Clin Exp Med 29, 441-448, doi:10.17219/acem/116754 (2020).

36 Collorone, S. et al. Reduced neurite density in the brain and cervical spinal cord in relapsing-remitting multiple sclerosis: A NODDI study. Mult Scler 26, 1647-1657, doi:10.1177/1352458519885107 (2020).

37 Kerbrat, A. et al. Multiple sclerosis lesions in motor tracts from brain to cervical cord: spatial distribution and correlation with disability. Brain 143, 2089-2105, doi:10.1093/brain/awaa162 (2020).

38 Lin, X., Blumhardt, L. D. & Constantinescu, C. S. The relationship of brain and cervical cord volume to disability in clinical subtypes of multiple sclerosis: a three-dimensional MRI study. Acta Neurol Scand 108, 401-406, doi:10.1034/j.1600-0404.2003.00160.x (2003).

39 Eden, D. et al. Spatial distribution of multiple sclerosis lesions in the cervical spinal cord. Brain 142, 633-646, doi:10.1093/brain/awy352 (2019).

40 Dekker, I. et al. Infratentorial and spinal cord lesions: Cumulative predictors of long-term disability? Mult Scler 26, 1381-1391, doi:10.1177/1352458519864933 (2020).

41 Freund, P. et al. Disability, atrophy and cortical reorganization following spinal cord injury. Brain 134, 1610-1622, doi:10.1093/brain/awr093 (2011).

42 Klein, J. P. et al. A 3T MR imaging investigation of the topography of whole spinal cord atrophy in multiple sclerosis. AJNR Am J Neuroradiol 32, 1138-1142, doi:10.3174/ajnr.A2459 (2011).

43 Rocca, M. A. et al. A multicenter assessment of cervical cord atrophy among MS clinical phenotypes. Neurology 76, 2096-2102, doi:10.1212/WNL.0b013e31821f46b8 (2011).

44 Brex, P. A. et al. Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis. J Neurol Neurosurg Psychiatry 70, 544-547, doi:10.1136/jnnp.70.4.544 (2001).

45 Zeydan, B. et al. Cervical spinal cord atrophy: An early marker of progressive MS onset. Neurol Neuroimmunol Neuroinflamm 5, e435, doi:10.1212/NXI.0000000000000435 (2018).

46 Smith, S. M. Fast robust automated brain extraction. Hum Brain Mapp 17, 143-155, doi:10.1002/hbm.10062 (2002).

47 Popescu, V. et al. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. Neuroimage 61, 1484-1494, doi:10.1016/j.neuroimage.2012.03.074 (2012).

48 Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20, 45-57, doi:10.1109/42.906424 (2001).

49 Valverde, S. et al. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage 155, 159-168, doi:10.1016/j.neuroimage.2017.04.034 (2017).

50 Valverde, S. et al. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. Neuroimage Clin 21, 101638, doi:10.1016/j.nicl.2018.101638 (2019).

51 Weeda, M. M. et al. Comparing lesion segmentation methods in multiple sclerosis: Input from one manually delineated subject is sufficient for accurate lesion segmentation. Neuroimage Clin 24, 102074, doi:10.1016/j.nicl.2019.102074 (2019).

52 Chard, D. T., Jackson, J. S., Miller, D. H. & Wheeler-Kingshott, C. A. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 32, 223-228, doi:10.1002/jmri.22214 (2010).

53 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179-194, doi:10.1006/nimg.1998.0395 (1999).

54 Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195-207, doi:10.1006/nimg.1998.0396 (1999).

55 Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402-1418, doi:10.1016/j.neuroimage.2012.02.084 (2012).

56 Guo, C. J., Ferreira, D., Fink, K., Westman, E. & Granberg, T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. Eur Radiol 29, 1355-1364, doi:10.1007/s00330-018-5710-x (2019).

CHAPTER 4

Grey and white matter damage in the MS brain

CHAPTER 4.1

Damage in the thalamocortical tracts is associated with subsequent thalamus atrophy in early multiple sclerosis

Merlin M. Weeda
Ilanah J. Pruis
Aimée S.R. Westerveld
Iman Brouwer
Barbara Bellenberg
Frederik Barkhof
Hugo Vrenken
Carsten Lukas
Ruth Schneider[#]
Petra J.W. Pouwels[#]

[#] *Shared last authors*

Frontiers in Neurology 11 (2020)

Abstract

Background

In early multiple sclerosis (MS), thalamus atrophy and decreased integrity of the thalamocortical white matter (WM) tracts have been observed.

Objective

To investigate the temporal association between thalamus volume and WM damage in the thalamocortical tract in subjects with early MS.

Methods

At two time points, 72 subjects with early MS underwent T1, FLAIR and diffusion tensor imaging. Thalamocortical tracts were identified with probabilistic tractography using left and right thalamus as seed regions. Regression analysis was performed to identify predictors of annual percentage change in both thalamus volumes and integrity of the connected tracts.

Results

Significant atrophy was seen in left and right thalamus ($p<0.001$) over the follow-up period (13.7 ± 4.8 months), whereas fractional anisotropy (FA) and mean diffusivity (MD) changes of the left and right thalamus tracts were not significant, although large inter-subject variability was seen. Annual percentage change in left thalamus volume was significantly predicted by baseline FA of the left thalamus tracts $F(1,71)=4.284$, $p=0.042$; while no such relation was found for the right thalamus. Annual percentage change in FA or MD of the thalamus tracts was not predicted by thalamus volume or any of the demographic parameters.

Conclusion

Over a short follow-up time, thalamus atrophy could be predicted by decreased integrity of the thalamic tracts, but changes in the integrity of the thalamic tracts could not be predicted by thalamus volume. This is the first study showing directionality in the association between thalamus atrophy and connected WM tract damage. These results need to be verified over longer follow-up periods.

Introduction

Multiple sclerosis (MS) is an autoimmune disorder affecting the central nervous system. MS is characterized by atrophy in the grey matter (GM) and widespread pathology in the white matter (WM), which can be investigated with diffusion tensor imaging (DTI). From DTI, fractional anisotropy (FA) and mean diffusivity (MD) in WM can be estimated, which are both simplified measures for the overall integrity of WM[1,2].

The thalamus, a central structure in the brain that acts as a convergence location as well as a gateway to the cortex[3], has numerous reciprocal WM connections to the cortex and subcortical structures, also referred to as the thalamocortical projections. These thalamocortical projections connect distinct thalamic nuclei to cortical areas, such as to the prefrontal and temporal cortex, to the posterior parietal cortex and to the somatosensory and motor / premotor cortices[4]. Due to its multitude of cortical connections the thalamus is involved in numerous neurological functions, such as motor, sensory, executive and higher cortical functions. Furthermore the thalamus plays a significant role in other functions, such as memory, emotion, attention and the regulation of sleep and wakefulness[3].

In subjects with MS, thalamic volumes are reduced when compared to healthy controls[5,6] and have been found to decrease further, even early in the disease and over short follow-up periods[7-9]. Moreover, thalamus atrophy is correlated with clinical disability[10,11].

The thalamocortical projections exhibit higher lesion loads and more WM loss than non-thalamocortical projections in subjects with early MS[12,13]. In addition, subjects with early MS showed a correlation between reduced thalamic volume and reduced FA in the WM adjacent to the thalamus[14] or damage in the connected WM tracts, as seen by decreased FA and increased MD as well as increased WM lesion volume in the thalamocortical projections[13]. Although some studies show a relationship between global WM damage and deep GM atrophy[15-17], no longitudinal studies have focused on damage of the WM specifically connected to these deep GM structures. Moreover, investigation of the directionality of the relationship between GM atrophy and WM damage remains scarce and inconclusive[13,18-20].

Therefore, this study aims to investigate the longitudinal association between thalamus volume and WM damage in the connected tracts in subjects with early MS over a short follow-up. We hypothesize that damage in the thalamocortical WM is a predictor of thalamic atrophy, but that lower baseline thalamus volume does not predict increased thalamocortical WM damage over time.

Methods

Subjects

MS patients from a single center who were enrolled in a longitudinal MS cohort study[21,22] and who underwent at least two MRI examinations including DTI were included for this study. Subjects included were all above 18 years of age and diagnosed with either RRMS based on the 2005 McDonald criteria within 2.5 years before the first MRI with DTI, or diagnosed with CIS within 1.3 years before the first MRI with DTI. Clinical performance was scored using the

Expanded Disability Status Scale (EDSS)[23]. Other parameters used for this study were: age, sex, disease duration (i.e. time from first clinical manifestation), treatment type (if applicable), and treatment duration (i.e. time on current treatment). The study was approved by the local institutional regulatory board (Reg.-No.:3714-10), and written informed consent was obtained from all individuals, according to the Declaration of Helsinki.

MR image acquisition

Brain images were acquired on a 3 Tesla scanner (Philips Achieva, Best, The Netherlands) with a 32-channel head coil following a standardized imaging protocol with sequences covering the entire brain[24]. For this analysis, we included the following acquisitions prior to contrast injection: (1) sagittal 3D T1-weighted fast field echo (FFE) (repetition time (TR)/ echo time (TE)=8.5-10ms/4.6ms; 180 slices; resolution=1.0x1.0x1.0mm^3; flip angle=8°); (2) sagittal 3D Fluid-attenuation Inversion Recovery (FLAIR) (TR/TE=4800ms/286-323.5ms; inversion time (TI)=1650ms; resolution=1.0x1.0x1.0mm^3); and (3) transversal 2D echo-planar DTI (TR/TE=7000ms/90ms; 2.5mm slices; in-plane resolution=2.5x2.5mm^2, 32 volumes with b-value=900s/mm^2 and one volume, 2 averages, without diffusion weighting).

MR imaging data analysis

MR imaging data analysis was performed with FMRIB Software Library (FSL) version 5.08 and FreeSurfer version 6.0.

Structural imaging

Brain extraction was performed on all T1w and FLAIR images using FSL BET[25], optimized for the current scan type and specifications[26]. For lesion segmentation and subsequent filling on the T1w images with the signal intensity of the surrounding WM, Lesion Segmentation Toolbox with Lesion Prediction Algorithm (LST-LPA) was used[27,28]. Lesion-filled T1w images were processed with the longitudinal pipeline of FreeSurfer[29-31] and whole brain, WM, GM and left and right thalamus volumes were extracted. Volumes of whole brain, brain structures, and lesions were not normalized. An example of lesion segmentation on FLAIR images and FreeSurfer segmentation of brain volumes on T1w images is shown in **Figure 1**. Using FreeSurfer's transform matrices, segmentations were registered back to T1 native space with nearest neighbour interpolation, and subsequently linearly registered to DTI space with nearest neighbour interpolation.

All processed images were checked visually for potential errors in the segmentation of FreeSurfer and/or LST-LPA.

Diffusion imaging

After correction for head movement and eddy current distortions[32] in the diffusion-weighted scans, the diffusion tensor was fitted with weighted least squares to obtain maps of FA and MD. The two-fiber model bedpostx[33] was used for estimation of the voxel-wise diffusion parameter distribution. Probabilistic tractography (probtrackx2)[34] was performed on the whole brain with the left and right thalami as separate seed regions.

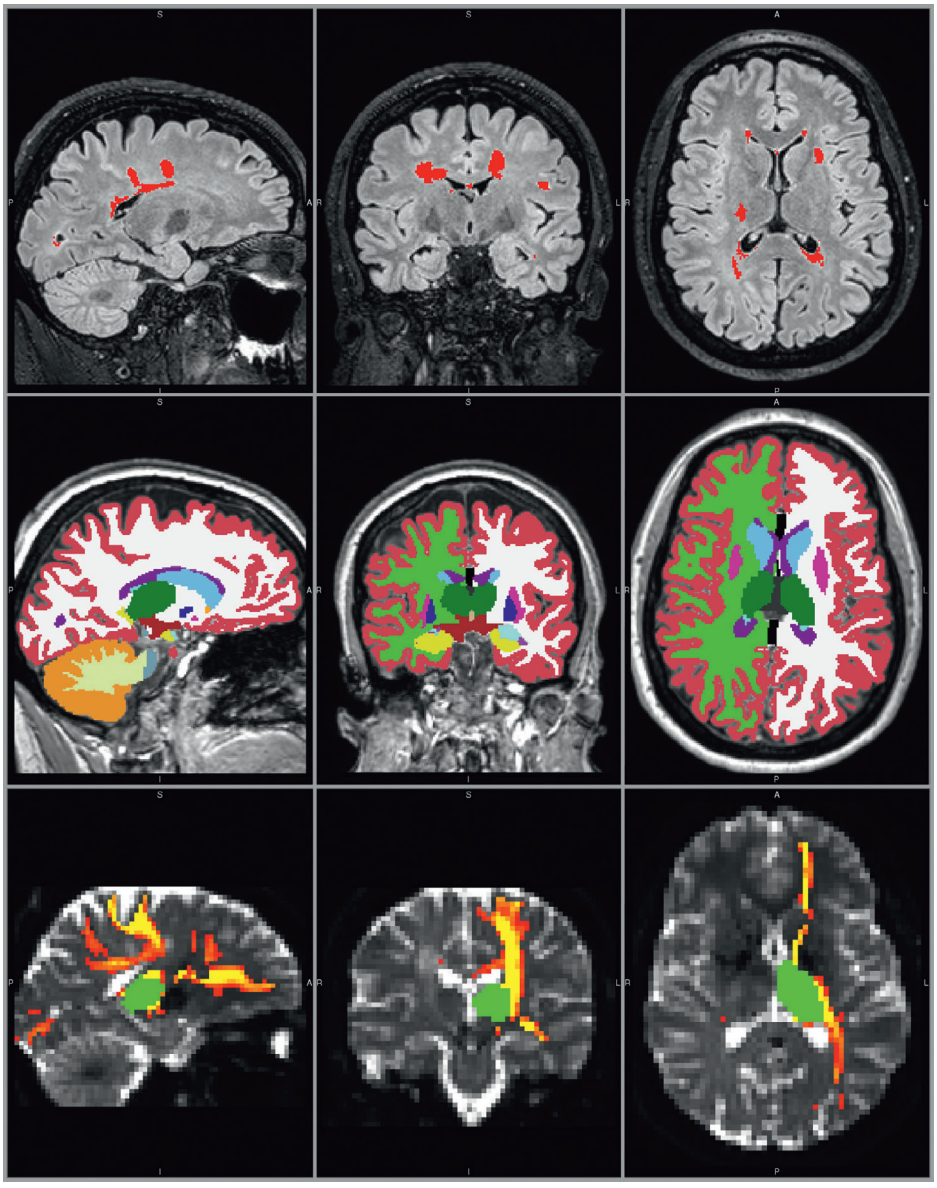

Figure 1 – MRI processing example in a 28 year old female with RRMS.

Top panels: T2-FLAIR weighted image with LST LPA lesion segmentation (red). Middle panels: T1-weighted lesion filled image with FreeSurfer segmentation, including left and right thalamus (dark green). Bottom panels: non-diffusion weighted image with the registered left thalamus (green) and the probabilistic connectivity distribution map of the corresponding thalamocortical tracts (red-yellow).

This probabilistic algorithm is capable of tracking fibers in the situation of multiple fiber orientations or crossing fibers. Its suitability has been specifically shown for tracking the thalamocortical projections[4]. Streamlines crossing through the midline of the brain – apart from the corpus callosum, fornix, and brainstem – were rejected using a midline mask. Subsequently,

a cerebral WM mask was applied to ensure no information from (sub)cortical or infratentorial structures was present in the thalamocortical tracts. An example of the resulting probabilistic connectivity distribution is depicted in **Figure 1**. This probabilistic connectivity distribution was used to obtain weighted mean FA and MD values per tract, in order to emphasize the major parts of the tract and decrease the effect of spurious streamlines. The entire tract (i.e. NAWM and lesions included) was used for further analysis.

Statistics

Statistical analysis was performed in IBM SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, N.Y., USA) and results were considered statistically significant upon $p<0.05$.
To ensure that baseline differences would minimally influence the longitudinal measures, changes over time were calculated as a relative change with regard to baseline values. For equal comparison between subjects with different follow-up periods, these changes over time were converted to annual rates.
All parameters were tested for normality with a Kolmogorov-Smirnov test. Differences between baseline and follow-up were calculated using a paired samples *t*-test (nominal data) or a Wilcoxon signed ranks tests (non-nominal data). Differences between subjects with CIS and RRMS and between the left and right brain hemispheres were calculated using independent *t*-test (nominal data), Mann-Whitney U-test (non-nominal data) or Chi-square test (binomial data). Forward linear regression was performed to see whether annual percentage change in thalamus volume could be predicted by baseline FA or MD of the thalamic tracts (left and right separately); and to see whether annual percentage change in FA or MD of the thalamic tracts could be predicted by baseline thalamus volume (left and right separately). All baseline demographics (age, sex, disease type and duration, treatment type and duration, EDSS) were included in the regression model as well, to correct for possible confounders.

Results

Demographics

From the total dataset of 82, ten subjects were excluded (received second line therapy $n=6$; movement observed during MRI $n=3$; erroneous lesion segmentation $n=1$) resulting in a total of 72 subjects with early MS with a mean disease duration of 10.7 ± 8.9 months (range: 0.8 to 31.1 months) and a mean follow-up time of 13.7 ± 4.8 months (range: 5.1 to 27.3 months). The baseline demographics are shown in **Table 1**. Subjects with CIS had a significantly shorter disease and treatment duration than subjects with RRMS, but there were no differences between the disease types for any of the other demographics.

Changes in brain volumes and WM tract integrity from baseline to follow-up

Baseline and follow-up brain volumes and white matter measures are shown in **Table 2**. None of the measures differed between subjects with CIS or RRMS, therefore all analyses were performed in the entire cohort of subjects.

Table 1 – Baseline demographics.

	Total *n*=72	**CIS** *n*=28	**RRMS** *n*=44	***p*-value**
Age in years, mean ± SD	37.4 ± 10.9	37.5 ± 9.4	37.4 ± 12.0	*p*=0.891[a]
Sex, m/f (%f)	28/44 (61%)	13/17 (57%)	15/27 (64%)	*p*=0.513[b]
Disease duration in months, mean ± SD (range)	10.7 ± 8.9 (0.8 – 31.1)	4.2 ± 3.7 (0.8 – 14.0)	15.2 ± 8.7 (1.1 – 31.1)	***p*<0.001**[a]
Treatment, *n* (%)				*p*=0.544[b]
none	17 (24%)	9 (30%)	8 (19%)	
interferon	34 (47%)	15 (50%)	19 (45%)	
glatiramer acetate	15 (21%)	4 (13%)	11 (26%)	
dimethyl fumarate	5 (7%)	2 (7%)	3 (7%)	
teriflunomide	1 (1%)	0 (0%)	1 (2%)	
Treatment duration in months, mean ± SD (range)	4.1 ± 6.9 (0.0 – 27.3)	1.2 ± 3.1 (0.0 – 13.1)	6.2 ± 8.1 (0.0 – 27.3)	***p*=0.011**[a]
EDSS, median with IQR	1.5 (1.5 – 2.5)	1.5 (1.5 – 2.5)	1.5 (1.0 – 2.5)	*p*=0.494[a]
Follow-up time in months, mean ± SD (range)	13.7 ± 4.8 (5.1 – 27.3)	12.5 ± 3.1 (7.9 – 25.4)	14.5 ± 5.7 (5.1 – 27.3)	*p*=0.346[a]

Abbreviations: SD = standard deviation; IQR = interquartile range. Differences between CIS and RRMS tested by: [a]Mann Whitney U test, [b]Chi Square test.

Table 2 – Baseline and follow-up brain volumes and the FA and MD of the tracts connected to the thalamus, shown as mean with standard deviation, as well as the annual percentage change observed.

	Baseline *n*=72	**Follow-up** *n*=72	***p*-value** paired samples *t*-test	**Annual change** (%, median with IQR)
Whole brain volume	1140.3 ± 126.7	1131.5 ± 124.6	***p* < 0.001**	-0.65 (-1.52 – -0.01)
WM volume	446.5 ± 58.6	441.7 ± 58.1	***p* < 0.001**	-0.83 (-2.42 – +0.44)
GM volume	637.4 ± 68.6	631.9 ± 67.4	***p* = 0.002**	-0.93 (-2.03 – +0.44)
Thalamus volume				
left	7.15 ± 0.91	7.06 ± 0.91	***p* < 0.001**	-1.34 (-2.14 – +0.43)
right	6.88 ± 0.84	6.78 ± 0.83	***p* < 0.001**	-0.88 (-2.34 – +0.14)
Lesion volume	12.26 ± 10.10	11.73 ± 10.44	*p* = 0.053	-8.52 (-27.65 – +8.87)
FA				
left	0.545 ± 0.027	0.546 ± 0.030	*p* = 0.624	+0.30 (-1.60 – +1.88)
right	0.560 ± 0.030	0.558 ± 0.030	*p* = 0.493	-0.23 (-2.93 – +2.33)
MD				
left	63.98 ± 3.63	64.39 ± 4.08	*p* = 0.354	+0.22 (-1.83 – +3.06)
right	62.32 ± 3.90	62.91 ± 4.14	*p* = 0.148	+0.61 (-2.58 – +3.71)

Legend: Volumes are in ml; MD is in 10^{-5} mm^2/s.

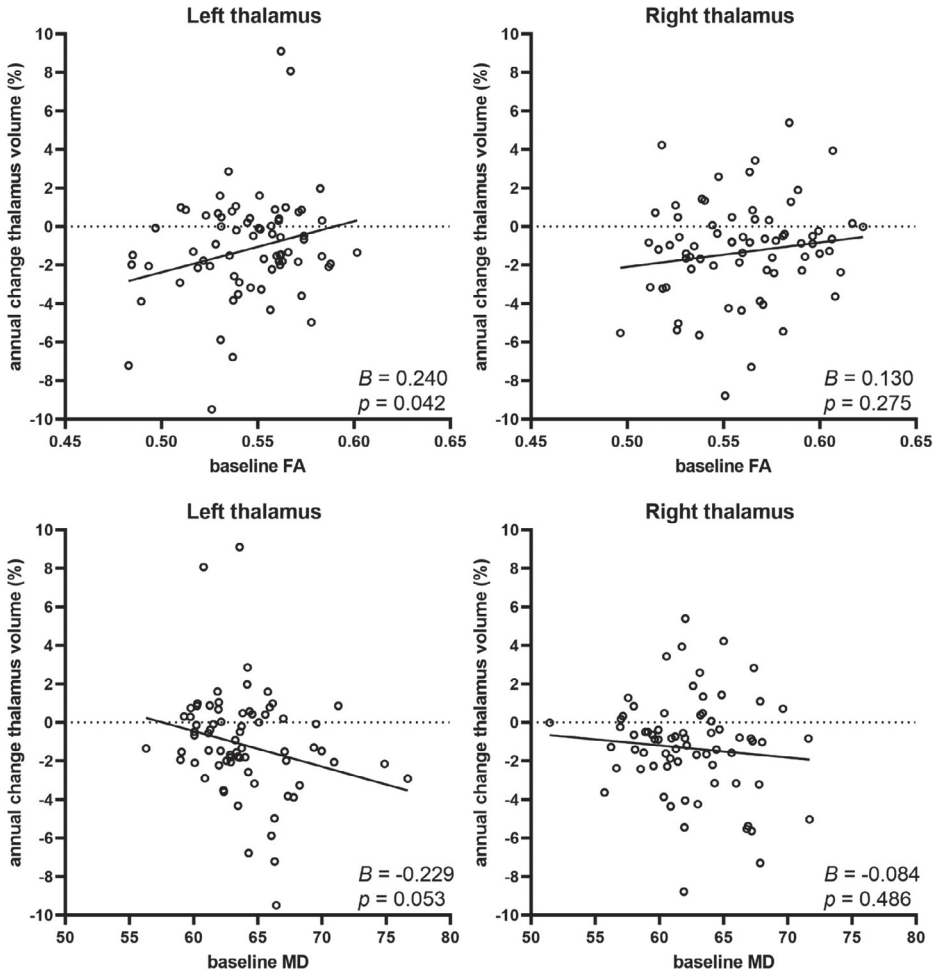

Figure 2 – Relation between baseline FA (top row) or baseline MD (depicted as 10^{-5} mm^2/s; bottom row) of the thalamocortical tracts and the annual percentage change in thalamus volume (left and right).

Lower baseline FA (B=0.240, p=0.042) was predictive of more negative annual percentage change of thalamus volume (i.e. atrophy), and a trend was seen for higher baseline MD (B=-0.229, p=0.053, not significant) with increasing thalamus atrophy, but only in the left thalamus.

Between baseline and follow-up, there were significant differences in whole brain volume (p<0.001), WM volume (p<0.001), GM volume (p=0.002), and left and right thalamus volume (both p<0.001). No significant differences over time were found for FA or MD of the thalamic tracts. Due to varying follow-up times between the subjects, differences between baseline and follow-up were converted to annual percentage change, shown in the last column of **Table 2**, showing large inter-subject variation for the annual percentage change of thalamus volume and FA/MD of the connected tracts.

Furthermore, differences between the left and right hemisphere were observed for thalamus volume, FA and MD at baseline (all p<0.001) and at follow-up (all p<0.001). However, no

significant differences between hemispheres were found for annual percentage change of thalamus volume, FA or MD.

Relation between thalamus volume and WM integrity of connected tracts

Linear regression analysis yielded a significant regression equation where annual percentage change in left thalamus volume was only predicted by baseline FA in the tracts connected to the left thalamus (B=0.240, p=0.042), and not by any of the demographic factors or by baseline MD, although a trend was seen for the latter (B=-0.229, p=0.053). The resulting regression equation showed that lower baseline FA of the left thalamic tracts was predictive of more negative annual percentage change in left thalamus volume (i.e. atrophy): $F(1,71)=4.284$, p=0.042 with adjusted R^2=0.044. Scatterplots showing the relationship between baseline FA and MD with annual percentage change in thalamus volume is shown for left and right thalamus in **Figure 2**. No significant regression equation could be computed for annual percentage change in right thalamus volume associated with baseline FA or MD, neither left nor right. Furthermore, we found no associations between the volumes of the thalami at baseline and annual percentage change in FA or MD. Thus, thalamus volume was not related to changes in white matter integrity over time.

Discussion

This study aimed to investigate the longitudinal association between thalamus volume and white matter damage in the connected tracts. We hypothesized that damage in the thalamocortical WM is a predictor of thalamic atrophy, but that lower baseline thalamus volume does not predict increased thalamocortical WM damage over time.

Our results show that relative thalamus atrophy could be predicted only by baseline FA of the connected tracts, and not by any of the demographic parameters. This was found only for the left, and not the right, thalamus. Relative change in FA or MD of the thalamic tracts could not be predicted by baseline thalamus volume, or any of the other demographic parameters, neither left nor right.

Whole brain volume loss around 0.5% per year in our cohort was consistent with previous research (for meta-analysis, see[35]), as was the annual rate of thalamus atrophy of around 1.0%[9,36] in early MS stages.

Previous cross-sectional studies showed a correlation between lower thalamus volumes and more damage in the white matter[13-15,37-39], but no causality could be extracted from these studies. Regression analysis showed that lower white matter integrity (i.e. lower FA, and a trend for higher MD) could predict increased thalamus atrophy over time in our cohort. This suggests that damage of thalamocortical WM could have an effect on thalamic atrophy. Although this relation between baseline WM integrity and thalamic atrophy was not very strong, it points toward a disease mechanism of retrograde degeneration in subjects with MS, which has been proposed before[17,20,40-42]. Beside the suggested phenomenon of Wallerian degeneration from distant WM lesions[43], our results in early MS patients with relatively low lesion volume, are in line with the indication of ongoing silent destructive processes of efferent and afferent thalamic projections which might induce measurable atrophy[14].

Since baseline thalamus volume and FA and MD of the connected tracts differed significantly between both hemispheres, this might explain why the relationship described above was only found for the left, and not right, thalamus. Leftward thalamic asymmetry could be demonstrated in healthy individuals and seems to corresponds to the physiological conditions[44]. Interhemispheric asymmetry of brain diffusivity was described in normal individuals[45], as well as in MS, whereby increased apparent diffusion coefficients (ADCs) in the left thalamus were described[46], which is in line with our results demonstrating lower FA and higher MD values in the left thalamus. Longer follow-up periods are essential to further study the difference observed here between left and right thalamic tracts.

This study has some limitations. Since we had no healthy control group, we were not able to distinguish atrophy due to normal ageing from pathological atrophy, although previous studies have shown that normal ageing shows a lower atrophy rate (around 0.2 to 0.5% per year) than seen in our early MS cohort[47].

Second, we did not investigate normal appearing white matter and lesions separately, because this would lead to additional statistical tests and loss of power. However, we defined tract integrity by using FA and MD-values weighted by connectivity probability, such that the average weighted value is representative for the whole tract without lesion exclusion. Further, tensor elements were estimated with a slightly higher variance than ideal, since the DTI sequence contained one volume (2 averages) without diffusion weighting (b=0 s/mm^2), instead of an optimal number of 3 or 4 averages, considering the 32 volumes with diffusion weighting[48].

Last, this study was mainly explorative to find directionality and causality in the relationship between WM damage and GM atrophy. Longer follow-up is needed to confirm these results and to elucidate the differences between left and right thalamus. Future studies including larger sample sizes and extended follow-up could also be suited to elucidate the pathophysiological background in greater detail by investigating radial diffusivity and axial diffusivity as indicators for myelination and axonal damage.

In conclusion, this preliminary study in early MS with a short follow-up indicates that thalamus atrophy progresses faster upon more baseline damage in the connected thalamocortical white matter tracts. No significant relation was found for annual percentage change in FA or MD with thalamus volumes at baseline. However, the effect is small and therefore results should be confirmed over longer follow-up periods.

Funding

This work was supported by the German Federal Ministry for Education and Research, BMBF, German Competence Network Multiple Sclerosis (KKNMS), grant no.01GI1601I and the Dutch MS Research Foundation, grant number 14-876.

FB is supported by the NIHR biomedical research centre at UCLH.

Declaration of interest

The authors declare that there was no conflict of interest related to this research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the local ethics committee of Ruhr-University Bochum (Approval No. 3714-10). The patients/participants provided their written informed consent to participate in this study and for the use and publication of their anonymized images and data.

References

1 Jones, D. K. & Cercignani, M. Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed 23, 803-820, doi:10.1002/nbm.1543 (2010).

2 Jones, D. K., Knosche, T. R. & Turner, R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. Neuroimage 73, 239-254, doi:10.1016/j.neuroimage.2012.06.081 (2013).

3 Minagar, A. et al. The thalamus and multiple sclerosis: modern views on pathologic, imaging, and clinical aspects. Neurology 80, 210-219, doi:10.1212/WNL.0b013e31827b910b (2013).

4 Behrens, T. E. et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature neuroscience 6, 750-757, doi:10.1038/nn1075 (2003).

5 Cifelli, A. et al. Thalamic neurodegeneration in multiple sclerosis. Ann Neurol 52, 650-653, doi:10.1002/ana.10326 (2002).

6 Wylezinska, M. et al. Thalamic neurodegeneration in relapsing-remitting multiple sclerosis. Neurology 60, 1949-1954 (2003).

7 Rocca, M. A. et al. Thalamic damage and long-term progression of disability in multiple sclerosis. Radiology 257, 463-469, doi:10.1148/radiol.10100326 (2010).

8 Eshaghi, A. et al. Progression of regional grey matter atrophy in multiple sclerosis. Brain 141, 1665-1677, doi:10.1093/brain/awy088 (2018).

9 Azevedo, C. J. et al. Thalamic atrophy in multiple sclerosis: A magnetic resonance imaging marker of neurodegeneration throughout disease. Ann Neurol 83, 223-234, doi:10.1002/ana.25150 (2018).

10 Schoonheim, M. M. et al. Subcortical atrophy and cognition: sex effects in multiple sclerosis. Neurology 79, 1754-1761, doi:10.1212/WNL.0b013e3182703f46 (2012).

11 Batista, S. et al. Basal ganglia, thalamus and neocortical atrophy predicting slowed cognitive processing in multiple sclerosis. J Neurol 259, 139-146, doi:10.1007/s00415-011-6147-1 (2012).

12 Riccitelli, G. et al. Mapping regional grey and white matter atrophy in relapsing-remitting multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England) 18, 1027-1037, doi:10.1177/1352458512439239 (2012).

13 Henry, R. G. et al. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. J Neurol Sci 282, 61-66, doi:10.1016/j.jns.2009.02.379 (2009).

14 Deppe, M. et al. Early silent microstructural degeneration and atrophy of the thalamocortical network in multiple sclerosis. Hum Brain Mapp 37, 1866-1879, doi:10.1002/hbm.23144 (2016).

15 Pontillo, G. et al. Determinants of Deep Gray Matter Atrophy in Multiple Sclerosis: A Multimodal MRI Study. Am J Neuroradiol 40, 99-106, doi:10.3174/ajnr.A5915 (2019).

16 Johnen, A. et al. Resolving the cognitive clinico-radiological paradox - Microstructural degeneration of fronto-striatal-thalamic loops in early active multiple sclerosis. Cortex 121, 239-252, doi:10.1016/j.cortex.2019.08.022 (2019).

17 Haider, L. et al. Multiple sclerosis deep grey matter: the relation between demyelination, neurodegeneration, inflammation and iron. J Neurol Neurosur Ps 85, 1386-1395, doi:10.1136/jnnp-2014-307712 (2014).

18 Steenwijk, M. D. et al. Unraveling the relationship between regional gray matter atrophy and pathology in connected white matter tracts in long-standing multiple sclerosis. Hum Brain Mapp 36, 1796-1807, doi:10.1002/hbm.22738 (2015).

19 Bodini, B. et al. Exploring the Relationship Between White Matter and Gray Matter Damage in Early Primary Progressive Multiple Sclerosis: An In Vivo Study With TBSS and VBM. Human Brain Mapping 30, 2852-2861, doi:10.1002/hbm.20713 (2009).

20 Calabrese, M. et al. Exploring the origins of grey matter damage in multiple sclerosis. Nat Rev Neurosci 16, 147-158, doi:10.1038/nrn3900 (2015).

21 Kompetenznetz. Kompetenznetz Multiple Sklerose. Available from: http://www.kompetenznetz-multiplesklerose.de/en/, <http://www.kompetenznetz-multiplesklerose.de/en/> (

22 von Bismarck, O. et al. Treatment choices and neuropsychological symptoms of a large cohort of early MS. Neurol Neuroimmunol Neuroinflamm 5, e446, doi:10.1212/NXI.0000000000000446 (2018).

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

24 Schneider, R. et al. Temporal Dynamics of Diffusion Metrics in Early Multiple Sclerosis and Clinically Isolated Syndrome: A 2-Year Follow-Up Tract-Based Spatial Statistics Study. Front Neurol 10, 1165, doi:10.3389/fneur.2019.01165 (2019).

25 Smith, S. M. Fast robust automated brain extraction. Hum Brain Mapp 17, 143-155, doi:10.1002/hbm.10062 (2002).

26 Popescu, V. et al. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. Neuroimage 61, 1484-1494, doi:10.1016/j.neuroimage.2012.03.074 (2012).

27 Schmidt, P. et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. Neuroimage 59, 3774-3783, doi:10.1016/j.neuroimage.2011.11.032 (2012).

28 Schmidt, P. Bayesian Inference for Structured Additive Regression Models for Large-scale Problems with Applications to Medical Imaging, LMU München, (2017).

29 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179-194, doi:10.1006/nimg.1998.0395 (1999).

30 Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195-207, doi:10.1006/nimg.1998.0396 (1999).

31 Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402-1418, doi:10.1016/j.neuroimage.2012.02.084 (2012).

32 Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 125, 1063-1078, doi:10.1016/j.neuroimage.2015.10.019 (2016).

33 Behrens, T. E. et al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 50, 1077-1088, doi:10.1002/mrm.10609 (2003).

34 Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F. & Woolrich, M. W. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34, 144-155, doi:10.1016/j.neuroimage.2006.09.018 (2007).

35 Vollmer, T. et al. The natural history of brain volume loss among patients with multiple sclerosis: a systematic literature review and meta-analysis. J Neurol Sci 357, 8-18, doi:10.1016/j.jns.2015.07.014 (2015).

36 Zivadinov, R. et al. Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years. AJNR Am J Neuroradiol 34, 1931-1939, doi:10.3174/ajnr.A3503 (2013).

37 Bergsland, N. et al. Subcortical and cortical gray matter atrophy in a large sample of patients with clinically isolated syndrome and early relapsing-remitting multiple sclerosis. AJNR Am J Neuroradiol 33, 1573-1578, doi:10.3174/ajnr.A3086 (2012).

38 Muhlau, M. et al. White-matter lesions drive deep gray-matter atrophy in early multiple sclerosis: support from structural MRI. Multiple sclerosis (Houndmills, Basingstoke, England) 19, 1485-1492, doi:10.1177/1352458513478673 (2013).

39 Amiri, H. et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin 19, 466-475, doi:10.1016/j.nicl.2018.04.023 (2018).

40 Steenwijk, M. D. et al. What explains gray matter atrophy in long-standing multiple sclerosis? Radiology 272, 832-842, doi:10.1148/radiol.14132708 (2014).

41 Balk, L. J. et al. A dam for retrograde axonal degeneration in multiple sclerosis? J Neurol Neurosurg Psychiatry 85, 782-789, doi:10.1136/jnnp-2013-306902 (2014).

42 Haider, L. et al. The topograpy of demyelination and neurodegeneration in the multiple sclerosis brain. Brain 139, 807-815, doi:10.1093/brain/awv398 (2016).

43 Louapre, C. et al. Heterogeneous pathological processes account for thalamic degeneration in multiple sclerosis: Insights from 7 T imaging. Multiple sclerosis (Houndmills, Basingstoke, England) 24, 1433-1444, doi:10.1177/1352458517726382 (2018).

44 Guadalupe, T. et al. Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex. Brain imaging and behavior 11, 1497-1514, doi:10.1007/s11682-016-9629-z (2017).

45 Fabiano, A. J., Horsfield, M. A. & Bakshi, R. Interhemispheric asymmetry of brain diffusivity in normal individuals: a diffusion-weighted MR imaging study. AJNR. American journal of neuroradiology 26, 1089-1094 (2005).

46 Fabiano, A. J. et al. Thalamic involvement in multiple sclerosis: a diffusion-weighted magnetic resonance imaging study. Journal of neuroimaging : official journal of the American Society of Neuroimaging 13, 307-314 (2003).

47 Hedman, A. M., van Haren, N. E., Schnack, H. G., Kahn, R. S. & Hulshoff Pol, H. E. Human brain changes across the life span: a review of 56 longitudinal magnetic resonance imaging studies. Hum Brain Mapp 33, 1987-2002, doi:10.1002/hbm.21334 (2012).

48 Jones, D. K., Horsfield, M. A. & Simmons, A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn Reson Med 42, 515-525 (1999).

CHAPTER 4.2

Multimodal MRI study on the relation between WM integrity and connected GM atrophy and its effect on disability in early multiple sclerosis

Merlin M. Weeda
David R. van Nederpelt
Jos W.R. Twisk
Iman Brouwer
Joost P.A. Kuijer
Maureen van Dam
Hanneke E. Hulst
Joep Killestein
Frederik Barkhof
Hugo Vrenken
Petra J.W. Pouwels

Journal of Neurology, in press (2023)

Abstract

Background

Multiple sclerosis (MS) is characterized by pathology in white matter(WM) and atrophy of grey matter (GM), but it remains unclear how these processes are related, or how they influence clinical progression.

Objective

To study the spatial and temporal relationship between GM atrophy and damage in connected WM in relapsing-remitting (RR)MS in relation to clinical progression.

Methods

Healthy control (HC) and early RRMS subjects visited our center twice with a 1-year interval for MRI and clinical examinations, including the Expanded Disability Status Scale (EDSS) and Multiple Sclerosis Functional Composite (MSFC) scores. RRMS subjects were categorized as MSFC decliners or non-decliners based on ΔMSFC over time. Ten deep (D)GM and 62 cortical (C)GM structures were segmented and probabilistic tractography was performed to identify the connected WM. WM integrity was determined per tract with, amongst others, fractional anisotropy (FA), mean diffusivity (MD), neurite density index (NDI), and myelin water fraction (MWF). Linear mixed models (LMMs) were used to investigate GM and WM differences between HC and RRMS, and between MSFC decliners and non-decliners. LMM was also used to test associations between baseline WM z-scores and changes in connected GM z-scores, and between baseline GM z-scores and changes in connected WM z-scores, in HC/RRMS subjects and in MSFC decliners/non-decliners.

Results

We included 13 HCs and 31 RRMS subjects with an average disease duration of 3.5 years and a median EDSS of 3.0. Fifteen RRMS subjects showed declining MSFC scores over time, and they showed higher atrophy rates and greater WM integrity loss compared to non-decliners. Lower baseline WM integrity was associated with increased CGM atrophy over time in RRMS, but not in HC subjects. This effect was only seen in MSFC decliners, especially when an extended WM z-score was used, which included FA, MD, NDI and MWF. Baseline GM measures were not significantly related to WM integrity changes over time in any of the groups.

Discussion

Lower baseline WM integrity was related to more cortical atrophy in RRMS subjects that showed clinical progression over a 1-year follow-up, while baseline GM did not affect WM integrity changes over time. WM damage therefore seems to drive atrophy more than conversely.

Introduction

In multiple sclerosis (MS), the most common pathological brain changes are widespread pathology in the white matter (WM) and atrophy of the grey matter (GM)[1]. Although GM atrophy shows stronger associations with clinical dysfunction than WM atrophy[2], large variability between patients' atrophy rates and disability progression have been reported[3,4]. It has been suggested that this inter-patient variability of GM atrophy rates arises to a large part because of the different distribution and severity of WM damage between patients, indicating the importance of looking at the GM-WM relationship in anatomically connected regions[5-8]. However, most studies investigating this had a cross-sectional design and could therefore not draw any conclusions on whether GM atrophy precedes of follows WM damage, and how this effects disability progression in MS.

A recent systematic review by Lie *et al*.[9] that included 90 studies on the relationship between WM lesions and GM volume showed an inverse association between the two, particularly in early (relapsing) MS, and less so in progressive MS, suggesting that GM neurodegeneration is mostly secondary to WM damage in the form of lesions in early stages of the disease. Still, a knowledge gap is present since WM damage in MS is not confined to focal lesions, but to microstructural damage as well. Microstructural damage in the WM can be assessed indirectly through quantitative magnetic resonance (MR) measures, such as fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD), which are simplified measures obtained from the diffusion tensor for the overall integrity of the WM[10]. Multi-shell diffusion weighted imaging (DWI) can provide more biophysical properties such like neurite orientation dispersion and density imaging (NODDI)[11], thereby providing more information on microstructural damage in the WM. Recent studies have shown that NODDI parameters such as neurite density index (NDI) and orientation dispersion index (ODI) enable additional characterization of the WM in MS subjects[12,13]. In particular, NDI can be considered an axonal marker[14]. Other imaging techniques that can be used for further characterization of damage in the WM are quantitative susceptibility mapping (QSM) and myelin water imaging (MWI). QSM may enable us to visualize WM damage and lesion formation before conventional structural imaging could[15], and from MWI the myelin water fraction (MWF) can be calculated, which can provide information on the myelin content of the WM[16]. Combining these imaging measures may enable us to visualize more subtle WM damage patterns that occur in the early stages of the disease.

In this study, we aimed to investigate whether in early RRMS, the amount and/or type of damage in the WM tracts connected to the GM influences neurodegeneration, or whether damage in the GM influences damage in the connected WM tracts over time, and how this may relate to disability. For this, we studied longitudinal multimodal imaging data to better understand the spatiotemporal relationship between WM and GM damage in early RRMS.

Methods

Subjects

The institutional review board approved the study protocol and all participants gave written informed consent prior to participation, according to the Declaration of Helsinki. To be included in the study, patients had to be diagnosed with clinically definite RRMS[17] and have a disease duration of no more than five years. They could be included if they were using no treatment or first-line treatment, but not if they were using more advanced treatment. Their clinical disability levels had to be limited, with a maximum allowed EDSS score of 5.0. In case of switching of treatment, MRI examinations were planned with at least 4–6 months delay[18]. When steroids were used, MRI was delayed by 3 months[19]. Patients were excluded (over the course of the study) in case of (switching to) second-line treatment to avoid spurious (pseudo)atrophy effects. Patients were seen at one-year intervals, for extensive MR imaging and evaluation of clinical and neuropsychological performance.

A group of age-, gender- and education-matched healthy controls (HCs) was also included and seen at one-year intervals for extensive MR imaging Exclusion criteria for both MS and HC subjects were inability to undergo MRI examination; and past or current clinically relevant neurological, psychiatric or (auto)immune disorders other than MS. The data used in the current study was part of a larger cohort study, some results of which have been reported previously[20].

MR imaging - acquisition

This was a single-center study in which a single MR scanner was used without intermediate upgrades. Imaging was performed on a 3T whole-body scanner (Discovery MR750, GE Healthcare, Milwaukee, WI., USA) and an eight-channel phased-array head coil. The MR protocol included a sagittal 3D T1-weighted fast spoiled gradient echo sequence (FSPGR with TR/TE/TI = 8.2/3.2/450 ms and 1.0 mm isotropic resolution) and a sagittal 3D T2-weighted fluid attenuated inversion recovery sequence (FLAIR with TR/TE/TI = 8000/130/2338 ms at resolution 1.0x1.0x1.2 mm).

Axial 2D DWI acquisitions covering the entire brain (echo planar imaging, TR/TE = 6200/86 ms with 2.0 mm isotropic resolution) with accompanying reference scans with reversed phase-encoding direction were performed. The multi-shell DWI consisted of 7 volumes without diffusion weighting (b=0 s/mm^2) and 88 volumes with non-colinear diffusion gradients (29 images with b=500 s/mm^2 and 59 images with b=2000 s/mm^2). The b-values were interleaved to improve post-processing steps like motion correction[21].

Quantitative susceptibility mapping (QSM) images were acquired using an axial 3D multiple echo GRE sequence, with 7 TEs (TR/TEstart/TEdelta/TEmax = 49.24/5.44/6.55/44.73 ms at resolution 0.5x0.5x1.6 mm).

Lastly, multicomponent driven equilibrium single pulse observation of T1 and T2 (mcDESPOT) was acquired in an axial 3D slab with 2.5 mm isotropic resolution, using a spoiled gradient recalled sequence (SPGR with TR/TE = 8/3.02 ms with multiple flip angles [3,4,5,6,7,9,13,18 degrees]), a balanced steady-state free precession sequence (SSFP with TR/TE = 4.736/2.368 ms, with multiple flip angles [12,16,21,27,33,40,51,65 degrees], with phases 0 and 180 degrees

for the odd flip angles and phases 90 and 270 degrees for the even flip angles), and a 2D axial B1 map (TR/TE = 17/12.8 ms at lower resolution 4.0x4.0x5.0 mm). High-order shimming was performed for multi-shell DWI, QSM, and mcDESPOT acquisitions.

MR imaging – analysis

An overview of the different MRI post-processing streams used in this study is shown in **Figure 1** and analyses are described below.

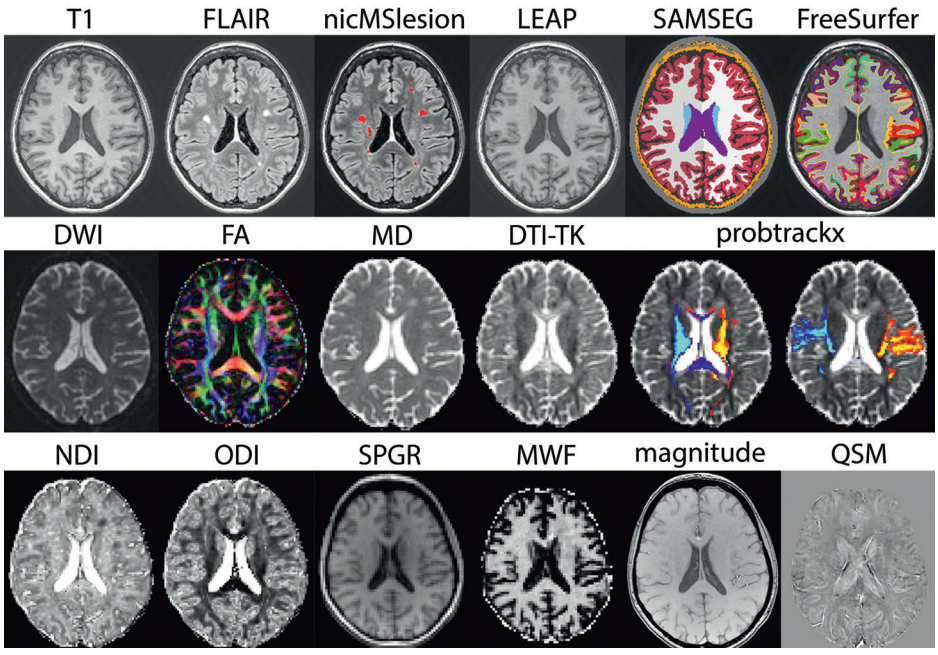

Figure 1 – Overview of MR imaging analysis.

Top row: T1 and FLAIR images were used to create a lesion mask (nicMSlesions) and lesion filled (LEAP) image, which was used as input for longitudinal SAMSEG for subcortical segmentation and longitudinal FreeSurfer for cortical parcellation. Middle row: multi-shell diffusion weighted images (DWI) were used to obtain FA and MD in the WM mask. A tensor image from DTI-TK was used to register FreeSurfer ROIs, and probtrackx was used for tractography (example from thalamus tracts and postcentral tracts). Bottom row: NODDI was used to obtain NDI and ODI maps; SPGR images were used to create MWF maps; magnitude and phase (not shown) images were used to create QSM maps.

Structural imaging analysis

Brain extraction was performed using FMRIB Software Library (FSL) version 5.0.10 brain extraction tool BET[22,23] and N3 bias field correction with 3T optimized parameters was performed with FreeSurfer version 6.0[24]. Lesion segmentation was performed with the deep-learning algorithm nicMSlesions version 0.2[25,26] which was optimized for our data in an earlier study[27]. In summary, full re-training of the nicMSlesions neural network (11 layers) was done with the use of manual lesion segmentations available from fourteen subjects with MS. All

parameters were set at default. On the resulting lesion probability map, the optimized probability threshold of 0.4 was applied and lesions smaller than five voxels were removed.
Lesion filling was performed with LEsion Automated Processing (LEAP)[28] and lesion filled images were processed with the longitudinal pipeline of FreeSurfer[29-31], using a template-driven approach to provide a detailed parcellation and segmentation of the cortex and subcortical structures[32,33]. Cortical thickness was obtained from FreeSurfer and brain volumes were obtained from the longitudinal pipeline of Sequence Adaptive Multimodal SEGmentation (SAMSEG)[34,35]. To control for differences in head size, whole brain (WB), WM and DGM brain volumes were normalized and calculated as a percentage of the segmentation-based total intracranial volume (normalized WB [nWB], normalized WM [nWM], and normalized DGM [nDGM], respectively). In addition, nDGM volume and CGM thickness were converted into z-scores for each structure separately, based on the entire cohort. In our analysis, two DGM structures (nucleus accumbens and amygdala) and three CGM structures (entorhinal cortex, frontal pole and temporal pole) were excluded from analyses due to their known measurement variabilities[36].

Diffusion imaging analysis

All images were corrected for susceptibility induced geometric distortions using FSL topup[37,38], and for movement and eddy currents with FSL eddy[39]. The diffusion tensor was fitted on all b=0 mm/s^2 and b=2000 s/mm^2 images to obtain the FA, MD, AD and RD values.
The multi-shell ball and stick model of bedpostx[40] was used to estimate the voxel-wise diffusion parameter distribution. Probabilistic tractography using 5000 streamlines was performed with probtrackx2[41-43]. After linear registration of DWI to 3DT1, the inverse transformation was used to register the FreeSurfer regions of interest (ROIs) to DWI space using nearest-neighbour interpolation; thus per hemisphere, five DGM and 31 CGM regions were available as seed and target ROIs for tractography. A midline exclusion mask was used to ensure tracts could not cross hemispheres, except through the corpus callosum, fornix and brainstem. Moreover, the target ROIs were used as both waypoint and termination masks.
Probabilistic tractography was performed on year-1 scans, after which the tracts were propagated to the year-2 scans using Diffusion Tensor Imaging ToolKit (DTI-TK) version 2.3.1[44,45], to create a spatially normalized tensor within-subjects template. DTI-TK makes use of an affine registration algorithm with explicit tensor reorientation optimization[46]. The longitudinal pipeline as proposed by Keihaninejad *et al.*[47] was used. A DTI-TK within-subject template was created for each subject from the b=2000 mm/s^2 images.. With these templates, the year-1 probabilistic connectivity distribution was registered to year-2 images for each subject.
Neurite Orientation Dispersion and Density Index (NODDI) Matlab Toolbox version 1.0.1 was used to obtain the neurite density index (NDI) and orientation dispersion index (ODI). Voxels for which the NODDI model failed (i.e. all voxels with f_{iso}>0.99; f_{icvf}>0.99; kappa=0.05; and all error-code voxels)[11] were filtered out to create a tissue mask image to apply to the NDI and ODI images.

QSM imaging analysis

SuscEptibility mapping PIpeline tool for phAse image (SEPIA) version 0.7.2[48] containing the Susceptibiltiy Tensor Imaging (STI) Suite[49] in MATLAB (The MathWorks, Natick, MA) was

used for processing of magnitude and phase images after applying a brain mask obtained with FSL BET. After phase unwrapping with the optimum weights-laplacian based method and background field removal with variable radius kernel, QSM images were estimated with iterative sparse linear equation and least square (iLSQR) with susceptibility of the whole brain defined as 0 parts per million (ppm). QSM maps were registered to DWI space of the corresponding time point.

Myelin water imaging analysis

From B1, SPGR and SSFP images, myelin water fraction (MWF) maps were estimated using qimcdespot (from Quantitative Imaging Tool [QUIT][50] version 2.0.2), assuming a three-component model with exchange. Upper and lower bounds were slightly adapted based on inspection of histograms and in-house optimization within a group of HC subjects. MWF maps were registered to DWI space of the corresponding time point.

WM measures in probabilistic tracts

The probabilistic tracts were multiplied by the binary WM mask, and for all measures (i.e. FA, MD, AD, RD, NDI, ODI, MWF and QSM), mean values per tract were extracted, weighted by the connectivity probability to emphasize the tract center, and decrease the effect of spurious tracts. We determined all WM measures for the whole tract, consisting of normal-appearing (NA)WM and lesional WM, but also perilesional WM.

In order to combine WM measures, z-scores were calculated per tract for each WM measure, based on the entire cohort, and z-scores were combined depending on which information was present. We calculated a diffusion z-score (WM-Diffusion = (zFA-zMD-zRD)/3) as well as an extended WM z-score (WM-Extended = (zFA-zMD-zRD+zNDI+zMWF)/5), when data was available.

Clinical and neuropsychological evaluation

Patients' medical history was taken, including the occurrence of relapses and changes in therapy. Neuropsychological evaluation included the Symbol Digit Modalities Test (SDMT)[51] to measure information processing speed. Two parallel test versions were used to minimize learning effects. SDMT scores were corrected for age, sex and educational level and transformed into z-scores, based on a normative sample of Dutch healthy controls (n=407)[52].

Patients' physical disability was measured with the Expanded Disability Status Scale (EDSS) questionnaire[53], the 9 Hole Peg Test (9-HPT) and the 25 Foot Walk Test (25-FWT)[54]. From these three, EDSS+ progression was determined as a descriptive, binary, marker of physical disability progression[55].

To also take cognitive disability into account in this group of early RRMS subjects with relatively low EDSS scores, Multiple Sclerosis Functional Composite (MSFC)[54] scores were calculated from 9-HPT, 25-FWT and SDMT z-scores. For each of the three sections, results were considered impaired upon z-score ≤ -1.5. Over time, RRMS subjects were classified as MSFC decliner when MSFC scores were lower at year-2 compared to year-1, and as MSFC non-decliner when MSFC scores were equal or higher at year-2 compared to year-1.

Statistics

HC vs RRMS comparisons for baseline demographics were performed by independent samples *t*-test, Mann Whitney *U*-test, or Chi Square test, when appropriate. Differences in the RRMS group over time were analyzed with paired *t*-test, Wilcoxon Signed Ranks test, or McNemar test, when appropriate.

To analyze HC and RRMS differences over time in normalized brain volumes, cortical thickness, and z-scores for nDGM, CGM or WM, linear mixed models (LMM) with subject as random intercept were used, with fixed factors time, type (i.e. HC/RRMS) and time*type. When appropriate, LMM analysis was also performed to assess baseline differences between the two groups. Since application of an LMM requires constant variance in errors, z-scores for both WM and GM measures were used.

LMM analysis with subject as random intercept and type (i.e. HC/RRMS or MSFC non-decliner/decliner) as fixed factor was used for longitudinal analysis of DGM/CGM z-score relations with WM z-scores in the connected tracts. For the longitudinal relations, we analyzed both the effect of baseline WM z-scores on change in the connected GM z-scores (i.e. baseline WM to ΔGM), and the effect of baseline GM z-scores on change in WM z-scores in the connected tracts (i.e. baseline GM to ΔWM); both with type (HC/RRMS) or MSFC group (declining/non-declining) as covariates. When an interaction effects were present, the cohort was split based on type or MSFC group and LMM analysis were performed as described above. Analyses for DGM and CGM were performed separately.

Exploratory binary logistic regression was used to predict whether a subject would belong to the MSFC non-declining or MSFC declining group over time by looking at baseline WM or baseline GM values only; as well as combining these measures with demographics such as sex, age, education, treatment type and EDSS at baseline. Statistical analyses were performed using SPSS26 (IBM SPSS, Chicago, USA) and *p*-values were considered statistically relevant upon $p \leq 0.05$. Since LMM analysis was performed multiple times (i.e. for every GM or WM measure as possible outcome separately), *p*-values were considered statistically significant upon $p \leq 0.01$.

Results

Demographics

In total, 40 subjects with early RRMS and 15 age-and-sex matched HCs were included in the study. A total of 11 subjects did not complete both year-1 and year-2 measurements and were therefore excluded from analyses in the current study (two HC subjects [inability to undergo MRI examination] and nine RRMS subjects [$n=5$ switch to second-line therapy; $n=1$ switch to first-line therapy just prior to examination; $n=3$ inability to undergo MRI examination unrelated to RRMS]). This resulted in a total of 44 age, sex, and education matched subjects (13 HC, 31 RRMS) with full data available for analysis, for whom demographics are depicted in **Table 1**. Follow-up time for RRMS and HC subjects was similar (1.00±0.10 years and 0.99±0.10 years, respectively).

Table 1 – Demographics for HC and RRMS subjects; and RRMS clinical characteristics over time.

Baseline demographics	**HC** *n*=13	**RRMS** *n*=31
Age in years, mean ± SD	37.5 ± 12.8	37.3 ± 7.5
Sex, m/f (% m)	4/9 (31)	7/24 (23)
Education level[1], median (IQR)	6 (5 – 7)	6 (5 – 7)
Follow-up time in years, mean ± SD	0.99 ± 0.10	1.00 ± 0.10
Disease duration in years, mean ± SD	-	3.5 ± 1.4
ARR		
median (range)	-	0 (0 – 2)
≥ 1 relapse in previous year, *n* (%)		9 (29)
Treatment[2], *n* (%)		
none	-	10 (32)
interferon		4 (13)
glatiramer acetate		5 (16)
dimethyl fumarate		11 (36)
teriflunomide		1 (3)
Treatment duration in years, mean ± SD	-	2.9 ± 1.1
EDSS		
median (range)	3.0 (0 – 5.0)	3.0 (0 – 6.0)
EDSS+ progression, *n* (%)	-	6 (19)
9-HPT		
in seconds, mean ± SD	19.66 ± 2.95	19.44 ± 3.22
z-score, mean ± SD	-	0.108 ± 1.09
25-FWT		
in seconds, mean ± SD	4.40 ± 1.25	4.90 ± 2.19
z-score, mean ± SD	-	-0.412 ± 1.785
SDMT		
n correct, mean ± SD	65 ± 14	66 ± 15
z-score, mean ± SD	-0.387 ± 1.284	-0.157 ± 1.289
MSFC score		
mean ± SD	-0.129 ± 0.696	-0.153 ± 0.880
≥1 impaired section, *n* (%)	10 (32)	7 (23)
MSFC decline over time, *n* (%)	n.a.	15 (48)

Abbreviations: SD = standard deviation; IQR = interquartile range; ARR = annualized relapse rate; EDSS = expanded disability status scale; 9-HPT = 9-hole peg test; 25-FWT = 25-foot walk test; SDMT = symbol digit modalities test; MSFC = multiple sclerosis functional composite.
Statistics: * $p<0.05$. Z-scores from 9-HPT and 25-FWT were calculated from year-1 data, and are therefore not depicted at year-1 since they will, by default, have a mean of 0 and a SD of 1.
Foot notes: [1] Education level according to Verhage scale (0 to 7); [2] Treatment group interferon consists of interferon beta-1a (Avonex®, Rebif®), beta-1b (Betaferon®) and peginterferonbeta.

RRMS subjects had an average disease duration of 3.5 ± 1.4 years at baseline measurements with median EDSS 3.0 (range 0.0 – 5.0). A total of 6 subjects (19%) showed EDSS+ progression during follow-up. 9-HPT and SDMT scores did not significantly differ from year-1 to year-2, but worsening of 25-FWT results was seen (Z=-2.930, p=0.003). A total of 10 subjects (32%) had at least 1 impaired MSFC section (i.e. 9-HPT, 25-FWT and/or SDMT) at year-1, and 7 subjects (23%) at year-2. Total MSFC score did not change significantly over time, but a total of *n*=15

subjects showed an overall decline in MSFC score from year-1 to year-2 (MSFC decliners), whereas *n*=16 did not (MSFC non-decliners). No significant differences in age, sex, education, disease duration, treatment time/type, or number of relapses in the previous year were found between the MSFC decliners and non-decliners.

MRI characteristics

MWI data was not available for 2 of 13 HC subjects and 7 of 31 RRMS subjects at year-1 (MSFC non-decliners *n*=2 and MSFC decliners *n*=5, respectively), and for one RRMS subject (MSFC non-decliner) at year-2. QSM data was not available for one RRMS subject (MSFC decliner) at year-1, and for one RRMS subject (MSFC non-decliner) at year-2.

HC vs RRMS subjects

MRI characteristics for HC and RRMS subjects over time are shown in **Table 2**. Larger atrophy rates in RRMS compared to HC subjects were found for normalized WB, WM and DGM volumes (p=0.024, p=0.018 and p=0.026, respectively). RRMS subjects showed generally lower z-scores of DGM volume, CGM thickness, and measures of WM integrity (e.g. lower FA, higher MD; or more specifically lower NDI as axonal damage marker, and lower MWF as myelin damage marker), but no significant differences were found between HC and RRMS, neither cross-sectionally nor longitudinally.

In both DGM- and CGM-seeded tracts, WM-Extended z-scores decreased less over time in RRMS compared to HC subjects (DGM: B=0.290, SE=0.090, p<0.001; CGM: B=0.122, SE=0.037, p<0.001), but the WM-Extended z-scores were also already lower at year-1 in RRMS compared to HC subjects (DGM: B=0.464, SE=0.200, p=0.024; CGM: B=0.358, SE=0.190, p=0.066), although not reaching significance.

MSFC non-declining vs MSFC declining subjects

Comparing the RRMS subjects based on their declining or non-declining MSFC score over time (**Table 3**), we found significantly higher atrophy rates for normalized WB, WM, GM and DGM volumes (p=0.013, p=0.029, p=0.022 and p=0.021, respectively) in the MSFC declining group compared to non-declining group. CGM thickness as well as CGM z-score also decreased faster over time in MSFC declining compared to non-declining subjects (p=0.010 and p=0.019, respectively).

In both DGM and CGM tracts, several WM z-scores declined faster in MSFC declining compared to non-declining subjects (e.g. WM-Extended: B=0.259, SE=0.094, p=0.006 for DGM, and B=0.176, SE=0.040, p<0.001 for CGM, respectively). Interestingly, this was highly significant for changes in myelin-related measures (MWF and RD), but not for the axonal marker NDI. For the CGM, but not DGM, overall WM integrity was significantly lower at both time points in MSFC declining subjects compared to MSFC non-declining subjects (e.g. WM-Extended: B=-0.436, SE=0.204, p=0.040).

Table 2 – MRI characteristics of HC and RRMS subjects over time.

	HC *n*=13		RRMS *n*=31		Statistics Time * Type interaction
	year-1	**year-2**	**year-1**	**year-2**	
WB volume					
unnormalized (ml), mean ± SD	1134 ± 122	1123 ± 129	1071 ± 89	1066 ± 93	*n.a.*
normalized (%ICV), mean ± SD	70.4 ± 2.1	70.8 ± 2.2	70.0 ± 2.0	69.8 ± 2.3	***F*(1,42)=5.473, *p*=0.024 ***
WM volume					
unnormalized (ml), mean ± SD	419 ± 59	415 ± 61	393 ± 39	392 ± 42	*n.a.*
normalized (%ICV), mean ± SD	26.0 ± 1.1	26.1 ± 1.0	25.7 ± 1.2	25.6 ± 1.4	***F*(1,42)=6.093, *p*=0.018 ***
GM volume	679 ± 64	673 ± 68	644 ± 52	641 ± 53	*n.a.*
unnormalized (ml), mean ± SD	42.2 ± 2.0	42.4 ± 2.1	42.1 ± 4.4	42.0 ± 1.6	*F*(1,42)=3.589, *p*=0.065
normalized (%ICV), mean ± SD					
DGM volume[1]					
unnormalized (ml), mean ± SD	44.13 ± 3.08	44.16 ± 3.10	41.44 ± 3.68	41.25 ± 3.70	*n.a.*
normalized (%ICV), mean ± SD	2.75 ± 0.19	2.76 ± 0.19	2.72 ± 0.17	2.70 ± 0.17	***F*(1,42)=5.341, *p*=0.026 ***
z-score, mean ± SD	0.145 ± 1.078	0.158 ± 1.067	-0.040 ± 0.947	-0.087 ± 0.972	*B*=-0.061, *p*=0.571, CI [-0.271, 0.149]
CGM thickness[2]					
mm, mean ± SD	2.58 ± 0.10	2.57 ± 0.11	2.55 ± 0.10	2.54 ± 0.10	*F*(1,86)=0.551, *p*=0.460
z-score, mean ± SD	0.187 ± 1.022	0.099 ± 1.036	-0.038 ± 0.967	-0.083 ± 0.986	*B*=0.043, *p*=0.334, CI [-0.044, 0.129]
WM z-scores in DGM tracts[3]					
WM-Diffusion	0.442 ± 0.829	0.105 ± 1.096	-0.050 ± 0.867	-0.180 ± 0.872	***B*=0.207, *p*=0.023, CI [0.028, 0.387] ***
WM-Extended[4]	0.384 ± 0.688	0.177 ± 0.898	-0.138 ± 0.839	-0.169 ± 0.772	***B*=0.290, *p*=0.001, CI [0.114, 0.467] ****
FA	0.504 ± 1.028	0.123 ± 1.094	-0.013 ± 0.941	-0.250 ± 0.911	*B*=0.144, *p*=0.131, CI [-0.043, 0.332]
MD	-0.362 ± 0.817	-0.014 ± 1.207	0.090 ± 0.952	0.105 ± 0.980	***B*=-0.234, *p*=0.022, CI [-0.452, -0.035] ***
AD	-0.006 ± 0.960	-0.079 ± 1.127	0.132 ± 0.982	-0.096 ± 0.961	*B*=-0.155, *p*=0.183, CI [-0.382, 0.073]
RD	-0.462 ± 0.879	-0.088 ± 1.224	0.046 ± 0.927	0.184 ± 0.946	***B*=-0.235, *p*=0.020, CI [-0.433, -0.037] ***
NDI	0.434 ± 0.82	0.221 ± 0.962	-0.150 ± 0.985	-0.125 ± 1.006	***B*=0.238, *p*=0.013, CI [0.051, 0.425] ***
ODI	-0.336 ± 0.982	0.032 ± 1.043	-0.124 ± 0.965	0.251 ± 0.958	*B*=0.007, *p*=0.947, CI [-0.211, 0.226]
MWF[4]	0.514 ± 0.592	0.347 ± 0.616	-0.250 ± 1.424	-0.139 ± 0.679	***B*=0.342, *p*=0.004, CI [0.108, 0.577] ****
QSM[5]	0.053 ± 1.019	0.041 ± 0.943	-0.052 ± 0.952	0.011 ± 1.057	*B*=0.050, *p*=0.706, CI [-0.209, 0.309]

[continued on next page]

Table 2 – *[continued]*

	HC *n*=13		RRMS *n*=31		Statistics Time * Type interaction
	year-1	year-2	year-1	year-2	
WM z-scores in CGM tracts[3]					
WM-Diffusion	0.320 ± 0.877	0.045 ± 1.002	0.033 ± 0.807	-0.186 ± 0.919	*B*=0.055, *p*=0.136, CI [-0.017, 0.128]
WM-Extended[4]	0.250 ± 0.788	0.123 ± 0.850	-0.079 ± 0.743	-0.189 ± 0.819	***B*=0.122, *p*<0.001, CI [0.050, 0.195] *****
FA	0.408 ± 1.001	-0.068 ± 1.027	0.143 ± 0.940	-0.286 ± 0.952	*B*=0.048, *p*=0.300, CI [-0.043, 0.138]
MD	-0.228 ± 0.971	-0.154 ± 1.131	0.073 ± 0.880	0.087 ± 1.037	*B*=-0.060, *p*=0.132, CI [-0.139, 0.018]
AD	0.022 ± 1.045	-0.264 ± 1.072	0.212 ± 0.928	-0.111 ± 0.967	*B*=-0.037, *p*=0.425, CI [-0.128, 0.054]
RD	-0.323 ± 0.970	-0.049 ± 1.143	-0.030 ± 0.873	0.186 ± 1.020	*B*=-0.058, *p*=0.161, CI [-0.139, 0.023]
NDI	0.293 ± 0.952	0.205 ± 0.970	-0.073 ± 0.977	-0.136 ± 1.009	*B*=0.025, *p*=0.469, CI [-0.042, 0.092]
ODI	-0.261 ± 1.014	0.368 ± 1.101	-0.311 ± 0.879	0.266 ± 0.927	*B*=-0.052, *p*=0.293, CI [-0.149, 0.045]
MWF[4]	0.495 ± 0.810	0.274 ± 0.792	-0.171 ± 1.245	-0.164 ± 0.813	***B*=0.312, *p*<0.001, CI [0.212, 0.412] *****
QSM[5]	0.154 ± 1.024	0.211 ± 0.938	-0.118 ± 0.982	-0.040 ± 1.004	*B*=0.010, *p*=0.852, CI [-0.096, 0.116]

Abbreviations: SD = standard deviation; %ICV = percentage of total intra cranial volume; B = estimate; CI = 95% confidence interval.

Legend: * *p*<0.05; ** *p*<0.01; *** *p*<0.001. Time*Type interaction calculated with HC as reference.

Foot notes: [1] DGM volume is the total of the 10 structures (i.e. excluding left and right accumbens and amygdala); [2] CGM thickness is calculated from the mean left and mean right cortical thickness output from FreeSurfer; [3] WM measures are depicted as a mean z-score over all DGM c.q. CGM tracts, respectively, and statistics calculated with linear mixed models to correct for the amount of tracts; [4] MWF was available for 11/13 HCs and 24/31 RRMS subjects at year-1, and for 13/13 HC and 30/31 RRMS subjects at year-2; [5] QSM was available for 13/13 HC and 30/31 RRMS subjects at year-1, and for 13/13 HC and 30/31 RRMS subjects at year-2.

Table 3 – MRI characteristics of RRMS subjects with non-declining or declining MSFC score over time.

	MSFC non-declining *n*=16		MSFC declining *n*=15		Statistics Time * Type interaction
	year-1	year-2	year-1	year-2	
WB volume					
unnormalized (ml), mean ± SD	1073 ± 93	1075 ± 96	1064 ± 94	1058 ± 92	*n.a.*
normalized (%ICV), mean ± SD	70.3 ± 2.0	70.4 ± 2.0	69.6 ± 2.1	69.2 ± 2.4	***F*(1,29)=7.029 , *p*=0.013 ***
WM volume					
unnormalized (ml), mean ± SD	392 ± 44	393 ± 45	392 ± 40	390 ± 39	*n.a.*
normalized (%ICV), mean ± SD	25.6 ± 1.2	25.7 ± 1.3	25.6 ± 1.4	25.5 ± 1.5	***F*(1,29)=5.266 , *p*=0.029 ***
GM volume					
unnormalized (ml), mean ± SD	648 ± 49	649 ± 51	638 ± 58	634 ± 56	*n.a.*
normalized (%ICV), mean ± SD	42.5 ± 1.4	42.5 ± 1.4	41.7 ± 1.5	41.4 ± 1.6	***F*(1,29)=5.895 , *p*=0.022 ***
DGM volume[1]					
unnormalized (ml), mean ± SD	42.12 ± 3.67	42.05 ± 3.71	40.71 ± 3.68	40.41 ± 3.62	*n.a.*
normalized (%ICV), mean ± SD	2.76 ± 0.17	2.76 ± 0.17	2.67 ± 0.16	2.65 ± 0.16	***F*(1,29)=5.470, *p*=0.026 ***
z-score, mean ± SD	0.164 ± 0.911	0.152 ± 0.918	-0.257 ± 0.940	-0.342 ± 0.967	*B*=0.074, *p*=0.511, CI [-0.147, 0.296]
CGM thickness[2]					
mm, mean ± SD	2.56 ± 0.09	2.56 ± 0.09	2.53 ± 0.10	2.52 ± 0.10	***F*(1,60)=7.165, *p*=0.010 ****
z-score, mean ± SD	0.024 ± 0.949	0.033 ± 0.959	-0.104 ± 0.983	-0.207 ± 0.999	***B*=0.113, *p*=0.019, CI [0.018, 0.207] ***
WM z-scores in DGM tracts[3]					
WM-Diffusion	0.045 ± 0.949	0.030 ± 0.870	-0.151 ± 0.759	-0.404 ± 0.820	***B*=0.238, *p*=0.009, CI [0.061, 0.415] ****
WM-Extended[4]	-0.115 ± 0.915	0.032 ± 0.751	-0.171 ± 0.722	-0.369 ± 0.742	***B*=0.259, *p*=0.006, CI [0.074, 0.444] ****
FA	0.033 ± 0.992	-0.110 ± 0.963	-0.062 ± 0.883	-0.398 ± 0.830	***B*=0.194, *p*=0.043, CI [0.006, 0.382] ***
MD	-0.046 ± 1.037	-0.144 ± 0.920	0.235 ± 0.831	0.371 ± 0.974	***B*=-0.235, *p*=0.026, CI [-0.442, -0.028] ***
AD	0.004 ± 1.009	-0.262 ± 0.892	0.269 ± 0.936	0.080 ± 1.002	*B*=-0.077, *p*=0.517, CI [-0.310, 0.156]
RD	-0.057 ± 1.012	-0.056 ± 0.926	0.156 ± 0.815	0.441 ± 0.903	***B*=-0.285, *p*=0.004, CI [-0.480, -0.089] ****
NDI	-0.020 ± 0.996	0.076 ± 0.951	-0.289 ± 0.956	-0.339 ± 1.022	*B*=0.146, *p*=0.132, CI [-0.044, 0.337]
ODI	-0.129 ± 0.966	0.260 ± 0.915	-0.118 ± 0.967	0.242 ± 1.004	*B*=0.030, *p*=0.800, CI [-0.201, 0.261]
MWF[4]	-0.373 ± 1.711	0.018 ± 0.607	-0.079 ± 0.856	-0.296 ± 0.713	***B*=0.602, *p*<0.001, CI [0.307, 0.898] *****
QSM[5]	0.045 ± 0.990	0.105 ± 1.062	-0.163 ± 0.896	-0.082 ± 1.047	*B*=0.005, *p*=0.973, CI [-0.283, 0.293]

[continued on next page]

Table 3 – *[continued]*

	MSFC non-declining *n*=16		MSFC declining *n*=15		Statistics Time * Type interaction
	year-1	year-2	year-1	year-2	
WM z-scores in CGM tracts[3]					
WM-Diffusion	0.185 ± 0.781	0.055 ± 0.788	-0.128 ± 0.803	-0.444 ± 0.977	***B*=0.186, *p*<0.001, CI [0.111, 0.261] ***
WM-Extended[4]	0.010 ± 0.728	0.027 ± 0.698	-0.204 ± 0.747	-0.404 ± 0.874	***B*=0.176, *p*<0.001, CI [0.098, 0.254] ***
FA	0.248 ± 0.952	-0.128 ± 0.922	0.030 ± 0.913	-0.453 ± 0.955	***B*=0.107, *p*=0.030, CI [0.011, 0.203] ***
MD	-0.112 ± 0.845	-0.206 ± 0.850	0.270 ± 0.874	0.399 ± 1.1123	***B*=-0.223, *p*<0.001, CI [-0.304, -0.143] ***
AD	0.065 ± 0.938	-0.314 ± 0.894	0.370 ± 0.892	0.106 ± 0.995	***B*=-0.116, *p*=0.019, CI [-0.213, -0.019] ***
RD	-0.194 ± 0.842	-0.088 ± 0.849	0.145 ± 0.871	0.478 ± 1.104	***B*=-0.227, *p*<0.001, CI [-0.310, -0.144] ***
NDI	0.084 ± 0.949	0.056 ± 0.958	-0.240 ± 0.979	-0.341 ± 1.022	*B*=0.073, *p*=0.047, CI [0.001, 0.145]
ODI	-0.332 ± 0.906	0.229 ± 0.916	-0.288 ± 0.848	0.306 ± 0.938	*B*=-0.033, *p*=0.541, CI [-0.137, 0.072]
MWF[4]	-0.218 ± 1.433	0.020 ± 0.623	-0.104 ± 0.916	-0.347 ± 0.931	***B*=0.486, *p*<0.001, CI [0.368, 0.604] ***
QSM[5]	-0.114 ± 0.919	-0.071 ± 0.908	-0.123 ± 1.049	-0.010 ± 1.091	*B*=-0.050, *p*=0.418, CI [-0.170, 0.071]

Abbreviations: SD = standard deviation; %ICV = percentage of total intra cranial volume; B = estimate; CI = 95% confidence interval.
Legend: * *p*<0.05; ** *p*<0.01; *** *p*<0.001. Time*Type interaction calculated with non-declining MSFC as reference.
Foot notes: [1] DGM volume is the total of the 10 structures (i.e. excluding left and right accumbens and amygdala); [2] CGM thickness is calculated from the mean left and mean right cortical thickness output from FreeSurfer; [3] WM measures are depicted as a mean z-score over all DGM c.q. CGM tracts, respectively, and statistics calculated with linear mixed models to correct for the amount of tracts; [4] MWF was available for 14/16 MSFC non-declining subjects and 10/15 MSFC declining subjects at year-1, and for 15/16 MSFC non-declining subjects and 15/15 MSFC declining subjects at year-2; [5] QSM was available for 16/16 MSFC non-declining and 14/15 MSFC declining subjects at year-1, and for 15/16 MSFC non-declining and 15/15 MSFC declining subjects at year-2.

Effect of baseline WM integrity on change in connected GM over time

The relation between baseline WM integrity and changes in DGM volume and CGM thickness over time is shown in **Table 4**. **Table 4A** shows significant relations between baseline WM-Diffusion, WM-Extended, FA and RD z-scores and the change in CGM thickness in the entire cohort (e.g. WM-Extended: *B*=-0.055, SE=0.019, *p*=0.005). This relationship between baseline WM and CGM thickness change was stronger in RRMS compared to HC subjects for WM-Diffusion, WM-Extended, MD, RD and NDI (e.g. WM-Diffusion*Type: *B*=0.088, SE=0.025, *p*<0.001), thus including both axonal and myelin markers. When the MSFC non-declining group was compared with the MSFC declining group, similar effects were observed as when comparing HC with RRMS (**Table 4B**): also here the baseline WM z-scores had a WM*Type relationship in CGM (e.g. WM-Extended*MSFC type: *B*=0.120, SE=0.031, *p*=0.001), where the relationship between WM damage and CGM thickness change was stronger in MSFC decliners than in non-decliners, and involved both markers of myelin damage in the WM (FA, RD) and the marker of axonal damage (NDI). No significant relations between WM z-scores and change in DGM volumes over time were seen, neither over the entire cohort, nor for the MSFC groups.

Post-hoc analysis in HC and RRMS groups separately (**Table 4C**) showed that lower WM integrity related to a decrease in cortical thickness over time in RRMS subjects (e.g. WM-Extended: *B*=0.033, SE=0.015, *p*=0.030), whereas HC subjects showed an opposite relationship (e.g. WM-Extended: *B*=-0.055, SE=0.019, *p*=0.004). Post-hoc analysis splitting in the two MSFC groups showed that the WM-GM relationship was only significant in subjects whose MSFC score declined over time (e.g. WM-Extended: *B*=0.098, SE=0.025, *p*<0.001), and not in subjects whose MSFC did not (e.g. WM-Extended: *B*=-0.021, SE=0.019, *p*=0.272).

Effect of baseline GM on change in connected WM integrity over time

The relation between baseline DGM volume or CGM thickness and changes in WM integrity over time is shown in **Table 5**. **Table 5A** shows there were no significant relations between baseline GM z-scores and WM z-scores over time, except for a different association in HC and RRMS subjects of baseline DGM volume on AD change over time, which was also seen between the two MSFC groups (**Table 5B**). CGM thickness z-scores did not relate significantly to WM change over time.

Post-hoc analysis in HC and RRMS groups separately (**Table 5C**) showed that in RRMS subjects, lower DGM volumes related to a decrease in AD in the connected tracts over time (*B*=0.109, SE=0.042, *p*=0.009), an effect that was not found in HC subjects (*B*=-0.130, SE=0.075, *p*=0.086). Furthermore, this was only seen in subjects whose MSFC would decline over time (*B*=0.203, SE=0.050, *p*<0.001), and not in subjects whose MSFC would not decline over time (*B*=-0.002, SE=0.066, *p*=0.977).

Table 4 – Effect of baseline WM integrity z-score on change in DGM volume z-score and CGM thickness z-score over time; compared in HC/RRMS (A), MSFC non-declining/declining (B), and split per group (C; CGM only).

A. HC vs MS		ΔDGM			ΔCGM		
		WM	**Type**	**WM*Type**	**WM**	**Type**	**WM*Type**
WM-Diffusion	*B*, *p* 95% CI	0.026, 0.201 [-0.014, 0.066]	-0.049, 0.076 [-0.102, 0.005]	-0.014, 0.551 [-0.061, 0.033]	**-0.051, 0.002** ** **[-0.083, -0.018]**	0.026, 0.622 [-0.079, 0.131]	**0.073, <0.001** *** **[0.033, 0.113]**
WM-Extended[1]	*B*, *p* 95% CI	0.006, 0.789 [-0.041, 0.054]	-0.042, 0.099 [-0.093, 0.008]	-0.004, 0.897 [-0.058, 0.051]	**-0.055, 0.005** ** **[-0.093, -0.017]**	0.024, 0.697 [-0.101, 0.149]	**0.088, <0.001** *** **[0.040, 0.136]**
FA	*B*, *p* 95% CI	0.031, 0.059 [-0.001, 0.063]	-0.045, 0.097 [-0.099, 0.008]	-0.023, 0.254 [-0.063, 0.017]	**-0.038, 0.003** ** **[-0.064, -0.013]**	0.027, 0.608 [-0.079, 0.133]	0.038, 0.015 [0.008, 0.069]
MD	*B*, *p* 95% CI	-0.003, 0.885 [-0.041, 0.035]	-0.059, 0.028 [-0.111, -0.006]	-0.007, 0.766 [-0.051, 0.037]	0.036, 0.024 [0.005, 0.068]	0.037, 0.479 [-0.068, 0.142	**-0.074, <0.001** *** **[-0.112, -0.035]**
AD	*B*, *p* 95% CI	0.035, 0.028 [0.004, 0.065]	-0.060, 0.017 [-0.109, -0.011]	-0.040, 0.035 [-0.076, -0.003]	-0.004, 0.769 [-0.030, 0.022]	0.048, 0.363 [-0.058, 0.154]	-0.022, 0.155 [-0.053, 0.009]
RD	*B*, *p* 95% CI	-0.024, 0.209 [-0.062, 0.014]	-0.049, 0.073 [-0.103, 0.005]	0.012, 0.598 [-0.032, 0.056]	**0.041, 0.007** ** **[0.011, 0.071]**	0.029, 0.583 [-0.076, 0.134]	**-0.062, <0.001** *** **[-0.098, -0.025]**
NDI	*B*, *p* 95% CI	-0.008, 0.659 [-0.042, 0.027]	-0.064, 0.021 [-0.118, -0.010]	0.007, 0.720 [-0.034, 0.048]	-0.039, 0.024 [-0.073, -0.005]	0.034, 0.516 [-0.070, 0.138]	**0.073, <0.001** *** **[0.033, 0.114]**
ODI	*B*, *p* 95% CI	-0.040, 0.012 [-0.071, -0.009]	-0.048, 0.066 [-0.100, 0.003]	0.033, 0.080 [-0.004, 0.070]	0.021, 0.088 [-0.003, 0.046]	0.042, 0.434 [-0.065, 0.149]	-0.006, 0.672 [-0.036, 0.023]
MWF[1]	*B*, *p* 95% CI	-0.031, 0.254 [-0.086, 0.023]	-0.061, 0.026 [-0.115, -0.008]	0.032, 0.268 [-0.025, 0.089]	-0.015, 0.393 [-0.049, 0.019]	0.032, 0.610 [-0.095, 0.160]	0.041, 0.045 [0.001, 0.081]
QSM[2]	*B*, *p* 95% CI	0.009, 0.536 [-0.019, 0.037]	-0.059, 0.028 [-0.111, -0.007]	0.008, 0.625 [-0.026, 0.043]	-0.005, 0.659 [-0.029, 0.018]	0.039, 0.470 [-0.069, 0.147]	0.010, 0.461 [-0.017, 0.038]

[continued on next page]

Table 4 – *[continued]*

B. MSFC non-declining vs MSFC declining		ΔDGM			ΔCGM		
		WM	**Type**	**WM*Type**	**WM**	**Type**	**WM*Type**
WM-Diffusion	*B*, *p* 95% CI	-0.003, 0.850 [-0.036, 0.029]	-0.070, 0.018 [-0.127, -0.013]	0.029, 0.266 [-0.023, 0.081]	-0.022, 0.188 [-0.056, 0.011]	-0.109, 0.0510 [-0.219, 0.000]	**0.083, <0.001** *** **[0.037, 0.129]**
WM-Extended[1]	*B*, *p* 95% CI	-0.004, 0.800 [-0.038, 0.029]	-0.049, 0.079 [-0.105, 0.006]	0.019, 0.521 [-0.041, 0.079]	-0.021, 0.306 [-0.061, 0.019]	-0.090, 0.200 [-0.232, 0.051]	**0.120, <0.001** *** **[0.059, 0.180]**
FA	*B*, *p* 95% CI	-0.003, 0.874 [-0.034, 0.029]	-0.073, 0.017 [-0.132, -0.014]	0.022, 0.369 [-0.026, 0.069]	-0.026, 0.029 [-0.049, 0.003]	-0.120, 0.034 [-0.230, -0.010]	**0.056, 0.001** ** **[0.021, 0.090]**
MD	*B*, *p* 95% CI	0.003, 0.834 [-0.026, 0.032]	-0.069, 0.016 [-0.125, -0.014]	-0.023, 0.316 [-0.069, 0.023]	-0.009, 0.587 [-0.041, 0.023]	-0.096, 0.084 [-0.206, 0.014]	-0.049, 0.025 [-0.093, -0.006]
AD	*B*, *p* 95% CI	0.002, 0.912 [-0.026, 0.029]	-0.073, 0.013 [-0.129, -0.016]	-0.006, 0.777 [-0.047, 0.035]	-0.026, 0.025 [-0.049, -0.003]	-0.105, 0.060 [-0.216, 0.005]	0.002, 0.930 [-0.032, 0.035]
RD	*B*, *p* 95% CI	0.002, 0.894 [-0.028, 0.032]	-0.070, 0.017 [-0.127, -0.013]]	-0.028, 0.244 [-0.077, 0.020]	0.020, 0.205 [-0.011, 0.051]	-0.109, 0.051 [-0.219, 0.000]	**-0.074, <0.001** *** **[-0.116, -0.031]**
NDI	*B*, *p* 95% CI	-0.011, 0.453 [-0.041, 0.018]	-0.072, 0.017 [-0.131, -0.014]	0.017, 0.460 [-0.029, 0.63]	0.003, 0.855 [-0.029, 0.035]	-0.098, 0.085 [-0.210, 0.014]	**0.059, 0.010** ** **[0.014, 0.104]**
ODI	*B*, *p* 95% CI	-0.008, 0.578 [-0.036, 0.020]	-0.074, 0.014 [-0.131, -0.016]	0.002, 0.938 [-0.040, 0.043]	0.025, 0.029 [0.003, 0.047]	-0.120, 0.034 [-0.231, -0.009]	-0.021, 0.212 [-0.053, 0.012]
MWF[1]	*B*, *p* 95% CI	0.002, 0.837 [-0.017, 0.021]	-0.053, 0.057 [-0.108, 0.002]	-0.001, 0.980 [-0.047, 0.046]	0.008, 0.547 [-0.019, 0.035]	-0.106, 0.33 [-0.246, 0.035]	0.052, 0.025 [0.007, 0.097]
QSM[2]	*B*, *p* 95% CI	0.014, 0.261 [-0.011, 0.040]	-0.072, 0.018 [-0.131, -0.014]	0.002, 0.907 [-0.038, 0.043]	-0.014, 0.189 [-0.035, 0.007]	-0.120, 0.037 [-0.233,-0.008]	0.037, 0.013 [0.008, 0.066]

[continued on next page]

Table 4 – *[continued]*

C. Post hoc group split		ΔCGM			
		HC *n*=13	**RRMS** *n*=31	**MSFC non-declining** *n*=16	**MSFC declining** *n*=15
WM-Diffusion	*B*, *p* 95% CI	-0.051, 0.003 [-0.084, -0.018]	0.022, 0.057 [-0.001, 0.046]	-0.022, 0.150 [-0.053, 0.008]	**0.030, <0.001 ***** **[0.026. 0.095]**
WM-Extended [1]	*B*, *p* 95% CI	**-0.055, 0.005 **** **[-0.093, -0.017]**	0.033, 0.030 [0.003, 0.063]	-0.021. 0.272 [-0.059. 0.017]	**0.098. <0.001 ***** **[0.050, 0.147]**
FA	*B*, *p* 95% CI	**-0.038, 0.004 **** **[-0.064, -0.012]**	0.000, 0.990 [-0.017, 0.017]	-0.026, 0.016 [-0.047, -0.005]	0.030, 0.032 [0.003, 0.058]
MD	*B*, *p* 95% CI	0.036, 0.027 [0.004, 0.069]	**-0.037, <0.001 ***** **[-0.059, -0.016]**	-0.009, 0.539 [-0.039, 0.020]	**-0.058, <0.001 ***** **[-0.090, -0.027]**
AD	*B*, *p* 95% CI	-0.004, 0.770 [-0.030, 0.023]	**-0.026, 0.002 **** **[-0.043, -0.010]**	-0.027, 0.014 [-0.048, -0.005]	-0.025, 0.065 [-0.052, 0.002]
RD	*B*, *p* 95% CI	**0.041, 0.008 **** **[0.011, 0.071]**	-0.021, 0.057 [-0.042, 0.001]	0.020, 0.168 [-0.008, 0.048]	**-0.053, <0.001 ***** **[-0.085, -0.022]**
NDI	*B*, *p* 95% CI	-0.039, 0.027 [-0.073, -0.004]	**0.034, 0.003 **** **[0.012, 0.057]**	0.002, 0.883 [-0.027, 0.032]	**0.062, <0.001 ******* **[0.028, 0.096]**
ODI	*B*, *p* 95% CI	0.021, 0.094 [-0.004, 0.046]	0.015, 0.070 [-0.001, 0.031]	0.025, 0.016 [0.005, 0.045]	0.004, 0.771 [-0.022, 0.030]
MWF[1]	*B*, *p* 95% CI	-0.015, 0.393 [-0.049, 0.019]	0.026, 0.017 [0.005, 0.048]	0.008, 0.530 [-0.017, 0.033]	**0.060, 0.003 **** **[0.021, 0.099]**
QSM[2]	*B*, *p* 95% CI	-0.005, 0.665 [-0.029, 0.019]	0.005, 0.492 [-0.009, 0.020]	-0.014, 0.151 [-0.033, 0.005]	0.023, 0.042 [0.001, 0.045]

Legend: B = estimate; CI = confidence interval; ** $p<0.01$; *** $p<0.001$.
Foot notes: [1] MWF was available for 11/13 HCs and 24/31 RRMS subjects (MSFC non-declining *n*=15, MSFC declining *n*=10) at year-1, and for 13/13 HC and 30/31 RRMS subjects (MSFC non-declining *n*=15, MSFC declining *n*=15) at year-2; [2] QSM was available for 13/13 HC and 30/31 RRMS subjects (MSFC non-declining *n*=16, MSFC declining *n*=14) at year-1, and for 13/13 HC and 30/31 RRMS subjects (MSFC non-declining *n*=15, MSFC declining *n*=15) at year-2.

Table 5 – Effect of baseline DGM volume z-score or CGM thickness z-score on change in WM integrity z-score over time; compared in HC/RRMS (A), MSFC non-declining/declining (B), and split per group (C; DGM only).

A. HC vs MS		**DGM**			**CGM**		
		WM	**Type**	**WM*Type**	**WM**	**Type**	**WM*Type**
ΔWM-Diffusion	*B*, *p* 95% CI	0.009, 0.867 [-0.095, 0.113]	0.209, 0.201 [-0.115, 0.533]	-0.007, 0.911 [-0.136, 0.121]	-0.001, 0.965 [-0.042, 0.040]	0.055, 0.720 [-0.253, 0.364]	0.008, 0.760 [-0.042, 0.057]
ΔWM-Extended [1]	*B*, *p* 95% CI	0.015, 0.791 [-0.094, 0.123]	0.306, 0.089 [-0.049, 0.660]	-0.060, 0.389 [-0.195, 0.076]	-0.010, 0.642 [-0.051, 0.032]	0.121, 0.465 [-0.213, 0.456]	0.025, 0.341 [-0.026, 0.075]
ΔFA	*B*, *p* 95% CI	-0.046, 0.330 [-0.140, 0.047]	0.139, 0.252 [-0.103, 0.381]	0.083, 0.160 [-0.033, 0.198]	0.011, 0.651 [-0.036, 0.058]	0.049, 0.663 [-0.178, 0.277]	-0.020, 0.478 [-0.076, 0.036]
ΔMD	*B*, *p* 95% CI	-0.068, 0.286 [-0.193, 0.057]	-0.251, 0.189 [-0.631, 0.128]	0.107, 0.174 [-0.047, 0.261]	0.010, 0.684 [-0.038, 0.057]	-0.059, 0.745 [-0.424, 0.306]	-0.031, 0.281 [-0.088, 0.025]
ΔAD	*B*, *p* 95% CI	-0.129, 0.041 [-0.254, -0.005]	-0.169, 0.240 [-0.455, 0.117]	**0.235, 0.003 **** **[0.081, 0.389]**	0.005, 0.841 [-0.046, 0.056]	-0.037, 0.754 [-0.273, 0.200]	-0.033, 0.287 [-0.094, 0.028]
ΔRD	*B*, *p* 95% CI	-0.008, 0.888 [-0.125, 0.108]	-0.237, 0.210 [-0.611, 0.138]	-0.001, 0.987 [-0.145, 0.143]	0.003, 0.903 [-0.044, 0.049]	-0.058, 0.753 [-0.424, 0.309]	-0.013, 0.656 [-0.068, 0.043]
ΔNDI	*B*, *p* 95% CI	0.018, 0.774 [-0.103, 0.138]	0.239, 0.068 [-0.018, 0.497]	-0.044, 0.562 [-0.192, 0.105]	-0.009, 0.670 [-0.049, 0.031]	0.024, 0.815 [-0.180, 0.227]	0.022, 0.356 [-0.025, 0.070]
ΔODI	*B*, *p* 95% CI	0.087, 0.099 [-0.016, 0.190]	0.017, 0.850 [-0.167, 0.202]	-0.150, 0.022 [-0.277, -0.022]	-0.036, 0.1169 [-0.086, 0.015]	-0.058, 0.474 [-0.221, 0.104]	0.053, 0.085 [-0.007, 0.114]
ΔMWF[1]	*B*, *p* 95% CI	0.028, 0.645 [-0.093, 0.150]	0.363, 0.407 [-0.517, 1.243]	-0.075, 0.333 [-0.228, 0.078]	-0.025, 0.465 [-0.093, 0.042]	0.308, 0.385 [-0.405, 1.021]	0.093, 0.026 [0.011, 0.176]
ΔQSM[2]	*B*, *p* 95% CI	0.068, 0.320 [-0.067, 0.203]	0.051, 0.593 [-0.139, 0.241]	-0.066, 0.440 [-0.235, 0.103]	-0.004, 0.914 [-0.069, 0.062]	0.009, 0.906 [-0.148, 0.167]	0.019, 0.643 [-0.060, 0.097]

[continued on next page]

Table 5 – *[continued]*

B. MSFC non-declining vs MSFC declining		**DGM**			**CGM**		
		WM	**Type**	**WM*Type**	**WM**	**Type**	**WM*Type**
ΔWM-Diffusion	*B*, *p* 95% CI	0.035, 0.459 [-0.059, 0.129]	-0.241, 0.100 [-531, 0.049]	-0.070, 0.277 [-0.196, 0.056]	0.028, 0.110 [-0.006, 0.062]	-0.0187, 0.148 [-0.444, 0.070]	-0.043, 0.086 [-0.092, 0.006]
ΔWM-Extended[1]	*B*, *p* 95% CI	-0.017, 0.719 [-0.111, 0.077]	-0.285, 0.114 [-0.646, 0.075]	-0.074, 0.274 [-0.207, 0.059]	0.032, 0.071 [-0.003, 0.066]	-0.172, 0.259 [-0.479, 0.136]	-0.037, 0.156 [-0.088, 0.014]
ΔFA	*B*, *p* 95% CI	0.040, 0.334 [-0.041, 0.120]	-0.181, 0.104 [-0.402, 0.040]	-0.016, 0.772 [-0.125, 0.093]	-0.006, 0.782 [-0.046, 0.035]	-0.108, 0.312 [-0.324, 0.107]	-0.007, 0.812 [-0.065, 0.051]
ΔMD	*B*, *p* 95% CI	-0.025, 0.664 [-0.139, 0.089]	0.257, 0.150 [-0.097, 0.610]	0.127, 0.107 [-0.027, 0.281]	-0.051, 0.011 [-0.090, -0.012]	0.223, 0.136 [-0.074, 0.521]	0.060, 0.035 [0.004, 0.116]
ΔAD	*B*, *p* 95% CI	-0.005, 0.940 [-0.125, 0.116]	0.131, 0.390 [-0.175, 0.437]	**0.216, 0.010 **** **[0.052, 0.380]**	-0.042, 0.057 [-0.086, 0.001]	0.114, 0.276 [-0.096, 0.323]	0.031, 0.334 [-0.032, 0.093]
ΔRD	*B*, *p* 95% CI	-0.042, 0.438 [-0.147, 0.064]	0.285, 0.087 [-0.044, 0.613]	0.068, 0.350 [-0.075, 0.210]	-0.040, 0.042 [-0.078, -0.002]	0.228, 0.137 [-0.077, 0.534]	0.061, 0.028 [0.007, 0.116]
ΔNDI	*B*, *p* 95% CI	-0.012, 0.831 [-0.125, 0.100]	-0.162, 0.191 [-0.408, 0.085]	-0.039, 0.613 [-0.193, 0.114]	0.037, 0.034 [0.003, 0.071]	-0.073, 0.470 [-0.278, 0.132]	-0.047, 0.056 [-0.096, 0.001]
ΔODI	*B*, *p* 95% CI	-0.023, 0.669 [-0.127, 0.081]	-0.062, 0.548 [-0.271, 0.147]	-0.089, 0.222 [-0.231, 0.054]	0.022, 0.343 [-0.023, 0.066]	0.034, 0.686 [-0.138, 0.207]	-0.008, 0.812 [-0.071, 0.056]
ΔMWF[1]	*B*, *p* 95% CI	-0.004, 0.957 [-0.141, 0.134]	-0.634, 0.284 [-1.835, 0.566]	-0.098,0.323 [-0.292, 0.097]	0.082, 0.014 [0.016, 0.147]	-0.468, 0.314 [-1.410, 0.475]	-0.031, 0.525 [-0.129, 0.066]
ΔQSM[2]	*B*, *p* 95% CI	-0.075, 0.296 [-0.218, 0.067]	-0.006, 0.958 [-0.219, 0.208]	0.156, 0.130 [-0.046, 0.358]	0.030, 0.340 [-0.032, 0.093]	0.049, 0.549 [-0.117, 0.215]	-0.033, 0.468 [-0.121, 0.056]

[continued on next page]

Table 5 – *[continued]*

C. Post hoc group split		**DGM**			
		HC *n*=13	**RRMS** *n*=31	**MSFC non-declining** *n*=16	**MSFC declining** *n*=15
ΔWM-Diffusion	*B*, *p* 95% CI	0.009, 0.900 [-0.129, 0.147]	0.001, 0.963 [-0.061, 0.064]	0.035, 0.493 [-0.065, 0.134]	-0.034, 0.401 [-0.113, 0.046]
ΔWM-Extended[1]	*B*, *p* 95% CI	0.017, 0.814 [-0.127, 0.161]	-0.049, 0.143 [-0.116, 0.017]	-0.022, 0.665 [-0.121, 0.077]	-0.083, 0.059 [-0.169, 0.003]
ΔFA	*B*, *p* 95% CI	-0.045, 0.486 [-0.174, 0.083]	0.036, 0.198 [-0.019, 0.090]	0.040, 0.355 [-0.045, 0.124]	0.024, 0.511 [-0.047, 0.094]
ΔMD	*B*, *p* 95% CI	-0.068, 0.410 [-0.230, 0.095]	0.039, 0.314 [-0.037, 0.116]	-0.024, 0.696 [-0.142, 0.095]	0.099, 0.050 [0.000, 0.198]
ΔAD	*B*, *p* 95% CI	-0.130, 0.086 [-0.279, 0.019]	**0.109, 0.009 **** **[0.027, 0.190]**	-0.002, 0.977 [-0.132, 0.128]	**0.203, <0.001 ***** **[0.103, 0.302]**
ΔRD	*B*, *p* 95% CI	-0.008, 0.919 [-0.162, 0.146]	-0.010, 0.783 [-0.081, 0.061]	-0.041, 0.468 [-0.153, 0.071]	0.025, 0.582 [-0.065, 0.115]
ΔNDI	*B*, *p* 95% CI	0.018, 0.820 [-0.135, 0.170]	-0.026, 0.497 [-0.102, 0.050]	-0.011, 0.856 [-0.133, 0.110]	-0.046, 0.338 [-0.139, 0.048]
ΔODI	*B*, *p* 95% CI	0.087, 0.150 [-0.032, 0.206]	-0.066, 0.066 [-0.136, 0.004]	-0.023, 0.679 [-0.133, 0.087]	-0.113, 0.017 [-0.205, -0.021]
ΔMWF[1]	*B*, *p* 95% CI	0.038, 0.493 [-0.071, 0.147]	-0.050, 0.307 [-0.148, 0.047]	-0.010, 0.898 [-0.162, 0.142]	-0.104, 0.079 [-0.220, 0.012]
ΔQSM[2]	*B*, *p* 95% CI	0.073, 0.28 [-0.075, 0.221]	0.001, 0.978 [-0.097, 0.100]	-0.078, 0.216 [-0.201, 0.046]	0.091, 0.263 [-0.070, 0.253]

Legend: B = estimate; CI = confidence interval; ** $p<0.01$; *** $p<0.001$.
Foot notes: [1] MWF was available for 11/13 HCs and 24/31 RRMS subjects (MSFC non-declining *n*=15, MSFC declining *n*=10) at year-1, and for 13/13 HC and 30/31, RRMS subjects (MSFC non-declining *n*=15, MSFC declining *n*=15) at year-2; [2] QSM was available for 13/13 HC and 30/31 RRMS subjects (MSFC non-declining *n*=16, MSFC declining *n*=14) at year-1, and for 13/13 HC and 30/31 RRMS subjects (MSFC non-declining *n*=15, MSFC declining *n*=15) at year-2.

Predicting MSFC decline based on baseline WM and GM parameters

Exploratory binary logistic regression to study the relation between baseline imaging parameters and the MSFC groups showed that WM-Extended z-scores were better predictors than WM-Diffusion z-scores (Nagelkerke R^2=0.594 versus 0.362, respectively), and that either of the WM z-scores were better predictors than GM z-scores only (Nagelkerke R^2=0.287) in a model including sex, age, education, treatment type and baseline EDSS score. A total of 78% of subjects could be classified correctly (MSFC-non-declining: 84%; MSFC-declining: 69%) in a model based on sex, age, education, treatment type, EDSS, GM z-score, WM-Diffusion z-score and WM-Extended z-score (Nagelkerke R^2=0.611).

Discussion

This longitudinal study in early RRMS explored the spatial and temporal relations between GM damage and WM integrity in the connected tracts using multimodal MRI. Our most important finding was that lower baseline WM integrity, as determined with DTI, NODDI and MWF, related to increasing atrophy of the connected cortical GM over time in subjects with RRMS, and especially for those who experienced increasing disability over the study period. In contrast, lower baseline cortical GM thickness or deep GM volume did not relate to WM integrity changes in the connected tracts over time.

These results suggest that in early RRMS, damage of the WM precedes atrophy of the cortical GM and that this relationship is clinically relevant since it was only found in subjects with worsening of their overall MSFC score over time. Cortical thinning as a result of preceding lower WM integrity in connected tracts has been suggested previously[56,57], and we extend those findings by studying this longitudinally in early RRMS. Accelerated CGM atrophy has previously also been found upon increasing clinical progression independent of relapse activity[58], and our results confirm this already in the very early stages of the disease. For progressive MS, it was earlier found that WM damage preceded cortical GM damage[6], and a recent combined MRI-histopathological study showed that this WM-integrity-related cortical thinning could be attributed to GM axonal density loss, rather than myelin or microglia density loss within GM[59]. Our study adds important new insights, by demonstrating that in WM tracts, both axonal damage and myelin damage are related to subsequent cortical thinning. That myelin damage, as assessed here through MWF and RD, would be implicated, may not come as a surprise given the relations between demyelinating focal WM lesions and GM atrophy in MS described previously and reviewed systematically in [9]. Nonetheless, the fact that quantitatively assessed myelin damage in whole WM tracts was related to subsequent thinning of the connected cortex in RRMS, is novel. While (higher) RD was highly significantly more strongly related to subsequent cortical thinning in RRMS than HC, (lower) MWF failed to reach significance. In the post-hoc analyses in RRMS alone, both RD and MWF just failed to meet the significance threshold. However, in the most affected patient group, those with RRMS who exhibited MSFC decline during the study, both RD and MWF at baseline were highly significantly related to subsequent thinning of the connected cortex. Future studies, ideally in larger groups but

maintaining the homogeneity of image acquisition adhered to here, should investigate the role of myelin damage in WM in the process of cortical thinning in RRMS in more detail.

Axonal damage is another important component of WM damage in MS. With NDI, this study included a quantitative marker of this damage. Lower NDI in WM tracts was significantly related to more subsequent thinning of the connected cortex in; whether assessing its effect in RRMS versus HC; in MSFC-declining versus non-declining RRMS; or in the post-hoc analyses on RRMS separately and in the MSFC-declining subgroup. The recurring strength of this marker suggests an important role for WM axonal damage in the development of subsequent cortical thinning.

The finding that this process may already be ongoing in early RRMS may also inform therapeutic choices: treatments that would be developed to target primarily neurodegeneration may have limited effect, since atrophy may be secondary to WM damage. In the current study we did not find any significant influences of WM integrity on atrophy of DGM in either HC or RRMS subjects; and also not in either of the two MSFC groups. Correlations between WM damage and connected DGM volumes have been shown in earlier studies, although these studies focused only on lesion volume and not on overall integrity of the entire tract[9,60]. Studies that did look at overall integrity of the tract at baseline and its effect on atrophy focused mainly on the thalamus[7,61,62]. Interestingly, a cross-sectional study showed that connectivity of thalamocortical tracts related to cortical, but not thalamic, atrophy[63], which may be in line with our results where we did not find any relations between WM and DGM atrophy, but only with CGM atrophy. It is known that MS DGM segmentation has some methodological issues[64], and therefore we took some precautions: subjects' brain volumes were measured at similar times of the day to limit diurnal volume fluctuations; specific longitudinal segmentation software optimized in subjects with MS was used[34]; amygdala and nucleus accumbens were excluded from analysis due to known difficulties in segmentation of these small structures[36]; and DGM volumes were normalized and converted into z-scores per structure to enable grouping of large and small DGM structures within one LMM. However, where our linear mixed model method was well suited for analysis of the CGM tracts (i.e. damage may occur in different tracts and different cortical areas between MS subjects) it may be less suitable for the analysis of DGM tracts, and analyses per DGM structure may be more informative. Due to the small sample size, we did not further investigate the separate DGM structures in the current study, but this remains of interest for future, larger, studies.

The fact that anterograde neurodegeneration from WM integrity loss was mainly found in subjects who showed clinical progression (i.e. MSFC decliners), further indicates the clinical relevance of these disease mechanisms. It should be noted that the distinction between decliner and non-decliner is only subtle, in the one-year follow-up of this cohort of 31 RRMS subjects. However, although the groups only consisted of 15 declining and 16 non-declining RRMS subjects, exploratory regression analysis suggested importance of baseline WM integrity (including NDI and MWF) over baseline GM scores, on MSFC progression over time. Although implications for clinical progression are subtle and preliminary, future studies may benefit from including measures of the integrity of the tract over lesion volume only, and especially from the

addition of NDI and MWF as important axonal and myelin markers, respectively, to standard DTI analyses.

This study has some limitations. First, our sample size was relatively small and for a few subjects MWF and QSM values were not available; this was mitigated by using mixed models analyses that take all data-points into account. Although our study follow-up time was short, significant atrophy patterns as well as WM-GM relations could be recognized already early in the disease. Larger sample sizes and longer follow-up times might increase the power of these findings.

Furthermore, while it would be interesting to study the relation with GM atrophy separately for damage in WM lesions in connected tracts and for damage in NAWM in connected tracts, we have chosen not to pursue such analyses in the present work. Because of the small numbers of lesions per person in our study group of people with early MS, and because the anatomical locations of those lesions vary from patient to patient, a large number of WM tracts will not contain any lesion. The small amount of data will lead to a low statistical power regarding the relation between that lesion-bound damage on the one hand, and atrophy of the connected GM regions on the other. By contrast, by quantifying the tract-specific damage through the quantitative MR measures for the whole tract, regardless of whether lesions were present, we were able to include all tracts in all patients in our analyses. This allowed for a comprehensive analysis, while also limiting the number of statistical tests. Nonetheless, the extraction of weighted mean values from the whole WM tract may be less optimal for QSM, due to opposite effects on susceptibility of demyelination and iron deposition[15], and so for QSM separate analysis of lesions, perilesional WM and NAWM may be informative. For future studies, it may be interesting to also look at the microstructural measures in the DGM and CGM itself. In this way, we can investigate not only the effect of WM integrity on atrophy of the connected GM, but also on the neurobiological processes underlying this neurodegeneration, such as demyelination and axonal loss.

Conclusion

Lower baseline WM integrity related to increasing cortical atrophy in RRMS subjects that show clinical progression over a 1-year follow-up. Baseline GM did not influence WM integrity changes over time. Collectively, this suggests that white matter tract damage drives cortical neurodegeneration in early RRMS.

Funding

This work was funded by the Dutch MS Research Foundation, grant number 14-876. FB is supported by the NIHR biomedical research centre at UCLH.

None of the funding sources had involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Ethics declarations

The authors declare that there is no conflict of interest.
The institutional ethical standards committee on human experimentation (METc VU University Medical Center) approved the study protocol (approval number 2016.314 NL57713.029.16). The study was registered in the Netherlands Trial Register as observational non-invasive non-randomized controlled trial (ID: NTR6103). All participants gave written informed consent prior to participation, according to the Declaration of Helsinki.

Acknowledgments

This research has been executed within the MS Center Amsterdam, Amsterdam UMC. We thank all who participated in the study.

Data availability

Anonymized data not published within this article will be made available by reasonable request from any qualified investigator.

References

1 Geurts, J. J., Calabrese, M., Fisher, E. & Rudick, R. A. Measurement and clinical effect of grey matter pathology in multiple sclerosis. Lancet Neurol 11, 1082-1092, doi:S1474-4422(12)70230-2 [pii];10.1016/S1474-4422(12)70230-2 [doi] (2012).

2 Benedict, R. H., Ramasamy, D., Munschauer, F., Weinstock-Guttman, B. & Zivadinov, R. Memory impairment in multiple sclerosis: correlation with deep grey matter and mesial temporal atrophy. J. Neurol. Neurosurg. Psychiatry 80, 201-206, doi:jnnp.2008.148403 [pii];10.1136/jnnp.2008.148403 [doi] (2009).

3 Zivadinov, R. et al. Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years. AJNR Am J Neuroradiol 34, 1931-1939, doi:10.3174/ajnr.A3503 (2013).

4 Uher, T. et al. Longitudinal MRI and neuropsychological assessment of patients with clinically isolated syndrome. J. Neurol 261, 1735-1744, doi:10.1007/s00415-014-7413-9 [doi] (2014).

5 Steenwijk, M. D. et al. Unraveling the relationship between regional gray matter atrophy and pathology in connected white matter tracts in long-standing multiple sclerosis. Hum. Brain Mapp 36, 1796-1807, doi:10.1002/hbm.22738 [doi] (2015).

6 Bodini, B. et al. White and gray matter damage in primary progressive MS: The chicken or the egg? Neurology 86, 170-176, doi:10.1212/WNL.0000000000002237 (2016).

7 Kolasinski, J. et al. A combined post-mortem magnetic resonance imaging and quantitative histological study of multiple sclerosis pathology. Brain 135, 2938-2951, doi:10.1093/brain/aws242 (2012).

8 Jehna, M. et al. An exploratory study on the spatial relationship between regional cortical volume changes and white matter integrity in multiple sclerosis. Brain Connect 3, 255-264, doi:10.1089/brain.2012.0108 (2013).

9 Lie, I. A. et al. Relationship Between White Matter Lesions and Gray Matter Atrophy in Multiple Sclerosis: A Systematic Review. Neurology 98, e1562-e1573, doi:10.1212/WNL.0000000000200006 (2022).

10 Jones, D. K., Knosche, T. R. & Turner, R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. Neuroimage 73, 239-254, doi:10.1016/j.neuroimage.2012.06.081 (2013).

11 Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61, 1000-1016, doi:10.1016/j.neuroimage.2012.03.072 (2012).

12 Margoni, M. et al. Quantification of normal-appearing white matter damage in early relapse-onset multiple sclerosis through neurite orientation dispersion and density imaging. Mult Scler Relat Disord 58, 103396, doi:10.1016/j.msard.2021.103396 (2022).

13 Alotaibi, A. et al. Investigating Microstructural Changes in White Matter in Multiple Sclerosis: A Systematic Review and Meta-Analysis of Neurite Orientation Dispersion and Density Imaging. Brain Sci 11, doi:10.3390/brainsci11091151 (2021).

14 Qian, W., Khattar, N., Cortina, L. E., Spencer, R. G. & Bouhrara, M. Nonlinear associations of neurite density and myelin content with age revealed using multicomponent diffusion and relaxometry magnetic resonance imaging. Neuroimage 223, 117369, doi:10.1016/j.neuroimage.2020.117369 (2020).

15 Zhang, Y. et al. Longitudinal change in magnetic susceptibility of new enhanced multiple sclerosis (MS) lesions measured on serial quantitative susceptibility mapping (QSM). J Magn Reson Imaging 44, 426-432, doi:10.1002/jmri.25144 (2016).

16 MacKay, A. L. & Laule, C. Magnetic Resonance of Myelin Water: An in vivo Marker for Myelin. Brain Plast 2, 71-91, doi:10.3233/BPL-160033 (2016).

17 Polman, C. H. et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 69, 292-302, doi:10.1002/ana.22366 (2011).

18 De Stefano, N. et al. Evidence of early cortical atrophy in MS: relevance to white matter changes and disability. Neurology 60, 1157-1162 (2003).

19 Rao, A. B. et al. Methylprednisolone effect on brain volume and enhancing lesions in MS before and during IFNbeta-1b. Neurology 59, 688-694, doi:10.1212/wnl.59.5.688 (2002).

20 Weeda, M. M. et al. Upper cervical cord atrophy is independent of cervical cord lesion volume in early multiple sclerosis: A two-year longitudinal study. Mult Scler Relat Disord 60, 103713, doi:10.1016/j.msard.2022.103713 (2022).

21 Wang, Y. et al. A comparison of readout segmented EPI and interleaved EPI in high-resolution diffusion weighted imaging. Magn Reson Imaging 47, 39-47, doi:10.1016/j.mri.2017.11.011 (2018).

22 Popescu, V. et al. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. Neuroimage 61, 1484-1494, doi:10.1016/j.neuroimage.2012.03.074 (2012).

23 Smith, S. M. et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 17, 479-489, doi:10.1006/nimg.2002.1040 (2002).

24 Zheng, W., Chee, M. W. & Zagorodnov, V. Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. Neuroimage 48, 73-83, doi:10.1016/j.neuroimage.2009.06.039 (2009).

25 Valverde, S. et al. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. Neuroimage Clin 21, 101638, doi:10.1016/j.nicl.2018.101638 (2019).

26 Valverde, S. et al. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage 155, 159-168, doi:10.1016/j.neuroimage.2017.04.034 (2017).

27 Weeda, M. M. et al. Comparing lesion segmentation methods in multiple sclerosis: Input from one manually delineated subject is sufficient for accurate lesion segmentation. Neuroimage Clin 24, 102074, doi:10.1016/j.nicl.2019.102074 (2019).

28 Chard, D. T., Jackson, J. S., Miller, D. H. & Wheeler-Kingshott, C. A. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 32, 223-228, doi:10.1002/jmri.22214 (2010).

29 Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402-1418, doi:10.1016/j.neuroimage.2012.02.084 (2012).

30 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179-194, doi:10.1006/nimg.1998.0395 (1999).

31 Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195-207, doi:10.1006/nimg.1998.0396 (1999).

32 Guo, C., Ferreira, D., Fink, K., Westman, E. & Granberg, T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. Eur Radiol 29, 1355-1364, doi:10.1007/s00330-018-5710-x (2019).

33 Klasson, N. et al. Valid and efficient manual estimates of intracranial volume from magnetic resonance images. Bmc Med Imaging 15, doi:ARTN 5 10.1186/s12880-015-0045-4 (2015).

34 Cerri, S. et al. An open-source tool for longitudinal whole-brain and white matter lesion segmentation. Neuroimage Clin 38, 103354, doi:10.1016/j.nicl.2023.103354 (2023).

35 Puonti, O., Iglesias, J. E. & Van Leemput, K. Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling. Neuroimage 143, 235-249, doi:10.1016/j.neuroimage.2016.09.011 (2016).

36 Gronenschild, E. H. et al. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS One 7, e38234, doi:10.1371/journal.pone.0038234 (2012).

37 Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208-S219, doi:10.1016/j.neuroimage.2004.07.051 (2004).

38 Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20, 870-888, doi:10.1016/S1053-8119(03)00336-7 (2003).

39 Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 125, 1063-1078, doi:10.1016/j.neuroimage.2015.10.019 (2016).

40 Jbabdi, S., Sotiropoulos, S. N., Savio, A. M., Graña, M. & Behrens, T. E. Model-based analysis of multishell diffusion MR data for tractography: how to get over fitting problems. Magnetic resonance in medicine 68, 1846-1855 (2012).

41 Hernández, M. et al. Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs. PloS one 8 (2013).

42 Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F. & Woolrich, M. W. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34, 144-155, doi:10.1016/j.neuroimage.2006.09.018 (2007).

43 Behrens, T. E. et al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 50, 1077-1088, doi:10.1002/mrm.10609 (2003).

44 Zhang, H., Yushkevich, P. A., Rueckert, D. & Gee, J. C. in International conference on medical image computing and computer-assisted intervention. 211-218 (Springer).

45 Zhang, H., Yushkevich, P. A., Alexander, D. C. & Gee, J. C. Deformable registration of diffusion tensor MR images with explicit orientation optimization. Medical image analysis 10, 764-785 (2006).

46 Wang, Y. et al. Evaluations of diffusion tensor image registration based on fiber tractography. Biomedical engineering online 16, 9 (2017).

47 Keihaninejad, S. et al. An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer's disease. Neuroimage 72, 153-163, doi:10.1016/j.neuroimage.2013.01.044 (2013).

48 Chan, K., Marques, J.P. in Proc. 27th Annual Meeting of the ISMRM (Montreal, Canada, 2019).

49 Li, W., Avram, A. V., Wu, B., Xiao, X. & Liu, C. Integrated Laplacian-based phase unwrapping and background phase removal for quantitative susceptibility mapping. NMR Biomed 27, 219-227, doi:10.1002/nbm.3056 (2014).

50 Wood, T. C. W. QUIT: QUantitative Imaging Tools. Journal of Open Source Software 3, 656 (2018).

51 Smith, A. Symbol digit modalities test. (Western Psychological Services Los Angeles, 1973).

52 Nauta, I. M. et al. Performance validity in outpatients with multiple sclerosis and cognitive complaints. Mult Scler 28, 642-653, doi:10.1177/13524585211025780 (2022).

53 Lechner-Scott, J. et al. Can the Expanded Disability Status Scale be assessed by telephone? Mult Scler 9, 154-159, doi:10.1191/1352458503ms884oa (2003).

54 Cohen, J. A. et al. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. Arch Neurol 58, 961-967, doi:10.1001/archneur.58.6.961 (2001).

55 Cadavid, D. et al. The EDSS-Plus, an improved endpoint for disability progression in secondary progressive multiple sclerosis. Mult Scler 23, 94-105, doi:10.1177/1352458516638941 (2017).

56 Dayan, M. et al. MRI Analysis of White Matter Myelin Water Content in Multiple Sclerosis: A Novel Approach Applied to Finding Correlates of Cortical Thinning. Front Neurosci 11, 284, doi:10.3389/fnins.2017.00284 (2017).

57 Steenwijk, M. D. et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain 139, 115-126, doi:10.1093/brain/awv337 (2016).

58 Cagol, A. et al. Association of Brain Atrophy With Disease Progression Independent of Relapse Activity in Patients With Relapsing Multiple Sclerosis. JAMA Neurol 79, 682-692, doi:10.1001/jamaneurol.2022.1025 (2022).

59 Kiljan, S. et al. Cortical axonal loss is associated with both gray matter demyelination and white matter tract pathology in progressive multiple sclerosis: Evidence from a combined MRI-histopathology study. Mult Scler 27, 380-390, doi:10.1177/1352458520918978 (2021).

60 Bussas, M. et al. Gray matter atrophy in relapsing-remitting multiple sclerosis is associated with white matter lesions in connecting fibers. Mult Scler 28, 900-909, doi:10.1177/13524585211044957 (2022).

61 Weeda, M. M. et al. Damage in the Thalamocortical Tracts is Associated With Subsequent Thalamus Atrophy in Early Multiple Sclerosis. Front Neurol 11, 575611, doi:10.3389/fneur.2020.575611 (2020).

62 Carolus, K. et al. Time course of lesion-induced atrophy in multiple sclerosis. J Neurol 269, 4478-4487, doi:10.1007/s00415-022-11094-y (2022).

63 Schoonheim, M. M. et al. Disability in multiple sclerosis is related to thalamic connectivity and cortical network atrophy. Mult Scler 28, 61-70, doi:10.1177/13524585211008743 (2022).

64 Amiri, H. et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin 19, 466-475, doi:10.1016/j.nicl.2018.04.023 (2018).

CHAPTER 5

Summary, discussion and concluding remarks

Summary, discussion and concluding remarks

The aim of this thesis was to study the temporal and spatial relationships between white matter (WM) damage and grey matter (GM) atrophy in early stages of multiple sclerosis (MS) using multimodal magnetic resonance imaging (MRI) techniques, in order to better understand how neurodegeneration is related to neuroinflammation and other pathological changes in MS. Our main research question was to investigate whether GM atrophy progressed faster when there is more damage in the connected WM, and how this related to disability, in subjects with early relapsing-remitting (RR)MS.

We first reviewed the current literature on cross-sectional and longitudinal relationships between WM lesions and GM atrophy in **Chapter 1.2**. To improve our methodology with regard to WM lesions, we focused in **Chapter 2** on challenges related to lesions by optimizing lesion segmentation and developing lesion simulation software. Since MS is not only a disease of the brain, but also the spinal cord, we studied WM lesions and GM atrophy in the MS upper cervical cord in **Chapter 3**. In **Chapter 4**, we investigated the relationship between GM atrophy and WM damage in the connected tracts in two early RRMS cohorts.

A summary and discussion of the contents of this thesis is given below, with recommendations for future research. The last section gives a pointwise summary of the main conclusions that could be drawn from our research.

Chapter 1

Systematic review on WM lesions and GM atrophy in different types of MS

In **Chapter 1.2**, we reviewed existing literature on the cross-sectional and longitudinal relationships between WM lesions and GM atrophy in the brain in different types of MS. A total of 90 studies were included in the systematic review, from which some general conclusions could be drawn. First, higher WM lesion numbers and/or volumes generally correlated to more atrophy of global, deep and cortical GM, in particular in cross-sectional studies. This was mainly observed in clinically isolated syndrome (CIS) and RRMS. WM lesions were less consistently related to GM atrophy in progressive disease: relations were found between global WM lesion load and global GM atrophy, but less so when looking specifically to deep GM or cortical GM. The 20-year time frame of included articles, during which MRI technology continued to improve, meant that earlier studies had to be evaluated in the light of the concurrently available technology and knowledge. Large variability in image acquisition, analysis methods, and outcome measures, combined with uncertainties about potential confounders such as treatment, inhibited an interpretable meta-analysis of the results reported in the reviewed articles.

Technical factors that are known to affect measurements of brain volumes have been reviewed previously[1], and include (but are not limited to) differences between MRI scanners and acquisitions (e.g. field strength, pulse sequence types, 2D or 3D acquisitions, spatial and timing parameters). In addition, physiological variability and therapeutic interventions were not consistently accounted for in the papers reviewed, whereas both type and duration of treatment may affect WM lesion size and primary or secondary GM atrophy[2]. In addition, steroid

use is known to introduce pseudo-atrophy effects[3,4], which may influence the investigated WM-GM relations in studies where the suggested MRI delay after steroid use[5] was not applied. Furthermore, some studies reported brain volumes only after administration of the contrast-enhancing agent gadolinium, which is known to introduce variability in brain volume outcomes as well[6,7]. Finally, large variety exists in image (pre)processing tools and analysis software, which may affect both lesion segmentation/identification[8] and WM and GM tissue segmentation[9].

To harmonize MRI studies in MS, the European network for Magnetic Resonance Imaging in MS (MAGNIMS), the Consortium of Multiple Sclerosis Centres (CMSC) and the North American Imaging in Multiple Sclerosis Cooperative (NAIMSC) have published consensus recommendations on the use of MRI in MS[10,11]. Recommendations include the use of isotropic 3D FLAIR imaging; the use of cervical cord area as a measure for atrophy; mitigation of the impact of pseudo-atrophy on outcome measures; appropriate management of scanner-related, physiological and MS-related factors; and careful evaluation of the (pre)processing and software used for brain volume and lesion measurements.

It should be noted that the conclusions in this review were drawn from studies with conventional imaging from the past 20 years. In the meantime, an increasing amount of research studies do not only report focal lesions, but also microstructural WM damage, as measured with diffusion weighted imaging (DWI) or myelin weighted imaging (MWI). These microstructural effects on GM atrophy may be of interest for a systematic review in more recently published studies.

Chapter 2

Challenges related to lesions in MS

In **Chapter 2**, we focused on one of the main challenges in MRI measurements in MS: lesions. Most importantly, accurate lesion segmentation is needed to robustly and reliably measure lesion load in a way that is less labor intensive or time consuming than manual segmentation. Second, MRI brain segmentation software is prone to errors when MS lesions are present[12,13] since they have different image characteristics than the surrounding (normal-appearing) tissue, and may therefore influence tissue segmentation and atrophy measures. Optimization of the lesion segmentation and brain segmentation in the presence of lesions processes are thus needed for accurate measurements of both WM and GM damage in MS.

In this thesis, we first optimized lesion segmentation in the brain in **Chapter 2.1**. For this, we looked at different MS lesion segmentation software packages[14,15] with different levels of supervision and training (i.e. LesionTOADS; LST-LPA without and with optimization; BIANCA; and nicMSlesions without and with optimization) using 3D T1 and T2-FLAIR images in fourteen early RRMS subjects. Manual lesion segmentation was used as gold standard for both volumetric and spatial accuracy of the software packages tested. Highest accuracy compared to manual segmentation was obtained when training the deep learning method nicMSlesions on our fourteen manual segmentation images. For the lesion segmentations in **Chapter 4.2**, this trained nicMSlesions version with optimized lesion probability threshold

was used. Interestingly, we also found that input from only one manual segmentation in the nicMSlesions algorithm already resulted in higher volumetric and spatial accuracy than the other tested (semi-)automatic lesion segmentation methods. However, this was only true when the manual input image had enough data points to re-train the last two layers of the network, which – in our data – was determined to be a lesion volume of at least 3.0 ml. Training of the full nicMSlesions algorithm with multiple subjects – and therefore all 11-layers of the network – was not influenced when lower lesion volumes were used as input.
Although this research resulted in an optimized lesion segmentation method for our dataset, this does not necessarily translate to datasets with even lower lesion load[16], in the case of multi-center data[8], or data from other MR vendors or field strengths. With the publicly available datasets for MRI based MS lesion segmentation evaluation (e.g. from MICCAI[17] or ISBI[15]), new segmentation methods are still being developed[18], mostly with deep learning techniques such as convolutional neural networks[19,20] with 2D and 3D U-Nets[21-24]. In addition, lesion segmentation may be improved by adding other MRI contrasts, such as double-inversion recovery (DIR)[25], artificial DIR[26] or phase-sensitive inversion recovery (PSIR)[27] for (juxta)cortical lesions, and susceptibility-weighted imaging (SWI) measures such as T2* and R2* for iron rim lesions[28,29].

Accurate lesion segmentation is also of importance for GM and WM segmentation, since they are influenced heavily by the presence of lesions[9]. To minimize these effects, filling of the lesions with the signal intensity of the normal-appearing WM is applied, resulting in more accurate GM and WM segmentations[30]. However, even with lesion filling, segmentation software measurements are often not accurate[31]. In order to develop software that is robust in the presence of lesions, we introduced the lesion simulation software LESIM to produce MS lesions in healthy control images with realistic shapes, locations, intensities and signal-to-noise ratio in **Chapter 2.2**. With this tool, software can be developed where brain measurements in the healthy brain (i.e. the so-called ground truth) equal measurements in the brain with lesions present (i.e. the healthy brain with LESIM applied).
Although LESIM produces more realistic lesions than current lesion simulation methods[32,33], the software has not yet been tested in larger studies for its effect on image processing tools, and adjustments of the algorithm for applicability in other WM lesion diseases may be needed.

Chapter 3

Atrophy and lesions in the MS spinal cord

Upper cervical cord area (UCCA) can be used as a measure of spinal cord atrophy, and several (semi-)automated methods exist to measure UCCA. Similar as for the brain, not all methods are as robust or reliable in clinical settings[34] and the influence of lesions on these segmentations is unknown. Therefore, our study aim in **Chapter 3.1** was twofold: first, we wanted to compare five software packages (i.e. SCT-PropSeg, SCT-DeepSeg, NeuroQLab, Xinapse JIM, and ITK SNAP) in clinical whole-brain MR images that include the upper cervical cord. Because lesions may influence segmentation of tissue in the brain, we also wanted to investigate the sensitivity of the UCCA measurements to lesions in the upper cervical cord.

For the first aim, we used 3D T1 brain MR images from 21 MS and 6 HC subjects that underwent six different MR scans: a scan and re-scan (with repositioning) on three different 3T MR scanners (GE, Toshiba and Philips). Both spatial and volumetric measurement accuracy was determined for between-scanner robustness and within-scanner reproducibility. We observed high variability between scanners and between software methods, whereas within-scanner reproducibility (i.e. agreement between scan and re-scan) was generally high. With the use of manual segmentations, we determined the overall most accurate and reproducible segmentation method to be Xinapse JIM.
Since the spatial resolution was almost identical in all three scanners, it is unlikely that this was of major influence on the large UCCA differences found. We speculate that small variability in contrast between spinal cord and surrounding CSF (arising from acquisition differences, especially timing parameters) could lead to differences in partial volume effects at the border of the spinal cord. Previous research in which acquisition was homogenized between centers still showed low between-scanner agreement[34,35], supporting our finding that cross-sectional MUCCA cannot be easily compared in multi-center, multi-scanner research. This emphasizes the importance of relative instead of absolute UCCA comparisons in multi-vendor studies, e.g. with normalized UCCA percent change over time as a measure for upper cervical cord atrophy[36,37]. This has also been included in the 2022 MAGNIMS recommendations for harmonization of MRI data in MS multicenter studies, together with the suggestion of including a within-subjects identical scanning protocol to have reliable quantification of longitudinal spinal cord atrophy in MS[38].

For our second aim, we studied 3D T1 brain images of fourteen MS subjects with, and seven subjects without, upper cervical cord lesions from another dataset. Manual segmentations were used as so-called ground truth measurement, and spatial and volumetric accuracy was compared between the subjects with and without lesions. Agreement of the automated methods to manual did differ between scanners and between methods, but not between subjects with or without lesions. This suggests that UCCA measurement is robust in the presence of lesions. Therefore, we further investigated the relation of cervical cord lesions on atrophy of the cervical cord, and the relation between UCCA and cord lesions with disability in **Chapter 3.2**. For this, manual delineation of cord lesions was obtained from cervical 2D T2 images from C1 to C4 level in 36 early RRMS and 14 HC subjects over a two-year interval. UCCA was measured using Xinapse JIM, the method that showed the highest reproducibility in our previous study. At baseline, UCCA was already lower in early RRMS subjects compared to HC. Linear mixed model analysis for the relations between lesions and UCCA showed no relation at any of the time points, or at any of the cervical levels, thus not indicating a temporal or spatial relationship between upper cord lesions and atrophy. It has been suggested that not lesion size or number, but rather local demyelination[39], increased diffusivity[40] and decreased axonal and neurite density[41,42], were substrates of decreased UCCA in MS. This indicates the need for multi-modal MRI, preferably including DWI, in the MS spinal cord as well[43]. Although this has been studied in research settings, for example in post mortem or high-field studies[44,45], this is more demanding in clinical studies due to the long acquisition times[46].

Regarding the relationship with clinical disability, we included the Expanded Disability Status Scale (EDSS) scores, 9 hole peg test (9-HPT; measuring upper extremity function) and 25 foot walk test (25-HPT; measuring lower extremity function) at both time points. Lower baseline UCCA was related to increasing disability of the upper and lower extremities, but not to overall EDSS scores, suggesting that cervical cord atrophy is correlated to pyramidal, and mainly manual, impairment[47-49], rather than overall disability (i.e. EDSS). No associations between cervical cord lesion volume and disability was found in our early MS cohort, where this relation was present in more progressive MS phenotypes with moderate to severe disability[50-52] or long-standing MS[53]. Interestingly, the relationship of lower UCCA with worsening upper extremity function has also been reported in spinal cord injury due to other causes[54], which seems to support the hypothesis that overall spinal cord damage rather than MS-related inflammation (e.g. cervical cord lesions) may be the physiological basis of upper extremity disability.

Chapter 4

Grey and white matter damage in the MS brain

In **Chapter 4**, we studied the relationship between GM damage and WM damage in connected tracts in two early RRMS cohorts. In **Chapter 4.1**, we first focused on the thalamus, which is known to act as a relay station or hub between subcortical and cortical areas of the brain. Diffusion tensor imaging (DTI) was used to quantify WM damage in the brain of 72 subjects with early CIS or RRMS over a one year follow-up. Already over this short time period, we found significant GM atrophy, but no significant WM integrity changes, as defined by fractional anisotropy (FA) and mean diffusivity (MD). Most importantly, more thalamus atrophy could be predicted by reduced WM integrity (i.e. lower FA and higher MD) of the thalamic tracts at baseline. This was only found for the left, and not right, thalamus. Since baseline thalamus volume and FA and MD of the connected tracts differed significantly between both hemispheres, this might explain why the relationship described above was only found in the left hemisphere. Leftward thalamic asymmetry could be demonstrated in healthy individuals and seems to correspond to physiological conditions[55]. Interhemispheric asymmetry of brain diffusivity was described in normal individuals[56], as well as in MS, whereby increased apparent diffusion coefficients (ADCs) in the left thalamus were described[57], which is in line with our results demonstrating lower FA and higher MD values in the left thalamus.

The other way around, baseline thalamus volume could not predict changes in WM integrity in the thalamic tracts over time. These results suggest that WM damage precedes GM damage, instead of the other way around, in early MS, pointing towards a disease mechanism of retrograde degeneration in MS, which has been proposed before[58-62]. Beside the suggested phenomenon of Wallerian degeneration from distant WM lesions[63], our results in early MS patients with relatively low lesion volume, are in line with the indication of ongoing silent destructive processes of efferent and afferent thalamic projections which might induce measurable atrophy[64].

To further study the relationships between GM damage and connected WM damage, we performed another prospective longitudinal cohort study in which 13 HC and 31 RRMS subjects visited our center twice with a one year interval for MRI and clinical examinations in **Chapter 4.2**. In addition to standard DTI measures of FA and mean, axial and radial diffusivity (MD, AD, RD), we also included neurite orientation dispersion and density imaging (NODDI), myelin water fraction (MWF), and quantitative susceptibility mapping (QSM) as measures to quantify damage in the WM anatomically connected to both deep and cortical GM areas. Disability was measured with the Multiple Sclerosis Functional Composite score, a score based on upper extremity function (9-HPT), lower extremity function (25-FWT) and cognition (information processing speed from the Symbol Digit Modalities Test[SDMT]), thereby giving more information on disability than EDSS scores only, especially in early MS.

Linear mixed models showed that lower baseline WM integrity related to faster cortical atrophy over time, but that baseline GM volume/thickness did not relate to differences in WM integrity of the connected tracts over time. In addition, we found that this WM-GM relation could be better determined upon including NODDI and MWF measurements in addition to the standard DTI measures. Furthermore, this relationship appeared to be clinically relevant, since it was only present in subjects who would show increasing disability over the follow-up period, and less so in subjects who would not show increasing disability. Longer follow-up may enable us to further examine the relation with disability for this WM-GM relationship.

Our results suggest that WM damage precedes GM atrophy in early RRMS. Although we only found this relationship for the cortical, and not deep, GM, we discuss that this may have methodological reasons: where our linear mixed model with z-scores approach may be well suited for analysis of the CGM tracts (i.e. damage may occur in different tracts and different cortical areas between MS subjects) it may be less suitable for the analysis of DGM tracts. Analyses per DGM structure may be more informative, since lower WM integrity has been correlated to thalamus atrophy in previous studies[65-67]. Because of to the small sample size, we did not further investigate the separate DGM structures in this study, but this remains of interest for future, larger, studies.

WM integrity scores that included NODDI and MWF parameters showed stronger correlations with GM atrophy than when only diffusion tensor parameters were used. The added value of these measures over more simple diffusion parameters such as FA, has been shown previously[68-71], since they show greater specificity for both myelin and axon damage as confirmed in post-mortem studies[72-74]. Multi-shell DWI and MWF sequences are not included in the standard MRI protocol for MS[11], but clinically compatible protocols are currently being developed for future combined use with other pulse sequences[75]. However, our results suggest that inclusion of multi-shell DWI, and if possible MWF, in the MS research MRI protocol is highly recommended for future studies.

The longitudinal relationship where GM atrophy progresses faster when more damage in the connected WM is present was mainly found in subjects who would show increasing disability over the follow-up period. Although the distinction between clinical decliners and non-decliners in our study was subtle, exploratory regression analysis suggested the importance of baseline WM integrity over baseline GM scores on disability progression over time, thereby underlining

the importance of baseline WM (and not GM) damage on increasing disability. This may inform therapeutic choices: treatments that would be developed to target primarily neurodegeneration may have limited effect, since some atrophy may be secondary to WM damage, and baseline WM damage seems to drive disability more than baseline GM damage.

Concluding remarks

The aim of this thesis was to contribute to answering the chicken-or-egg question that remains in MS research: whether and how grey and white matter damage in MS are connected. The research presented in this thesis has led to the following conclusions:

Chapter 1 – Introduction

- In non-progressive MS, more WM lesions (size and/or number) correlated to more GM atrophy, both globally and regionally. In progressive MS, these correlations were less consistent.
- Harmonization of methodology in MRI studies of MS is of importance to enable between-study comparisons of data.

Chapter 2 – Challenges related to lesions in MS

- Lesion segmentation methods vary largely in their volumetric and spatial accuracy compared to manual segmentation.
- Using only one manual segmentation image as input for deep learning software (such as nicMSlesions) can already outperform commonly used (semi-)automated methods.
- LESIM is a new lesion simulation tool that enables the user to introduce realistic looking MS lesions with correct shapes, locations, intensities and signal-to-noise ratios in healthy control images. This allows optimization and validation of new segmentation software in MS.

Chapter 3 – Atrophy and lesions in the MS spinal cord

- Upper cervical cord area (UCCA) cannot be robustly reproduced between scanners or between vendors and comparisons in multi-center studies should therefore be made with caution.
- UCCA can be robustly measured when cervical cord lesions are present.
- There was no temporal nor spatial relation between UCCA and cervical cord lesions in our cohort of early RRMS subjects.
- Increasing cord atrophy, and not cord lesions, were predictive of increasing disability over time.

Chapter 4 – Grey and white matter damage in the MS brain

- Thalamic atrophy progresses faster when more damage is present in the connected WM in early RRMS and CIS.

- Cortical GM atrophy progresses faster when more damage is present in the connected WM in early RRMS subjects with increasing disability.
- WM integrity measures such as NODDI and MWF had additional value to standard DTI parameters in determining this WM-GM relationship, and including these acquisitions is recommended for future MRI research studies in MS.

References

1 Amiri, H. et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin 19, 466-475, doi:10.1016/j.nicl.2018.04.023 (2018).

2 De Stefano, N. et al. Evidence of early cortical atrophy in MS: relevance to white matter changes and disability. Neurology 60, 1157-1162 (2003).

3 Zivadinov, R. et al. Mechanisms of action of disease-modifying agents and brain volume changes in multiple sclerosis. Neurology 71, 136-144, doi:10.1212/01.wnl.0000316810.01120.05 (2008).

4 Cheriyan, J., Kim, S., Wolansky, L. J., Cook, S. D. & Cadavid, D. Impact of inflammation on brain volume in multiple sclerosis. Arch Neurol 69, 82-88, doi:10.1001/archneurol.2011.674 (2012).

5 Rao, A. B. et al. Methylprednisolone effect on brain volume and enhancing lesions in MS before and during IFNbeta-1b. Neurology 59, 688-694, doi:10.1212/wnl.59.5.688 (2002).

6 Lie, I. A. et al. The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis. Eur Radiol 32, 3576-3587, doi:10.1007/s00330-021-08405-8 (2022).

7 Warntjes, J. B., Tisell, A., Landtblom, A. M. & Lundberg, P. Effects of gadolinium contrast agent administration on automatic brain tissue classification of patients with multiple sclerosis. AJNR Am J Neuroradiol 35, 1330-1336, doi:10.3174/ajnr.A3890 (2014).

8 de Sitter, A. et al. Performance of five research-domain automated WM lesion segmentation methods in a multi-center MS study. Neuroimage 163, 106-114, doi:10.1016/j.neuroimage.2017.09.011 (2017).

9 de Sitter, A. et al. Reduced accuracy of MRI deep grey matter segmentation in multiple sclerosis: an evaluation of four automated methods against manual reference segmentations in a multi-center cohort. J Neurol 267, 3541-3554, doi:10.1007/s00415-020-10023-1 (2020).

10 Sastre-Garriga, J. et al. MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. Nature Reviews Neurology 16, 171-182, doi:10.1038/s41582-020-0314-x (2020).

11 Wattjes, M. P. et al. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol 20, 653-670, doi:10.1016/S1474-4422(21)00095-8 (2021).

12 Gonzalez-Villa, S. et al. Evaluating the effect of multiple sclerosis lesions on automatic brain structure segmentation. Neuroimage Clin 15, 228-238, doi:10.1016/j.nicl.2017.05.003 (2017).

13 Mortazavi, D., Kouzani, A. Z. & Soltanian-Zadeh, H. Segmentation of multiple sclerosis lesions in MR images: a review. Neuroradiology 54, 299-320, doi:10.1007/s00234-011-0886-7 (2012).

14 Styner M., L. J., Chin B., Chin M.S., Commowick O., Tran H., Markovic-Plese S., Jewells V., Warfield S. 3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation. The MIDAS Journal - MS Lesion Segmentation (MICCAI 2008 Workshop) (Nov 2008).

15 Carass, A. et al. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. Neuroimage 148, 77-102, doi:10.1016/j.neuroimage.2016.12.064 (2017).

16 Tran, P. et al. Automatic segmentation of white matter hyperintensities: validation and comparison with state-of-the-art methods on both Multiple Sclerosis and elderly subjects. Neuroimage Clin 33, 102940, doi:10.1016/j.nicl.2022.102940 (2022).

17 Commowick, O. et al. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. Sci Rep-Uk 8, doi:ARTN 13650 10.1038/s41598-018-31911-7 (2018).

18 Danelakis, A., Theoharis, T. & Verganelakis, D. A. Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. Comput Med Imaging Graph 70, 83-100, doi:10.1016/j.compmedimag.2018.10.002 (2018).

19 Salem, M., Ryan, M. A., Oliver, A., Hussain, K. F. & Llado, X. Improving the detection of new lesions in multiple sclerosis with a cascaded 3D fully convolutional neural network approach. Front Neurosci 16, 1007619, doi:10.3389/fnins.2022.1007619 (2022).

20 Rocca, M. A. et al. Deep Learning on Conventional Magnetic Resonance Imaging Improves the Diagnosis of Multiple Sclerosis Mimics. Invest Radiol 56, 252-260, doi:10.1097/RLI.0000000000000735 (2021).

21 Krishnan, A. P. et al. Multi-arm U-Net with dense input and skip connectivity for T2 lesion segmentation in clinical trials of multiple sclerosis. Sci Rep 13, 4102, doi:10.1038/s41598-023-31207-5 (2023).

22 Ashtari, P., Barile, B., Van Huffel, S. & Sappey-Marinier, D. New multiple sclerosis lesion segmentation and detection using pre-activation U-Net. Front Neurosci 16, 975862, doi:10.3389/fnins.2022.975862 (2022).

23 Kruger, J. et al. Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks. Neuroimage Clin 28, 102445, doi:10.1016/j.nicl.2020.102445 (2020).

24 Sarica, B., Seker, D. Z. & Bayram, B. A dense residual U-net for multiple sclerosis lesions segmentation from multi-sequence 3D MR images. Int J Med Inform 170, 104965, doi:10.1016/j.ijmedinf.2022.104965 (2023).

25 Geurts, J. J. et al. Intracortical lesions in multiple sclerosis: improved detection with 3D double inversion-recovery MR imaging. Radiology 236, 254-260, doi:10.1148/radiol.2361040450 (2005).

26 Bouman, P. M. et al. Artificial double inversion recovery images for (juxta)cortical lesion visualization in multiple sclerosis. Mult Scler 28, 541-549, doi:10.1177/13524585211029860 (2022).

27 Sethi, V. et al. Improved detection of cortical MS lesions with phase-sensitive inversion recovery MRI. J Neurol Neurosurg Psychiatry 83, 877-882, doi:10.1136/jnnp-2012-303023 (2012).

28 Dal-Bianco, A. et al. Long-term evolution of multiple sclerosis iron rim lesions in 7 T MRI. Brain 144, 833-847, doi:10.1093/brain/awaa436 (2021).

29 Ng Kee Kwong, K. C. et al. Rim lesions are demonstrated in early relapsing-remitting multiple sclerosis using 3 T-based susceptibility-weighted imaging in a multi-institutional setting. Neuroradiology 64, 109-117, doi:10.1007/s00234-021-02768-x (2022).

30 Guo, C. J., Ferreira, D., Fink, K., Westman, E. & Granberg, T. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. Eur Radiol 29, 1355-1364, doi:10.1007/s00330-018-5710-x (2019).

31 Popescu, V. et al. Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. Neuroimage Clin 4, 366-373, doi:10.1016/j.nicl.2014.01.004 (2014).

32 Chard, D. T., Jackson, J. S., Miller, D. H. & Wheeler-Kingshott, C. A. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J Magn Reson Imaging 32, 223-228, doi:10.1002/jmri.22214 (2010).

33 Senra, A. C. D., Simozo, F. H., dos Santos, A. C. & Murta, L. O. Multiple Sclerosis multimodal lesion simulation tool (MS-MIST). Biomed Phys Eng Expr 5, doi:ARTN 035003 10.1088/2057-1976/ab08fc (2019).

34 Lukas, C. et al. Quantification of spinal cord atrophy in MS: which software, which vertebral level, spinal cord or brain MRI? A multi-centric, longitudinal comparison of three different volumetric approaches. Mult Scler J 24, 88-90 (2018).

35 Papinutto, N. et al. Gradient nonlinearity effects on upper cervical spinal cord area measurement from 3D T1 -weighted brain MRI acquisitions. Magn Reson Med 79, 1595-1601, doi:10.1002/mrm.26776 (2018).

36 Lukas, C. et al. Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. J Neurol Neurosurg Psychiatry 86, 410-418, doi:10.1136/jnnp-2014-308021 (2015).

37 Valsasina, P., Rocca, M. A., Horsfield, M. A., Copetti, M. & Filippi, M. A longitudinal MRI study of cervical cord atrophy in multiple sclerosis. J Neurol 262, 1622-1628, doi:10.1007/s00415-015-7754-z (2015).

38 De Stefano, N. et al. MAGNIMS recommendations for harmonization of MRI data in MS multicenter studies. Neuroimage Clin 34, 102972, doi:10.1016/j.nicl.2022.102972 (2022).

39 Lee, L. E. et al. Cervical cord myelin abnormality is associated with clinical disability in multiple sclerosis. Mult Scler J, doi:Artn 13524585211001780 10.1177/13524585211001780 (2021).

40 Wolanczyk, M. et al. Diffusion tensor imaging of normal-appearing cervical spinal cords in patients with multiple sclerosis: Correlations with clinical evaluation and cerebral diffusion tensor imaging changes. Preliminary experience. Adv Clin Exp Med 29, 441-448, doi:10.17219/acem/116754 (2020).

41 Pravata, E. et al. Influence of CNS T2-focal lesions on cervical cord atrophy and disability in multiple sclerosis. Mult Scler J 26, 1402-1409, doi:Artn 1352458519865989 10.1177/1352458519865989 (2020).

42 Collorone, S. et al. Reduced neurite density in the brain and cervical spinal cord in relapsing-remitting multiple sclerosis: A NODDI study. Mult Scler 26, 1647-1657, doi:10.1177/1352458519885107 (2020).

43 Hori, M. et al. Advanced Diffusion MR Imaging for Multiple Sclerosis in the Brain and Spinal Cord. Magn Reson Med Sci 21, 58-70, doi:10.2463/mrms.rev.2021-0091 (2022).

44 Schmierer, K. et al. Quantifying multiple sclerosis pathology in post mortem spinal cord using MRI. Neuroimage 182, 251-258, doi:10.1016/j.neuroimage.2018.01.052 (2018).

45 Kreiter, D. J., van den Hurk, J., Wiggins, C. J., Hupperts, R. M. M. & Gerlach, O. H. H. Ultra-high field spinal cord MRI in multiple sclerosis: Where are we standing? A literature review. Mult Scler Relat Disord 57, 103436, doi:10.1016/j.msard.2021.103436 (2022).

46 Combes, A. J. E., Clarke, M. A., O'Grady, K. P., Schilling, K. G. & Smith, S. A. Advanced spinal cord MRI in multiple sclerosis: Current techniques and future directions. Neuroimage Clin 36, 103244, doi:10.1016/j.nicl.2022.103244 (2022).

47 Valsasina, P. et al. Cervical Cord T1-weighted Hypointense Lesions at MR Imaging in Multiple Sclerosis: Relationship to Cord Atrophy and Disability. Radiology 288, 234-244, doi:10.1148/radiol.2018172311 (2018).

48 Zurawski, J. et al. The impact of cervical spinal cord atrophy on quality of life in multiple sclerosis. J Neurol Sci 403, 38-43, doi:10.1016/j.jns.2019.04.023 (2019).

49 Valsasina, P. et al. Characterizing 1-year development of cervical cord atrophy across different MS phenotypes: A voxel-wise, multicentre analysis. Mult Scler, 13524585211045545, doi:10.1177/13524585211045545 (2021).

50 Kearney, H. et al. Cervical cord lesion load is associated with disability independently from atrophy in MS. Neurology 84, 367-373, doi:10.1212/WNL.0000000000001186 (2015).

51 Kerbrat, A. et al. Multiple sclerosis lesions in motor tracts from brain to cervical cord: spatial distribution and correlation with disability. Brain 143, 2089-2105, doi:10.1093/brain/awaa162 (2020).

52 Eden, D. et al. Spatial distribution of multiple sclerosis lesions in the cervical spinal cord. Brain 142, 633-646, doi:10.1093/brain/awy352 (2019).

53 Dekker, I. et al. Infratentorial and spinal cord lesions: Cumulative predictors of long-term disability? Mult Scler 26, 1381-1391, doi:10.1177/1352458519864933 (2020).

54 Freund, P. et al. Disability, atrophy and cortical reorganization following spinal cord injury. Brain 134, 1610-1622, doi:10.1093/brain/awr093 (2011).

55 Guadalupe, T. et al. Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex. Brain imaging and behavior 11, 1497-1514, doi:10.1007/s11682-016-9629-z (2017).

56 Fabiano, A. J., Horsfield, M. A. & Bakshi, R. Interhemispheric asymmetry of brain diffusivity in normal individuals: a diffusion-weighted MR imaging study. AJNR. American journal of neuroradiology 26, 1089-1094 (2005).

57 Fabiano, A. J. et al. Thalamic involvement in multiple sclerosis: a diffusion-weighted magnetic resonance imaging study. Journal of neuroimaging : official journal of the American Society of Neuroimaging 13, 307-314 (2003).

58 Calabrese, M. et al. Exploring the origins of grey matter damage in multiple sclerosis. Nat Rev Neurosci 16, 147-158, doi:10.1038/nrn3900 (2015).

59 Steenwijk, M. D. et al. What explains gray matter atrophy in long-standing multiple sclerosis? Radiology 272, 832-842, doi:10.1148/radiol.14132708 (2014).

60 Balk, L. J. et al. A dam for retrograde axonal degeneration in multiple sclerosis? J Neurol Neurosurg Psychiatry 85, 782-789, doi:10.1136/jnnp-2013-306902 (2014).

61 Haider, L. et al. The topograpy of demyelination and neurodegeneration in the multiple sclerosis brain. Brain 139, 807-815, doi:10.1093/brain/awv398 (2016).

62 Haider, L. et al. Multiple sclerosis deep grey matter: the relation between demyelination, neurodegeneration, inflammation and iron. J Neurol Neurosur Ps 85, 1386-1395, doi:10.1136/jnnp-2014-307712 (2014).

63 Louapre, C. et al. Heterogeneous pathological processes account for thalamic degeneration in multiple sclerosis: Insights from 7 T imaging. Multiple sclerosis (Houndmills, Basingstoke, England) 24, 1433-1444, doi:10.1177/1352458517726382 (2018).

64 Deppe, M. et al. Early silent microstructural degeneration and atrophy of the thalamocortical network in multiple sclerosis. Hum Brain Mapp 37, 1866-1879, doi:10.1002/hbm.23144 (2016).

65 Weeda, M. M. et al. Damage in the Thalamocortical Tracts is Associated With Subsequent Thalamus Atrophy in Early Multiple Sclerosis. Front Neurol 11, 575611, doi:10.3389/fneur.2020.575611 (2020).

66 Carolus, K. et al. Time course of lesion-induced atrophy in multiple sclerosis. J Neurol 269, 4478-4487, doi:10.1007/s00415-022-11094-y (2022).

67 Kolasinski, J. et al. A combined post-mortem magnetic resonance imaging and quantitative histological study of multiple sclerosis pathology. Brain 135, 2938-2951, doi:10.1093/brain/aws242 (2012).

68 Weeda, M. M. NODDI enables detection of subtle WM changes that lead to thalamic atrophy in subjects with RRMS. Mult Scler J 27, 134-740, doi:10.1177/13524585211044667 (2021).

69 Margoni, M. et al. Quantification of normal-appearing white matter damage in early relapse-onset multiple sclerosis through neurite orientation dispersion and density imaging. Mult Scler Relat Disord 58, 103396, doi:10.1016/j.msard.2021.103396 (2022).

70 Hagiwara, A. et al. White Matter Abnormalities in Multiple Sclerosis Evaluated by Quantitative Synthetic MRI, Diffusion Tensor Imaging, and Neurite Orientation Dispersion and Density Imaging. AJNR Am J Neuroradiol 40, 1642-1648, doi:10.3174/ajnr.A6209 (2019).

71 Sacco, S. et al. Neurite Orientation Dispersion and Density Imaging for Assessing Acute Inflammation and Lesion Evolution in MS. AJNR Am J Neuroradiol 41, 2219-2226, doi:10.3174/ajnr.A6862 (2020).

72 Laule, C. & Moore, G. R. W. Myelin water imaging to detect demyelination and remyelination and its validation in pathology. Brain Pathol 28, 750-764, doi:10.1111/bpa.12645 (2018).

73 Kozlowski, P., Rosicka, P., Liu, J., Yung, A. C. & Tetzlaff, W. In vivo longitudinal Myelin Water Imaging in rat spinal cord following dorsal column transection injury. Magn Reson Imaging 32, 250-258, doi:10.1016/j.mri.2013.12.006 (2014).

74 Falangola, M. F. et al. Histological correlation of diffusional kurtosis and white matter modeling metrics in cuprizone-induced corpus callosum demyelination. NMR Biomed 27, 948-957, doi:10.1002/nbm.3140 (2014).

75 Rahmanzadeh, R. et al. Myelin and axon pathology in multiple sclerosis assessed by myelin water and multi-shell diffusion imaging. Brain 144, 1684-1696, doi:10.1093/brain/awab088 (2021).

76 Dayan, M. et al. MRI Analysis of White Matter Myelin Water Content in Multiple Sclerosis: A Novel Approach Applied to Finding Correlates of Cortical Thinning. Front Neurosci 11, 284, doi:10.3389/fnins.2017.00284 (2017).

77 Steenwijk, M. D. et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. Brain 139, 115-126, doi:10.1093/brain/awv337 (2016).

78 Cagol, A. et al. Association of Brain Atrophy With Disease Progression Independent of Relapse Activity in Patients With Relapsing Multiple Sclerosis. JAMA Neurol 79, 682-692, doi:10.1001/jamaneurol.2022.1025 (2022).

79 Bodini, B. et al. White and gray matter damage in primary progressive MS: The chicken or the egg? Neurology 86, 170-176, doi:10.1212/WNL.0000000000002237 (2016).

80 Kiljan, S. et al. Cortical axonal loss is associated with both gray matter demyelination and white matter tract pathology in progressive multiple sclerosis: Evidence from a combined MRI-histopathology study. Mult Scler 27, 380-390, doi:10.1177/1352458520918978 (2021).

APPENDICES

1. Nederlandse samenvatting
2. List of author affiliations
3. PhD portfolio
4. Dankwoord
5. About the author

Appendix 1 – Nederlandse samenvatting

Het doel van dit proefschrift was om de temporele en spatiële relatie tussen schade in de witte stof (white matter, WM) en atrofie (hersenkrimp) van de grijze stof (grey matter, GM) te onderzoeken in vroege multiple sclerose (MS) door gebruik te maken van multimodale MRI, om beter te begrijpen hoe neurodegeneratie is gerelateerd aan neuroinflammatie en andere pathologische veranderingen in MS. Ons hoofddoel was om te onderzoeken of GM-atrofie sneller ontwikkelt wanneer er meer schade is in de verbonden WM, en hoe dit relateert aan klinische achteruitgang, in personen met vroege relapsing-remitting (RR)MS.
In **Hoofdstuk 1.2** hebben we de bestaande literatuur over de relatie tussen WM-laesies en GM-schade systematisch beoordeeld. Om onze methodologie met betrekking tot WM-laesies te verbeteren, hebben we in **Hoofdstuk 2** gericht op de uitdagingen die laesies met zich mee brengen, door het segmenteren van laesies te verbeteren en software te ontwikkelen om laesies te simuleren. Omdat MS niet alleen een ziekte is van het brein, maar ook van het ruggenmerg (myelum), hebben we de relatie tussen WM-laesies en GM-atrofie in het bovenste deel van het cervicale myelum bestudeerd in **Hoofdstuk 3**. In **Hoofdstuk 4** hebben we de relatie tussen GM-atrofie en WM-schade in de verbonden banen onderzocht in twee vroege RRMS-cohorten.
In deze samenvatting worden de bevindingen van het proefschrift per hoofdstuk kort samengevat, met aan het eind een opsomming van de belangrijkste conclusies die uit dit proefschrift getrokken kunnen worden.

In **Hoofdstuk 1.2** hebben we de bestaande literatuur naar cross-sectionele en longitudinale relaties tussen WM-laesies en GM-atrofie bediscussieerd in verschillende typen MS. In totaal zijn 90 studies bestudeerd en systematisch beoordeeld, waar een aantal algemene conclusies uit getrokken konden worden. Allereerst vonden we dat een groter volume en/of een groter aantal WM-laesies over het algemeen correleerden met meer GM-atrofie, zowel globaal als voor de diepe en corticale GM, en vooral in cross-sectionele studies. Dit vonden we voornamelijk bij de niet-progressieve typen MS, dus bij clinically isolated syndroom (CIS) en RRMS. WM-laesies waren minder consequent gerelateerd aan GM-atrofie in progressieve ziektebeelden: relaties werden wel gevonden tussen de totale hoeveelheid WM-laesies en totale GM-schade, maar niet specifiek voor diepe of corticale GM. De review sloot af met een aantal suggesties voor toekomstig onderzoek, waaronder het belang om technische variabiliteit binnen en tussen studies minimaal te houden, en het belang van transparantie van therapeutische interventies in de onderzochte populaties.

In **Hoofdstuk 2** hebben we onze focus gelegd op één van de belangrijkste uitdagingen in MRI-metingen in MS: laesies. Allereerst is accurate laesiesegmentatie belangrijk voor robuuste en betrouwbare metingen van laesievolume, op een manier die minder tijd en inspanning kost dan het geval is bij manuele segmentatie. Daarnaast is software voor de segmentatie van brein MRI-beelden gevoelig voor fouten wanneer MS-laesies aanwezig zijn, omdat deze laesies andere beeldkenmerken hebben dan het omliggende (normale) weefsel, en

daardoor weefselsegmentatie en atrofiemetingen kunnen beïnvloeden. Optimalisatie van laesiesegmentatie, en van breinsegmentatie in de aanwezigheid van laesies, is dus van belang voor accurate metingen van zowel WM- als GM-schade in MS.
In dit proefschrift hebben we eerst de laesiesegmentatie in het brein geoptimaliseerd in **Hoofdstuk 2.1**. Hiervoor hebben we gekeken naar vier MS-laesiesegmentatie softwarepakketten, met verschillende niveaus van supervisie en training (i.e. LesionTOADS; LST-LPA zonder en met optimalisatie; BIANCA; en nicMSlesions zonder en met optimalisatie) in 3D T1 en T2 FLAIR MRI-beelden in veertien personen met vroege RRMS. Manuele laesiesegmentatie werd gebruikt als gouden standaard voor zowel volumetrische als spatiële nauwkeurigheid van de geteste softwarepakketten. De grootste nauwkeurigheid in vergelijking met manuele segmentatie werd gevonden wanneer de deep learning methode nicMSlesions op alle manuele segmentaties werd getraind. Voor de laesiesegmentaties in **Hoofdstuk 4.2** hebben we deze methode met optimalisatie toegepast. Interessant was dat zelfs bij training op slechts één manuele segmentatie, het nicMSlesions-algoritme al hogere volumetrische en spatiële nauwkeurigheid vertoonde dan de andere geteste (semi-)automatische laesiesegmentatie-methodes.

Zoals hierboven aangegeven is nauwkeurige laesiesegmentatie ook belangrijk voor GM- en WM-segmentatie. Om de effecten van laesies te minimaliseren, wordt gebruikt gemaakt van het maskeren van de laesies door laesievulling: de signaalintensiteit van normaal uitziend WM wordt toegepast op de laesies, waardoor GM- en WM-segmentatie verbetert. Zelfs met deze laesievulling zijn de metingen van de segmentatiesoftware helaas nog niet altijd even nauwkeurig. Om software te kunnen ontwikkelen die robuust is wanneer laesies aanwezig zijn, introduceerden we in **Hoofdstuk 2.2** de laesiesimulatiesoftware LESIM om MS-laesies te produceren in beelden van gezonde controles met realistische vorm, locatie, intensiteit, en signaal-ruis-verhoudingen. Met behulp van LESIM kan software ontwikkeld worden waar breinsegmentaties van het gezonde brein (zogenaamde “ground truth” meting) gelijk is aan breinsegmentatie met laesies in het weefsel (i.e. het gezonde brein met LESIM toegepast). De software is vrij beschikbaar voor gebruik in MS MRI-onderzoek.

In **Hoofdstuk 3** hebben we ons toegespitst op het ruggenmerg of myelum in MS. Als meting voor ruggenmergatrofie wordt vaak gebruik gemaakt van “upper cervical cord area” (UCCA; oppervlak van een loodrechte doorsnede van het bovenste deel van het cervicale myelum), en er bestaan verschillende (semi-)automatische methoden om UCCA te meten. Net als bij breinsegmentaties zijn ook niet alle myelumsegmentaties even robuust of betrouwbaar in klinische scans, en daarnaast is het onbekend of deze softwarepakketten beïnvloed worden door de aanwezigheid van laesies in het cervicale myelum. Onze onderzoeksvraag in **Hoofdstuk 3.1** bestond daarom uit twee delen: allereerst wilden we vijf softwarepakketten voor het meten van UCCA vergelijken (i.e. SCT-PropSeg, SCT-DeepSeg, NeuroQLab, Xinapse JIM, en ITK-SNAP) in klinische MRI-beelden van het brein waar het bovenste cervicale myelum geïncludeerd is. Omdat laesies van invloed kunnen zijn op breinsegmentatie, wilden we ook

onderzoeken of UCCA-metingen beïnvloed worden door de aanwezigheid van laesies in het cervicale myelum.
Voor onze eerste onderzoeksvraag hebben we gebruik gemaakt van 3D T1 brein MRI-beelden van 21 MS en 6 gezonde personen, die ieder zes verschillende MRI-scans ondergingen: een scan en herhaalscan (waarbij de proefpersoon tussendoor uit de scanner was geweest) op drie verschillende 3T MR-scanners (GE, Toshiba en Philips). We berekenden de spatiële en volumetrische meetnauwkeurigheid tussen scanners en softwaremethoden (robuustheid), en tussen de twee scans op dezelfde scanner (reproduceerbaarheid). We vonden grote variaties in UCCA-uitkomsten tussen scanners en tussen softwaremethoden: vergelijkingen van UCCA-metingen in multicenter studies met verschillende MR-scanners en/of verschillende meetmethoden moeten dus met grote voorzichtigheid worden geïnterpreteerd. De reproduceerbaarheid binnen scanners (i.e. de overeenkomst tussen scan en herhaal-scan) was over het algemeen hoog. Door vergelijking met manuele segmentaties, vonden we dat de segmentatiemethode met de grootste nauwkeurigheid en reproduceerbaarheid in onze dataset Xinapse JIM was.
Voor de tweede onderzoeksvraag werden 3D T1 brein MRI-beelden (van één MRI-scanner) van 21 mensen met MS bestudeerd, waarvan veertien met, en zeven zonder, laesies in het bovenste deel van het cervicale myelum. Manuele segmentatie werd gebruikt als zogenaamde gouden standaard, en spatiële en volumetrische nauwkeurigheid werd vergeleken tussen de personen met en zonder laesies. Ook hier vonden we grote verschillen tussen methoden, maar niet tussen personen met of zonder laesies. Dit suggereert dat UCCA-metingen robuust zijn in de aanwezigheid van laesies.

We vervolgden ons onderzoek in **Hoofdstuk 3.2** met het onderzoeken van de relatie tussen cervicale myelumlaesies en UCCA, en de relatie tussen laesies en/of UCCA met klinische achteruitgang. Hiervoor werd manuele laesiesegmentaties van het cervicale myelum uitgevoerd in 2D T2 myelum MRI-beelden van myelumlevel C1 tot C4 in 36 vroege RRMS en 14 gezonde proefpersonen over een interval van twee jaar. UCCA werd gemeten met Xinapse JIM, de methode die in onze vorige studie de grootste reproduceerbaarheid en nauwkeurigheid vertoonde. Bij de start van de studie was UCCA al lager in mensen met RRMS dan in gezonde proefpersonen. Lineaire mixed model analyses voor de relatie tussen laesies en UCCA lieten geen relaties zien op enig tijdspunt of enig cervicaal level, en er werd dus geen temporele of spatiële relatie gevonden tussen cervicale myelumlaesies en atrofie.
Voor de relatie tussen cervicale myelumlaesies en/of UCCA met klinische achteruitgang includeerden we de Expanded Disability Status Scale (EDSS) score, de 9-hole peg test (9-HPT, een meting van armfunctie) en 25-foot walk test (25-FWT, een meting van beenfunctie) op beide tijdspunten. Lagere UCCA bij de start van het onderzoek was gerelateerd aan een grotere klinische achteruitgang van zowel arm- als beenfunctie, maar niet van totale EDSS-score. Er werden geen associaties gevonden tussen laesievolumes en klinische achteruitgang in dit cohort met vroege RRMS-proefpersonen. Deze resultaten suggereren dat cervicale myelumatrofie, en niet cervicale myelumlaesies, voorspellers kunnen zijn voor bepaalde

typen van klinische achteruitgang, maar dat er geen relatie lijkt te bestaan tussen cervicale myelumatrofie en –laesies.

In **Hoofdstuk 4** onderzochten we de relatie tussen WM-schade in verbonden banen aan GM-schade in twee vroege RRMS cohorten. In **Hoofdstuk 4.1** richtten we ons eerst op de thalamus, een belangrijk schakelstation, of hub, tussen subcorticale en corticale breinregio's. Diffusion tensor imaging (DTI) werd gebruikt om schade in de WM verbonden aan de thalamus te bepalen in 72 vroege CIS/RRMS personen over één jaar tijd. Over deze korte tijdsperiode vonden we al significante GM-atrofie, maar geen significante verandering in WM-integriteit (gemeten met fractional anisotropy [FA] en mean diffusivity [MD]). Onze belangrijkste bevinding was dat toenemende atrofie van de thalamus kon worden voorspeld door lagere WM-integriteit (i.e. lagere FA en hogere MD) van de banen verbonden aan de thalamus bij de start van het onderzoek. We vonden dit alleen voor de linker, en niet voor de rechter, thalamus. Omgekeerd vonden we géén voorspellende waarde van thalamusvolume bij de start van het onderzoek voor veranderingen over tijd in WM-integriteit in de banen verbonden aan de thalamus. Deze resultaten suggereren dat WM-schade voorafgaat aan GM-schade, in plaats van andersom, in vroege RRMS.

Om deze relaties verder te bestuderen, hebben we een prospectieve, longitudinale cohort studie uitgevoerd waarbij 13 gezonde en 31 RRMS proefpersonen ons centrum twee keer bezochten met een interval van één jaar voor MRI- en klinisch onderzoek in **Hoofdstuk 4.2**. Naast de standaard DTI-metingen van fractional anisotropy (FA) en mean, axial en radial diffusivity (MD, AD, RD), hebben we ook neurite orientation dispersion and density index (NODDI), myelin water fraction (MWF) en quantitative susceptibility mapping (QSM) toegevoegd als metingen om schade te kwantificeren in de WM die anatomisch verbonden is aan diepe en/of cortical GM regio's. Klinische achteruitgang werd gemeten met de Multiple Sclerosis Functional Composte (MSFC) score, een score gebaseerd op armfunctie (9-HPT), beenfunctie (25-FWT) en cognitie (informatieverwerkingssnelheid, via de Symbol Digit Modalities Test [SDMT]). Deze score geeft, vooral in vroege MS, meer informatie over achteruitgang dan alleen EDSS-scores.
Lineaire mixed model analyses lieten zien dat lagere integriteit van de WM bij de start van het onderzoek een voorspeller was voor meer corticale atrofie over tijd, maar dat GM-volume of -dikte bij de start van het onderzoek geen voorspeller was voor verschillen in WM-integriteit in de verbonden banen over tijd. Verder vonden we dat deze WM-GM-relaties beter konden worden gedefinieerd wanneer NODDI- en MWF-metingen werden toegevoegd aan de (standaard) DTI-metingen. Bovenal vonden we dat de WM-GM-relatie een klinisch relevante relatie was: deze relatie was vooral te zien was in personen die klinische achteruitgang lieten zien over de duur van de studie, en minder in personen waarbij dit niet het geval was. Ook deze resultaten suggereren dat WM-schade voorafgaat aan GM-schade in vroege RRMS. Hierbij was tevens te zien dat WM-integriteit bij de start van het onderzoek een belangrijkere voorspeller voor klinische achteruitgang was dan GM-scores, wat van belang kan zijn bij beslissingen over therapeutische behandelingen.

In conclusie heeft dit proefschrift nieuwe inzichten gegeven in de relatie tussen WM- en GM-schade, zowel temporeel als spatieel, in het vroege RRMS-brein en -ruggenmerg door gebruik te maken van multimodale MRI-technieken. Daarnaast hebben we nieuwe inzichten betreffende methodologische uitdagingen van MS MRI-metingen gegeven. Samenvattend kunnen uit dit proefschrift de volgende conclusies getrokken worden:

Hoofdstuk 1 – Introductie

- In niet-progressieve MS correleert een groter volume en/of een groter aantal WM-laesies met meer GM-atrofie, zowel globaal als regionaal. In progressieve MS waren deze correlaties minder consistent.
- Harmonisatie van methodologie in MS MRI-studies is van belang om data tussen studies te kunnen vergelijken.

Hoofdstuk 2 – Uitdagingen gerelateerd aan laesies in MS

- Softwaremethoden voor laesiesegmentatie verschillen sterk in hun volumetrische en spatiële nauwkeurigheid wanneer ze vergeleken worden met manuele segmentatie
- Het gebruik van slechts één manuele segmentatie als input voor de deep learning software nicMSlesions laat al betere nauwkeurigheid zien dan bij veelgebruikte (semi-)automatische methoden.
- LESIM is een nieuwe tool voor laesiesimulatie die de gebruiker in staat stelt om realistisch lijkende MS-laesies met de juiste vorm, locatie, intensiteit en signaal-ruis ratio's te introduceren in MRI-beelden van gezonde controles. Dit maakt optimalisatie en validatie van nieuwe segmentatiesoftware in MS mogelijk.

Hoofdstuk 3 – Atrofie en laesies in het MS ruggenmerg

- Upper cervical cord area (UCCA; oppervlak van het bovenste cervicale myelum) kan niet robuust worden gereproduceerd tussen scanners of tussen methoden, en UCCA-vergelijkingen bij multicenterstudies moeten daarom met grote voorzichtigheid worden gemaakt.
- UCCA kan robuust gemeten worden in de aanwezigheid van laesies in het cervicale myelum.
- Er was geen temporele of spatiële relatie tussen UCCA en laesies in het cervicale myelum in ons vroege RRMS-cohort.
- Myelumatrofie, maar niet de aanwezigheid van myelumlaesies, was voorspellend voor meer klinische achteruitgang over tijd.

Hoofdstuk 4 – Grijze en witte stof schade in het MS brein

- Thalamusatrofie ontwikkelt zich sneller wanneer meer schade aanwezig is in de verbonden WM in vroege RRMS en CIS.
- Corticale GM-atrofie ontwikkelt zich sneller wanneer meer schade aanwezig is in de verbonden WM in vroege RRMS personen, met name bij klinische achteruitgang.

- WM-integriteitmetingen zoals NODDI en MWF hadden toegevoegde waarde op de standaard DTI-parameters in de bepalingen van deze WM-GM-relaties, en de inclusie van deze sequenties is aan te raden voor toekomstige MRI-studies in MS.

Appendix 2 – Author affiliations

Amiri, Houshang	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Barkhof, Frederik	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands Institutes of Neurology and Healthcare Engineering, UCL, London, United Kingdom
Bellenberg, Barbara	Institute of Neuroradiology, St. Josef Hospital, Ruhr-University Bochum, Bochum, Germany
de Boer, Menno M.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Bø, Lars	Department of Clinical Medicine, University of Bergen, Bergen, Norway Norwegian Multiple Sclerosis Competence Centre, Department of Neurology, Haukeland University, Bergen, Norway
Brouwer, Iman	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Daams, Marita	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
van Dam, Maureen	Department of Anatomy and Neuroscience, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Dekker, Iris	Department of Neurology, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Gallo, Paolo	Department of Neuroscience, University of Padova, Padova, Italy
Hulst, Hanneke E.	Neuropsychology Unit, Institute of Psychology, Leiden University, Leiden, the Netherlands Department of Anatomy and Neuroscience, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Killestein, Joep	Department of Neurology, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Kuijer, Joost P.A.	Department of Neurology, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands

Lie, Ingrid Anne	Department of Clinical Medicine, University of Bergen, Bergen, Norway Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Lukas, Carsten	Institute of Neuroradiology, St. Josef Hospital, Ruhr-University Bochum, Bochum, Germany
Mattiesing, Rozemarijn M.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Middelkoop, Sander M.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Mol, Marijke A.E.	Medical Library, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Moraal, Bastiaan	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Myhr, Kjell-Morten	Department of Clinical Medicine, University of Bergen, Bergen, Norway Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway
van Nederpelt, David R.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Pouwels, Petra J.W.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Pruis, Ilanah J.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Schneider, Ruth	Institute of Neuroradiology, St. Josef Hospital, Ruhr-University Bochum, Bochum, Germany Department of Neurology, St. Josef Hospital, Ruhr-University Bochum, Bochum, Germany
de Sitter, Alexandra	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Steenwijk, Martijn D.	Department of Anatomy and Neuroscience, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Torkildsen, Øivind	Department of Clinical Medicine, University of Bergen, Bergen, Norway Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Bergen, Norway

van Tuijl, Rick J.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Twisk, Jos W.R.	Department of Clinical Epidemiology and Biostatistics, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands Institute of Health Science of the Vrije Universiteit, Amsterdam, the Netherlands
Uitdehaag, Bernard M.J.	Department of Neurology, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
de Vos, Marlieke L.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Vrenken, Hugo	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
de Vries, Myrte S.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Westerveld, Aimée S.R.	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands
Zywicki, Sofia	Department of Radiology and Nuclear Medicine, Amsterdam UMC – location VUmc, Amsterdam, the Netherlands Department of Neuroscience, University of Padova, Padova, Italy

Appendix 3 – PhD portfolio

Department	Radiology and Nuclear Medicine
PhD period	April 2016 – May 2023
Promotor	prof.dr. F. Barkhof
Copromotores	dr.ir. H. Vrenken
	dr. P.J.W. Pouwels

1. PhD training	Year	ECTS
General courses		
- Scientific integrity	2016	1.0
- eBROK ('Basiscursus Regelgeving Klinisch Onderzoek')	2016	1.5
- Writing a METc proposal	2016	2.0
- Writing a data management plan	2018	1.0
Subject specific courses		
- *In vivo* NMR	2016	1.0
- FMRIB FSL	2017	1.0
- FreeSurfer	2017	1.0
Seminars, workshops and masterclasses		
- Scientific writing	2017	0.3
- Weekly meeting Physics and Medical Technology	2016-2017	1.0
- Weekly meeting Neurology	2016-2019	2.0
- Weekly meeting Anatomy and Neuroscience	2016-2019	2.0
- Weekly meeting Radiology and Nuclear Medicine	2018-2020	2.0
- Weekly meeting Structural Brain Imaging group	2016-2023	7.0
Presentations		
- Poster presentation, ECTRIMS, Berlin	2018	0.5
- Oral presentation, ISMRM Benelux, Leiden	2019	0.5
- Poster presentation, iQuantivision, Amsterdam	2019	0.5
- ePoster presentation, ECTRIMS, Stockholm	2019	0.5
- Poster presentation (2x), ISMRM, Montreal	2019	1.0
- ePoster presentation, ISMRM, Vancouver	2021	0.5
- ePoster presentation, ECTRIMS, Vienna	2021	0.5

[continued on next page]

(Inter)national conferences		
- Amsterdam Neuroscience (participant)	2016-2017	0.6
- MS research days (including presentation)	2016-2017	0.8
- ISMRM Benelux (participant)	2017	0.3
- OHBM (participant)	2017	1.0
- ECTRIMS/ACTRIMS (participant)	2017	1.0
- Translational Neuroscience Network (participant)	2017	0.3
- MAGNIMS (participant)	2018	0.3
Other		
- Journal club	2017-2019	3.0
- Writing a grant proposal	2017	0.5
- Peer-review manuscripts for journals	2017-2020	1.0

2. Teaching	**Year**	**ECTS**
Lecturing		
- Neuroimaging practical, VUmc bachelor of medicine	2017	0.1
- HOVO course ‘ De hersenen van cel tot aan gedrag’	2016-2017	0.2
Tutoring, mentoring		
- Volunteer at Stichting Brein in Beeld	2016-2020	1.0
Supervising		
- MSc Neuroscience (Aimée Westerveld)	2017	1.0
- MSc Neuroscience (Ilanah Pruis)	2017	1.0
- BSc Medical natural sciences (Luuk Geelen)	2017	0.3
- BSc Medical natural sciences (Rick van Tuijl)	2017	0.3
- MSc Biomedical sciences (Sander Middelkoop)	2017	1.0
- BSc Medical natural sciences (Myrte de Vries)	2018	0.6
- MSc Medical natural sciences (David van Nederpelt)	2019-2020	1.0
- Neurology resident (Sofia Zywicki)	2020	0.6
- BSc Psychobiology (Suzanne de Jong)	2020	1.0
Other		
- Volunteer at Netherlands Brain Bank	2016-2021	1.0
- Volunteer at MS Web	2016-2021	1.0

3. Parameters of esteem	**Year**
Grants	
- Monique Blom-de Wagt grant (3500 euros)	2017
Other	
- Selected for ECTRIMS Summer School	2019

4. Publications	Year
Peer reviewed	
Weeda MM, Middelkoop SM, Steenwijk MD, Daams M, Amiri H, Brouwer I, Killestein J, Uitdehaag BMJ, Dekker I, Lukas C, Bellenberg B, Barkhof F, Pouwels PJW, Vrenken H. **Validation of mean upper cervical cord area (MUCCA) measurement techniques in multiple sclerosis (MS): High reproducibility and robustness to lesions, but large software and scanner effects.** *NeuroImage Clinical* 2019 (24) doi:10.1016/j.nicl.2019.101962	2019
Weeda MM, Brouwer I, de Vos ML, de Vries MS, Barkhof F, Pouwels PJW, Vrenken H. **Comparing lesion segmentation methods in multiple sclerosis: input from one manually delineated subject is sufficient for accurate lesion segmentation.** *NeuroImage Clinical* 2019 (24) doi:10.1016/j.nicl.2019.102074	2019
Weeda MM, Pruis IJ, Westerveld ASR, Brouwer I, Bellenberg B, Barkhof F, Vrenken H, Lukas C, Schneider R, Pouwels PJW. **Damage in the thalamocortical tracts is associated with subsequent thalamus atrophy in early multiple sclerosis.** *Frontiers in Neurology* 2020 (11) doi:10.3389/fneur.2020.575611	2020
Lie IA, Weeda MM, Mattiesing RM, Mol MAE, Pouwels PJW, Barkhof F, Torkildsen Ø, Bø L, Myhr KM, Vrenken H. **Relationship between white matter lesions and gray matter atrophy in multiple sclerosis: a systematic review.** *Neurology* 2022 (98) doi:10.1212/WNL.0000000000200006	2022
Weeda MM, Zywicki S, Brouwer I, Moraal B, Killestein J, Gallo P, Barkhof F, Pouwels PJW, Vrenken H. **Upper cervical cord atrophy is independent of cervical cord lesion volume in early multiple sclerosis: a two-year longitudinal study.** *Multiple Sclerosis and Related Disorders* 2022 (60) doi:10.1016/j.msard.2022.103713	2022
Weeda MM, Nederpelt DR, Twisk JWR, Brouwer I, Kuijer JPA, van Dam M, Hulst HE, Killestein J, Barkhof F, Vrenken H, Pouwels PJW. **Multimodal MRI study on the relation between GM atrophy and connected WM integrity and its effect on disability in early multiple sclerosis.** *Journal of Neurology (in press)* doi: N.A.	2023
de Sitter A, Weeda MM, Brouwer I, de Boer MM, van Tuijl RJ, Pouwels PJW, Barkhof F, Vrenken H. **Lesion simulation software LESIM: A robust and flexible tool for realistic simulation of white matter lesions.** *Under revision* doi: N.A.	t.b.d.

Appendix 4 – Dankwoord

Dit proefschrift is tot stand gekomen door het harde werk en de tomeloze inzet van velen die ik in dit hoofdstuk graag wil bedanken voor hun bijdragen.

Proefpersonen

Allereerst mijn allergrootste dank aan alle **proefpersonen**, zowel met als zonder MS, voor hun deelname aan ons onderzoek. Het enthousiasme waarmee jullie mee wilden meewerken aan het begrijpelijk maken van de ontwikkeling van deze rotziekte was voor mij extra motivatie om dit onderzoek zo goed mogelijk uit te voeren. Het onderzoek was zwaar: jullie moesten drie keer naar Amsterdam komen, en ondergingen daar een uitgebreid traject van onderzoeken, waaronder een MRI-scan van bijna 60 minuten, en cognitieve testen en vragenlijsten die soms erg confronterend konden zijn. Het kostte velen van jullie dagen energie om dit mogelijk te maken, maar desalniettemin kwamen jullie met veel enthousiasme steeds terug op de poli. Dankzij jullie zijn we weer een stap dichterbij het begrijpen van MS gekomen. De allerhoogste lof voor jullie allemaal!

Sponsors

Daarnaast wil ik ook gelijk hier aan het begin mijn dank richten aan de **Stichting MS Research** die het onderzoek mogelijk heeft gemaakt, en mij de ruimte heeft gegeven om (ondanks langdurig ziekteverlof) mijn promotie af te kunnen ronden. Ook vanuit het **Amsterdam UMC, locatie VUmc** en de afdeling **Radiologie en Nucleaire Geneeskunde** in het bijzonder is er veel extra moeite, tijd en geld gestoken om mij dit promotieonderzoek te kunnen laten volbrengen. Ik heb me altijd gesteund gevoeld, met ruimte om eerst aan mijn gezondheid te kunnen werken voordat ik aan mijn onderzoek verder ging werken, en daar ben ik iedereen zeer dankbaar voor.

Promotiecommissie

Beste **prof.dr. Barkhof**, beste **Frederik**, als mijn promotor wist je me bij de les te houden als ik me liet verzinken in alle details. Je hebt het mogelijk gemaakt dat ik na zeven jaar alsnog mijn proefschrift kon afronden. Het was een eer om jou, als gelauwerde hoogleraar met zelfs je naam verbonden aan MS-MRI criteria, als mijn promotor te mogen hebben. Ondanks het grote aantal studenten en onderzoekers dat je begeleid bij zowel het Amsterdam UMC als bij University College London, was je altijd goed op de hoogte van mijn project, en gaf je vaak prikkelende ideeën om het onderzoek te verbeteren of te verhelderen. Bedankt voor alles wat je me hebt geleerd.

Beste **dr.ir. Vrenken** en **dr. Pouwels**, beste **Hugo** en **Petra**, jullie waren tijdens mijn promotietraject mijn co-promotoren, maar eigenlijk waren jullie mijn werkouders. En zoals echte ouders betaamd, zaten jullie niet altijd op één lijn. Maar één ding waren jullie het altijd over eens: als ik met ideeën kwam, dan werd er altijd naar geluisterd. Veel van die ideeën hebben jullie me ook al in de eerste maanden van mijn promotietijd laten toepassen (was dit

oorspronkelijk niet een onderzoek met alleen maar een korte MRI, bedacht voor een fysicus?) en dat heb ik altijd heel bijzonder gevonden. Ik kreeg de ruimte om het onderzoek “van mij” te maken, en heb veel geleerd van de vrijheid en zelfstandigheid die jullie mij toevertrouwden. Bedankt voor alles wat jullie de afgelopen jaren voor mij hebben gedaan.

Beste **Petra**, jij bent een voorbeeld voor wie ik als wetenschapper zou willen zijn: lief, geduldig, leergierig, en vooral ontzettend slim. Je hebt me onderwezen in de verdere details van MRI, en met name de “speciale” sequenties, en altijd met veel geduld: ook als ik weer even uit de rotatie was en het niet meer wist, legde je het gewoon nog een keer uit. Het meest bijzondere aan jou als wetenschapper, en vooral aan jou als persoon, is dat je altijd tijd maakt voor je studenten en collega’s. Al is het ’s avonds laat, of ben je al zeventien dingen tegelijk aan het doen, bij jou was er altijd tijd en ruimte. Onze trip in Montreal zie ik wel als hoogtepunt, om 7 uur ’s ochtends al diep in de QSM fysica duiken, en er samen nog uren over napraten wat het allemaal voor ons onderzoek zou kunnen betekenen. Met alle ideeën die wij samen hadden voor vervolgstudies zouden we nog tien proefschriften kunnen schrijven. Gelukkig heb je me ook geleerd om af te bakenen: soms is goed ook gewoon goed, want perfect bestaat niet altijd in onderzoek. Er moet immers altijd ruimte blijven voor future recommendations!

Beste **Hugo**, binnen je onderzoeksgroep was er altijd ruimte voor een (woord)grapje, een koffie (uiteraard boven), en een praatje over hoe het nou écht met je gaat. Er was veel ruimte om het niet alleen te hebben over het onderzoek, maar ook over de ontwikkeling die je als promovendus maakt als onderzoeker, en als mens. Wie wil je zijn? Wat wil je doen? Waar ligt je grens? Dat heb ik altijd erg gewaardeerd. Ook tijdens mijn ziekte was je erg begripvol: je hebt me nooit gehaast, nooit gepusht om dingen af te ronden als je het idee had dat ik er misschien niet toe in staat was. Sterker nog, jullie moesten me vaak waarschuwen of ik niet tóch mezelf voorbij aan het lopen was als ik weer enthousiast een weekend in het VUmc doorbracht. Je kwaliteiten als wetenschapper, en hoe je verbonden bent met andere MS-onderzoekers in binnen- en buitenland, is een voorbeeld voor hoe wij allemaal ons onderzoek zouden moeten uitvoeren: samen. Ondanks dat je zelf helaas ook met ziekte te maken gehad hebt, bleef je betrokken bij het onderzoek. Ik ben blij dat we aan het einde nog met het hele team aan de afronding hebben kunnen werken.

Leescommissie

Beste **prof.dr. Uitdehaag**, beste **Bernard**, bedankt voor de tijd die je hebt gemaakt om voorzitter van de leescommissie te zijn en voor je begeleiding tijdens de Neurologie-MS meetings. Ik leer nu ook mijn studenten om de presentatie te beginnen met de conclusie! Beste **prof.dr. Reneman**, **dr. de Jong** en **dr. Roosendaal**, dank voor het lezen en beoordelen van mijn proefschrift. Beste **prof.dr. Wattjes**, beste **Mike**, wat fijn dat je in mijn leescommissie wilde plaatsnemen als oud-collega. We zullen nooit vergeten wat je voor Marco en mij hebt gedaan in de tijd dat je nog als radioloog bij ons werkte. Beste **dr. Schoonheim**, beste **Menno**, zo zitten we te barbecueën op een balkon in Vancouver, en zo zit je in mijn leescommissie! Bedankt voor de fijne samenwerking van de afgelopen jaren.

Paranimfen

Lieve paranimfen **Alexandra** en **Iman**, bedankt dat jullie vandaag naast mij willen staan, en dat jullie al die jaren tijdens mijn promotie óók naast me hebben gestaan. **Alexandra**, we begonnen samen aan onze promotie, en ik heb ontzettend veel aan je gehad, niet alleen als collega, maar vooral als vriendin. Jij bent natuurlijk al lang klaar, maar: ik ben er nu ook! Misschien moeten we nu dan toch maar een keer over die boekwinkel met maccaroons in Parijs gaan nadenken… Hoewel de herinneringen aan onze trips in o.a. Berlijn en Parijs (en vooral het wijntje in Montmartre) natuurlijk mooi zijn, herinner ik me vooral de tijden op de kamer waarbij we steun hadden aan elkaar, of gewoon even "over het scherm heen konden klagen" zonder oordeel van de ander. Je hebt mijn promotie kleur gegeven!

En dan **Iman**, hoe vaak heb ik jou niet gemaild met als onderwerp: "Help me Ob-iman Kenobi, you're my only hope!"? De hoeveelheid werk die jij (vaak letterlijk en figuurlijk achter de schermen) hebt verricht, zijn ontzettend belangrijk geweest voor mijn onderzoek. Daarnaast heb je mij "jouw" Vancouver laten zien, waar we samen de FSL-cursus mochten doen. Ik kan jaren later nog steeds de spierpijn in mijn kuiten voelen van de trappen naar Wreck Beach… Je bent een ontzettend open en prettig persoon om mee samen te werken, met een script voor elk probleem. Ik heb veel van je geleerd, en ben dankbaar voor alle tijd en moeite die je in mijn projecten hebt gestoken.

Structural Brain Imaging Group

Aan mijn andere collega's van SBIG: **Ronald**, **Alle Meije**, **Viktor**, **Houshang**, **Fabian**, **Marita**, **Rozemarijn**, en **Ingrid Anne**, bedankt voor alle discussies, journal clubs, presentaties, suggesties, leermomenten, lunches, koffies en diners. In het bijzonder dank aan de studenten die onze groep in alle jaren hebben versterkt, en mij hebben ondersteund in het onderzoek: **Ilanah**, **Aimée**, **Luuk**, **Rick**, **Sander**, **Myrte** en **Suzanne**. Een speciaal bedankje voor **David**, begonnen als "mijn" student, en nu je eigen promotietraject aan het doorlopen. Onze samenwerking heeft geleid tot prachtige resultaten, en ik heb veel van je mogen leren op programmeervlak. We maakten elkaar, en daarmee het onderzoek, beter, en daar is het laatste paper het bewijs van! Dear **Sofia**, you were a part of our group for a couple of months as a neurology resident, and you did an amazing job on our research on cervical lesions and atrophy. Thank you for our collaboration!

Amsterdam UMC

Mijn dank aan alle collega's van de afdeling **Radiologie en Nucleaire Geneeskunde** en van de (niet meer bestaande) afdeling **Fysica en Medische Technologie**. Speciale dank voor **Ton Schweigmann** voor alle uren die je met mij achter de MRI hebt doorgebracht. Je hebt me geleerd hoe je moet scannen, liet me meedraaien in het patiëntenprogramma, en hebt zelfs weekenden opgeofferd om mij te helpen om van alle proefpersonen de MRI's te maken. Je was een fantastisch mens om mee samen te werken, en ik hoop dat je met volle teugen kan genieten van je pensioen en je boot! Ook veel dank aan **dr. Joost Kuijer** voor je hulp met de MRI-sequenties en analyses, van QSM in het bijzonder. Aan **Marlieke de Vos** en **dr. Bastiaan**

Moraal veel dank en hulde voor het maken van de manuele laesiesegmentaties in het brein en het cervicale myelum, wat ons geholpen heeft om de methodologie sterk te verbeteren en nauwkeuriger naar laesies te kunnen kijken.
Dit onderzoek is een samenwerking van vele afdelingen, en ik ben dan ook blij dat ik bij zowel de afdeling **Neurologie** als **Anatomie en Neurowetenschappen** (ANW) mocht meekijken en meedraaien. Allereerst dank aan **prof.dr. Joep Killestein** en **prof.dr. Hanneke Hulst** voor hun klinische blik bij het opzetten, uitvoeren en analyseren van het onderzoek. Speciale dank aan **Martijn**, voor het leggen van de basis van dit promotieonderzoek, en aan **Iris**, **Quinten**, **Kim** en **Anand** voor de MRI-lessen. Dank aan Neurologie collega's **dr. Brigit de Jong**, **dr. Bob van Oosten**, arts-onderzoekers **Danko**, **Cyra**, **Marloes**, en alle collega's van de data-unit, voor de hulp bij het werven van patiënten en het leren van klinisch onderzoek. Dank ook aan mijn collega's bij ANW, in het bijzonder van de wijze lessen over wetenschap in het algemeen en neuroscience in het bijzonder van **prof.dr. Jeroen Geurts** en voor de analyse van de cognitieve data van **Maureen**. Aan **dr. Geert Schenk**, **dr. Linda Douw**, en **dr. Laura Jonkman**, en mijn mede-promovendi **Jolanda**, **Shanna**, **Svenja**, en **Marijn**: dank voor het warme welkom in jullie ANW-KNW team.
Aan **prof.dr. Jos Twisk** veel dank voor de prettige samenwerking en alle adviezen die je hebt kunnen geven met ons statistisch ingewikkelde onderzoek.
Dank ook aan alle andere collega's van het **MS Centrum Amsterdam** voor alle samenwerkingen, borrels, en uiteraard als hoogtepunt de heerlijke ECTRIMS-diners: de verbondenheid van de verschillende afdelingen die samenwerken binnen dit centrum was bijzonder om mee te mogen maken.

Collaborators

To my colleagues from the **Ruhr University** in Bochum: liebe **prof.dr. Carsten Lukas**, **dr. Barbara Bellenberg** and **dr. Ruth Schneider**, vielen dank für Eure Hilfe und Mühe mit den beiden Projekten bei denen ihr mich unterstützt haben. Sharing your longitudinal dataset from KNNMP and your help with the cervical cord segmentations were of great value to my thesis.
For help with the systematic review I would like to thank my colleagues from the **University of Bergen**, **dr. Øivind Torkildsen**, **dr. Lars Bø** and **dr. Kjell-Morten Myhr**, en van de Medische Bibliotheek van het Amsterdam UMC dank aan **Marijke Mol**.
I would like to thank my colleagues from the **University Padova**, **prof.dr. Paola Gallo** and **Sofia Zywicki**, for their collaboration on our cervical lesions and atrophy project.

Nevenactiviteiten

Dank aan het post-mortem MRI team van de **Nederlands Breinbank** en de mortuarium-medewerkers en pathologen van het VUmc voor alle kennis die ik heb opgedaan met dit zware, maar ook mooie, vrijwilligerswerk.
Dank aan de **Stichting Brein in Beeld** voor de leuke lessen, en gezellige uitjes, die we hebben mogen doen om neurowetenschap naar het publiek, en in het bijzonder naar kinderen, toe te brengen.

Dank aan **MS Web** voor het helpen bij de werving van proefpersonen, het mooie interview wat jullie met mij hebben gedaan, en voor de samenwerking bij het maken van vertalingen van wetenschappelijke artikelen voor de website.

Persoonlijk

Aan mijn lieve **vrienden** uit Leiden en omstreken: bedankt voor de vriendschappen die door weer en wind hebben mogen blijven bestaan. Ik proost met jullie op nog meer mooie borrels, feestjes, diners, promoties, zwangerschappen, bruiloften, en vooral op heel veel gezondheid voor jullie allemaal.

Aan mijn lieve **familie** Ad, Joke, Niels, Lars en Milan, en mijn lieve **schoonfamilie** Carola, René, Edwin, Ilse, Anouk, Debby, Koen, oma en omi: dank voor alle hulp, steun, gezelligheid en warmte die jullie mij al die jaren hebben gegeven.

Lieve **papa** en **mama**, jullie zijn mijn superhelden en hebben me altijd onvoorwaardelijk geholpen, ongeacht wat ik nodig had. Ik ben een gelukkig mens met zulke prachtige ouders.

Mijn allerliefste superman **Marco**: jij bent mijn thuis, mijn rots, mijn rust. Jij hebt me de mooiste tien jaar van mijn leven gegeven, en ik hoop er als bijna-getrouwd stel nog heel veel gelukkige jaren aan vast te mogen plakken. Mijn promotietraject was niet altijd makkelijk, zeker niet in combinatie met jouw baan als arts. De vele dagen zonder elkaar door mijn of door jouw werk, de verhuizingen en verbouwingen, en de vaste klantenpas bij alle ziekenhuizen in de omgeving van ons allebei… Maar: we hebben het gehaald, we zijn nu samen dokter en doctor! Dit is voor jou. Van hier tot aan de maan en terug.

Appendix 5 – About the author

Merline Maria (Merlin) Weeda was born on Sunday the 7th of July, 1991 in Rotterdam, the Netherlands. She grew up with her parents and two brothers in Numansdorp, Hoeksche Waard. During her high school period at the CSG Willem van Oranje in Oud Beijerland, she was diagnosed with musculoskeletal disease, impacting her final years of high school. After graduation she was accepted for the *numerus fixus* of the bachelor Biomedical sciences at Leiden University. After a couple of months, she moved to Leiden and became an active member of the study association of the medical faculty (Medische Faculteit der Leidse Studenten, M.F.L.S.), with great interest in educational sciences. She finished her bachelor's with an internship at the Human Genetics department of the Leiden University Medical Center (LUMC) in the Duchenne muscular dystrophy group. She continued with a two-year research master in Biomedical sciences with the neuroscience specialization. In her first year, she performed her junior internship at the LUMC department of Human Genetics again, this time on retinal vasculopathy with cerebral leukodystrophy (RVCL). After her first year, she interrupted here studies for a gap year to become a full-time member of the M.F.L.S. board as Education Member Biomedical sciences. In her second year as a research master student, she continued to be involved in both LUMC and M.F.L.S. boards and committees, and was chair of the National Biomedical sciences Student Organization (BMSO). Her major internship was performed at the University of Utrecht, department of Animals in Science and Society, and the Utrecht University Medical Center, department of Biomedical MR Imaging and Spectroscopy, where she researched the predictive properties of resting-state MRI on alcoholism in rats. She graduated and got her MSc degree in November 2015 with a 8.1 GPA.

Interested in the human brain and MRI techniques, she applied for a position at the VU University Medical Center, where she would be able to apply and expand her knowledge on the clinical, anatomical, and engineering aspects of multiple sclerosis research. She started her PhD in April 2016 in the department of Radiology and Nuclear Medicine under supervision of prof. dr. Frederik Barkhof, dr. ir. Hugo Vrenken and dr. Petra J.W. Pouwels on a grant from the Dutch MS Research Foundation entitled: "Does MS grey matter atrophy progress faster in regions with more damage in the connected white matter? A longitudinal study". In addition, she also volunteered as a translator for the website MS Web, as a teaching assistant at Stichting Brein in Beeld, and in the *post mortem* MRI team of the Netherlands Brain Bank. In November 2017, she got awarded the first Monique Blom-de Wagt grant for a two-month work visit to Oxford University for a collaboration with prof. dr. Mark Jenkinson on multi-modal surface matching. From the start of 2018, increasing physical disabilities prevented her from continuing to work full-time, and in 2020 she took a year of absence to rehabilitate. After rehabilitation, she continued working part-time, and finished her PhD thesis in May 2023.